

THÈSE
présentée devant
L'UNIVERSITÉ TOULOUSE III — PAUL SABATIER
pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ TOULOUSE III
Discipline : Biologie moléculaire
présentée et soutenue
par
Hervé SEITZ

Empreinte génomique parentale et petits ARN non-codants

Soutenance publique dans la salle de conférences du bâtiment IBCG, le 25 octobre 2004
à 14h30 devant la commission d'examen :

CAVAILLÉ Jérôme, Chargé de recherche, CNRS, Toulouse	directeur de thèse
CHARLIER Carole, Professeur associé, Université de Liège	rapporteur
CLERC Philippe, Chef de laboratoire, Institut Pasteur, Paris	rapporteur
CRIBBS David, Professeur, Université Toulouse III	président du jury
FORNÉ Thierry, Chargé de recherche, CNRS, Montpellier	examineur
VOINNET Olivier, Chargé de recherche, CNRS, Strasbourg	examineur

À tous ceux que j'aime.



SAINT-SATURNIN-SUR-LOIRE (MAINE-ET-LOIRE), vu de la vallée

Non ! non, c'est bien plus beau lorsque c'est inutile !
(Edmond Rostand, *Cyrano de Bergerac*, V 6)

Remerciements

Un pour tous, tous bourrins !

(Serge Florens citant Régis Lachaume, annuaire des élèves de l'É.N.S. 1997-1998)

Je tiens à remercier les membres du jury de ma thèse : Carole Charlier, Philippe Clerc, David Cribbs, Thierry Forné et Olivier Voinnet, pour avoir accepté de juger mes travaux, et pour m'avoir aidé à améliorer le manuscrit.

Je remercie évidemment Jérôme, qui a encadré ma thèse avec dévouement et enthousiasme. On a un peu le même caractère, tous les deux, et on a passé des moments extraordinaires. Combien de discussions passionnées ! Forcément, quand on n'était pas du même avis, elles pouvaient durer tard dans la nuit (pardon à sa famille, qui l'attendait autour d'une soupière froide ...). Malgré nos goûts musicaux, euh, orthogonaux, on s'est idéalement bien entendus ; et quand ce n'était pas au labo., c'était sur le bitume des petites courses de la Haute-Garonne, ou des berges du canal ... Parce que vous en connaissez beaucoup, vous, des chefs qui vous traînent aux courses à pied le dimanche ? Il est allé jusqu'à me prêter sa voiture pour emmener une jeune fille chez moi (véridique) !

Avec ses encouragements personnalisés et ses devises motivantes (« Qui peut le plus peut le plus »), il a su créer une ambiance merveilleuse avec les étudiants qu'il a encadrés (nous étions, ses trois thésards — Élène, Patrice et moi — les « gladiateurs de l'ère post-génomique »), et le moins qu'on puisse dire, c'est qu'avec lui, on s'ennuie rarement ... Tiens, mon seul regret : je n'étais pas encore là à l'époque des manips avec l'opossum (la pauvre bête avait réussi l'évasion du siècle, et s'était planquée sous les paillasses ; elle avait dû rendre les armes au terme d'une battue improvisée dans le labo.). Par contre, j'ai été le témoin privilégié de beaucoup des mésaventures quotidiennes de Jérôme, toutes aussi cocasses qu'in vraisemblables ... Je ne les relaterai pas ici (la crédibilité d'un brillant chargé de recherche est en jeu) (et puis, le chapitre « Remerciements » serait plus gros que le reste de la thèse, donc bon), mais je m'en souviendrai longtemps ...

Je vais donc entamer, en janvier prochain, mon « ère post-Jérômique » ; c'est sûr, beaucoup de choses vont me manquer, et la personnalité de Jérôme ne sera pas la moindre.

Je remercie également Jean-Pierre, qui a co-encadré mon stage de D.E.A. et le début de ma thèse, et qui dirigeait notre équipe jusqu'à l'an dernier. Sa connaissance profonde de notre sujet force le respect, et je garde aussi le souvenir de discussions intéressantes, sur des thèmes pas tous scientifiques, et dans des conditions pas forcément scientifiques.

Les membres de notre équipe ne sont évidemment pas pour rien dans le plaisir que j'ai eu à préparer ma thèse : aucune ne m'a jamais refusé son aide, et la bonne humeur de toutes contribuait à l'excellente ambiance qui planait au-dessus des paillasses ... Yvette (qui, avec Élène et Annie, a souvent été prise pour « Mme Seitz » au téléphone par des agents immobiliers, à l'époque de ma recherche d'appartement), virtuose du séquençage et de la synthèse d'oligos (avec sourire et célérité) ; Nicole, dont le récit des aventures herboristiques et artistiques égayait nos lundis ; Marie-Hélène, avec qui j'ai eu l'immense plaisir d'enseigner (voir plus bas), et grâce au dévouement de qui j'avais pu sauter dans mon avion et rendre mon rapport de D.E.A. dans les temps, il y a quelques années ; Marie-Line, qui gère avec sérieux et efficacité le quotidien du labo. (et dont les talents de cuisinière sont unanimement appréciés des étudiants de Jérôme ...) ; Béatrice, qui avait pris en main l'organisation des séminaires d'équipe, et qui s'était spécialisée dans les organismes modèles aux noms qui torschent (*Haloferax volcanii*, *Pyrococcus furiosus* et tout le tremblement) ; et Annie, dont les horaires de labo. s'apparentaient aux miens, avec qui j'ai eu d'agréables discussions, scientifiques ou non. Je me souviens en particulier d'un dîner, il y a deux ans (« Tu verras, sa cheminée est grande comme mon appartement », *dixit* un brillant chargé de recherche dont je tairai le nom pour épargner la crédibilité).

Je remercie aussi mes compagnons d'infortune, Élène et Patrice, contraints eux aussi de préparer leurs thèses dans cet environnement hostile. Élène, qu'on a vue grandir (quand elle est arrivée, elle n'était qu'une petite stagiaire de maîtrise), avait adopté les mêmes horaires que

moi, et on a passé de longues soirées à bosser à l'heure où les gens honnêtes dorment. Pour mon bonheur, Éléne est la seule personne au monde, avec moi, à apprécier (enfin, tolérer) les cassettes que j'écoutais le soir pour égayer mes Northerns, et il a fallu que ce soit ma collègue de paillasse ! Impec'. Et toujours un sourire poli quand je sors une blague (de celles qui m'attirent plutôt des quolibets et des regards contrits, de la part de publics moins avertis) ; non, vraiment, Éléne est une collègue idéale. J'ai passé de très bons moments avec Damien et elle (et Eliot, maintenant !), chez eux, chez moi, ou dans la salle de Vic ... Patrice, quant à lui, est mon aîné de quelques mois dans le labo. ; il m'avait intégré, à mon arrivée, parmi les jeunes du L.B.M.E., et nous nous sommes vraiment bien entendus. Il est ensuite allé commencer sa thèse outre-Rhin, puis au pays de Mozart et de Heïdi, et nous fait le plaisir, en 2004, de revenir travailler dans l'équipe. Toujours disposé à aider tout le monde, et avec le sourire, c'est un collègue extraordinaire ; je ne lui connais que des amis.

Je remercie aussi — et comment ! — mon *alter ego* de l'autre côté du couloir, ma complice de toujours, Patricia-de-chez-Kiss-en-face. Encore une accro. de la paillasse, sans doute la pire de nous tous ; du coup, quand j'avais envie d'aller faire l'andouille après 22h, ça ne pouvait tomber que sur elle, il n'y avait plus que nous dans le bâtiment. On a refait le monde si souvent, pendant que les gels séchaient et que les PCR migraient ... Star des transfections et des soirées α -T, leader charismatique des repas thésards et des congrès *SIFRA*^{RN}, c'est une personnalité marquante, presque traumatisante. Et l'enseignement ! Bon sang, il faut que je parle de l'enseignement : on a fait des T.P. et des T.D. ensemble, et ils font partie de mes meilleurs souvenirs. On était parfaitement complémentaires (elle avait l'autorité, et moi, je me chargeais du divertissement), et un subtil dosage de nos interventions donnait des résultats miraculeux, captivait l'attention des petiot, qui ressortaient de la salle avec des clonages plein la tête et des étoiles plein les yeux. Et c'était tellement gratifiant ! Ils rigolaient : on se charriait gentiment, on préparait des petites mises en scène : on se serait cru au théâtre ! Et à la fin du T.D., après que la salle se soit vidée, on repartait au labo., de bonne humeur pour la journée ...

Je remercie aussi tous les autres membres du L.B.M.E., une sacrée tribu, qui s'organise en deux clans : les « p'tits ARN » et les « chromatine » ; voyez plutôt : les membres de l'équipe Ferrer, spécialistes de la biochimie des particules ribonucléoprotéiques. Sous la direction bienveillante de Michou (débordante d'énergie et de gentillesse), évoluent Yves (dont la culture impressionnante et la générosité illimitée sont légendaires), Coralie (qui a pris l'adorable habitude de m'apporter les sandwiches du C.A.E.S. qu'elle ne prend pas le temps de manger), Régine, Simon, Jean-Paul et Julien, toujours accueillants et disponibles quand je m'aventure dans leur bout de couloir.

Les membres de l'équipe Gas (équipe Gleizes depuis peu), esthètes de la microscopie et de ses avatars du troisième millénaire, au premier rang desquels, la fameuse technologie de la cryofixation ultrarapide (non ? si !). Autour de Pierre-Emmanuel (un compagnon de table bien agréable, avec Jérôme et Yves ; il a toujours des histoires extraordinaires à raconter — celle de ses étudiants, qui lui empruntent des microscopes pour les envoyer dans l'espace ; les aventures américaines autour du lac Tahoe, avec Jérôme et Anthony ; la raison secrète pour laquelle son module de maîtrise fait recette auprès des étudiantes ; ...) se pressent Isabelle, Jacqueline, Marlène, Almass-Houd (toujours agréables avec tout le monde), Jacques (que je n'ai jamais vu sans son sourire, je ne sais pas quelle tête il aurait sans ; c'est lui, l'homme à la métaphore de la mobylette, devant le parterre ébahi des généreux donateurs de la Ligue contre le cancer du Gers), Valérie (toujours serviable, elle est devenue experte des réactions de séquençage) et Rosy (l'an dernier, elle atterrissait à peine du Liban qu'on la trimbalait tout un après-midi dans Toulouse, avec Jérôme ; le lendemain, on l'amenait au semi-marathon de Pechbusque, puis je la recrutais pour m'aider à faire l'électricité de ma cuisine ... quand même, quand j'y repense, c'était un accueil un peu traumatisant, elle a bon fond, pour ne pas nous en vouloir ...).

Les thématiques de l'équipe Kiss (pas de jeu de mots sur le nom de leur équipe, ils aiment

pas) embrassent tous les aspects de la biologie des ARN. Tamás (toujours réservé — sauf quand il donne son avis sur des « bullshit experiments » — il parle peu, mais juste ; au L.B.M.E., il est respecté comme un patriarche) s’est entouré de la fine fleur de la science européenne : Beáta, dont le talent et la volonté tranchent avec la timidité et la gentillesse ; Patricia (voir plus haut) dont le rire, euh, reconnaissable, et les exclamations sonores (« Oh my God ! ») traversent régulièrement les couloirs ; Arnold, capable de décimer une armée de thésards lorsqu’il a un peu bu ; Vera, qui rivalise de décibels avec Patricia lorsque les manip (qui n’ont qu’à bien se tenir) ne marchent pas comme elle voudrait (Vera (de l’autre côté du couloir) : « Meeeeerde ! » ; Nicole (dans notre labo.), calmement : « Tiens, elle a de moins en moins d’accent, Vera. ») ; Michel, devenu incollable sur la génomique des petits ARN non-codants, et avec qui j’ai eu d’intéressantes discussions ; Bruno et Sylvain, les ermites du fond du couloir (on raconte aux p’tits nouveaux qu’ils sont un peu sorciers) ; ils viennent de recruter Élodie, qui n’aurait pas pu mieux tomber ...

L’équipe Käs s’articule autour d’Emmanuel (un orateur exceptionnel ; ses séminaires sont toujours de vrais plaisirs) et Caroline (qui s’assure de la sécurité des conditions de travail de ses collègues avec application ; elle a toujours un mot gentil ou un sourire quand on la croise dans les couloirs), flanqués de leurs deux loupiots, Roxane (dont les qualités d’écraseuse de cervelle d’asticot entre lame et lamelle font le bonheur de son équipe) et Guillaume (jamais avare de son temps ou de sa bonne humeur, encore un pilier de la vie extra-scientifique du labo. ; c’est lui qui m’a emmené à mon premier concert — à 26 ans, il était temps — il y a six mois, quand il a fallu « déniaiser le gamin »).

Dans l’équipe Trouche, Didier (très dynamique, il mène de front foot, rugby et direction d’équipe) canalise les ardeurs de Sandrine (qui commence sa thèse), Olivier (un collègue très agréable, et un voisin bien sympa : après que j’eus visité un appartement de son immeuble, pas le moins du monde étonné de me voir débouler chez lui, il m’a invité à dîner avec des amis à lui : la soirée était très plaisante, le voisinage, donc, plutôt affable : allez zou, j’ai pris l’appartement et on s’est retrouvés voisins), Laurence (encore une bonne copine, avec qui j’ai passé du temps à parler de tout et de rien), Martine (avec qui j’ai un peu enseigné, il y a deux ans), Lucas (digne thésard de Laurence, il a hérité de son tempérament acharné), Cathy (la grande prêtresse de la culture cellulaire), et les post-docs Jérôme (qui va maintenant titiller les disques durs), Laetitia et Vinciane (rançon du succès, elles ne vont pas rester chez Didier, toutes les deux recrutées à la rentrée comme enseignantes-chercheuses).

L’équipe Grigoriev se penche sur la migration des jonctions de Holliday (non, rien à voir avec la transhumance des vacanciers pendant les ponts du mois de mai) : Micha encadre les travaux de Cynthia et Christophe, ses deux thésards, experts de la recombinaison sous caméra. Kamila (arrivée cette année, elle s’est tout de suite intégrée ; elle a pris en charge la gestion des stocks de consommables, et il se trouve qu’on en consomme, justement !) vient compléter l’effectif.

Dans l’équipe Richard-Foy, Hélène (qui rechigne rarement à se mêler aux bandes d’étudiants qui discutent dans les coins) travaille avec Kerstin, Mathilde (docteur depuis début septembre, elle s’apprête à entamer une nouvelle carrière), Sylvain (il retient toutes les infos intéressantes qui ont pu vous échapper, ou que vous avez oubliées, c’est bien pratique ; corollaire : ne vous fiez pas à son air innocent ...) et Françoise (une autre spécialiste de la culture cellulaire, elle se charge de former les petits jeunes).

Enfin, la plupart des tâches communes sont effectuées par Adri et Fatima, avec qui il est toujours agréable de discuter au détour d’un couloir.

Une mention spéciale pour les bourrins du labo., les acharnés de la pipette, ceux qui tutoient les veilleurs de nuit, et pour qui week-ends et jours fériés sont avant tout synonymes de disponibilité des centris et des paillasse : Christophe, Ali (partis, depuis, bourriner sous d’autres cieux), Élène, Patricia, Laurence, Jérôme, Yves, Lucas, ...

De nombreux amis du labo. m'ont filé de fiers coups de main pour mon installation dans mon appartement (la grande affaire de ces dix-huit derniers mois) : Jérôme (qui ne me prête pas seulement sa voiture pour que je ramène des demoiselles : il me l'a aussi prêtée pour que je puisse trimballer une ponceuse à parquet), Damien, Patricia et Guillaume, qui ont accepté de jouer les camionneurs pour des plaques de placo. ou des sacs de plâtre, et toute la bande de jeunes de l'I.B.C.G., qui se sont cotisés pour m'aider à me meubler ...

Je n'oublie pas non plus Claire, avec qui on faisait la course à la rédaction, cet été.

Je remercie aussi l'association α -T, pour l'animation scientifique (ah, le colloque ...) et parascientifique (avec les soirées — j'avais tort de dire qu'Élène est la seule personne, avec moi, à apprécier la musique que j'aime : il faut y ajouter Pascal et Damien, DJ irremplaçables des soirées α -T — la journée sport, et la garden-party) des labos de bio toulousains. Je ne rentre pas dans les détails, ce n'est pas l'endroit, mais je me doute qu'ils seraient de nature à bien faire rigoler les étudiants du labo. quand, dans quelques décennies, je serai un vieux chef respecté, dont ils s'amuseront à lire les remerciements de la thèse ...

Je remercie les membres des labos du voisinage, et en particulier Dominique Morello, Pascale Mercier, Bernard Monsarrat et Paul Alvinerie, avec qui j'ai travaillé.

Je remercie également tous mes instituteurs, professeurs et responsables de stage, pour m'avoir toujours intéressé à leurs histoires ; c'est grâce à eux que j'ai pris tant de plaisir dans mes études. Ils sont trop nombreux pour être tous cités ici, mais je n'oublie pas tout ce que je leur dois. Parmi mes responsables de stage, je remercie en particulier Mireille : c'est elle qui m'a initié à la biologie moléculaire, puis, un an plus tard (alors que je prospectais des labos de thèse), c'est elle qui m'a parlé de Jean-Pierre et Jérôme.

Je remercie aussi Yves, chef d'orchestre (à double titre) de la section « Courir » du C.A.E.S. du C.N.R.S. ; j'ai notamment profité de l'organisation du week-end à Espelette pour la Course des Crêtes, l'an dernier, et j'en garde un excellent souvenir (on avait mangé dans un petit restau. du front de mer, à St-Jean-de-Luz ...) (et puis on avait couru dans les crêtes, aussi, c'était loin d'être mal!). Il ne sera pas dit que je ne me suis fait des amis que dans le labo. : grâce à la course à pied, j'ai pu découvrir de nombreux villages de la Haute-Garonne, et je me suis fait de très bons amis. Eux non plus ne sont évidemment pas pour rien dans mon bonheur de vivre à Toulouse.

Je l'ai déjà évoqué plus haut : j'ai eu l'immense bonheur d'enseigner à l'Université Paul Sabatier. Mes petiots ont toujours été adorables : intéressés, marrants, sympas ; c'était tellement rafraîchissant ! Parfois déroutant (comment faut-il réagir quand les pipettes du deuxième rang se collent les avant-bras les unes aux autres pour comparer leurs bronzages, pendant l'analyse des effectifs de la F2 ? quand une étudiante facétieuse pince le dormeur du troisième rang, qui avait pris l'habitude de pioncer, bouche ouverte, une dizaine de minutes au bout de la première heure de T.D. ? quand une étudiante, dont la blouse était restée chez les parents, propose de faire le T.P. en robe de chambre ? ...), mais toujours tellement gratifiant, quand ils sont contents ; l'enseignement, c'est une expérience à vivre ... À présent, mes petiots (cette classe d'âge, que j'ai pour ainsi dire suivie : quand j'ai commencé ma thèse, et le monitorat, ils étaient en deuxième année de DEUG ; j'en ai retrouvé plusieurs l'année suivante, alors que j'enseignais principalement en licence ; puis encore plusieurs de ces licences quand, l'an dernier, j'ai enseigné en maîtrise) arrivent en M2R (D.E.A., dans l'ancienne terminologie), et je les vois déferler dans les labos. Morbleu, un cycle s'achève, il est temps que je m'en aille, et eux, mes petiots, vont maintenant me remplacer ...

Je bénéficie toujours de mon compte informatique à l'École normale supérieure, qui me permet de garder le contact (en grande partie grâce au forum informatique) avec les nombreux

amis que je m'y suis faits. Outre l'intérêt humain évident (chaque fois que j'y vais, je m'y sens encore chez moi, et y retrouve toujours des copains), le forum est un point d'accès privilégié à ce creuset où se côtoient tant de spécialistes de tous les domaines ... En ce qui me concerne, j'ai souvent profité des conseils de la fine équipe de la salle S, quand je ne me sortais pas de mes problèmes d'installation de Cygwin, de compilateur gcc, de syntaxe du langage de shell csh, et de commandes $\text{\TeX}$; je remercie donc tous ceux qui m'ont aidé : David Monniaux, Luc Habert, David Madore, Marc Espie, Marc Glisse, Laurent Rineau, Josselin Noirel et Nicolas George. Ce dernier en particulier, connu pour sa rapidité de réponse, n'a pas failli à sa réputation en réglant en 2 minutes 48 le problème qui me tourmentait, à des centaines de kilomètres de distance, le 15 août entre minuit et une heure du matin (en plus, c'était un dimanche, mais je pense que ça n'ajoute plus grand'chose ...).

J'ai énormément profité, au cours de ces trois ans, de notre collaboration avec les membres du laboratoire d'Anne Ferguson-Smith, en Angleterre, qui m'ont abondamment fourni en ARN d'embryons de Souris avec des disomies uniparentales, et avec la — désormais fameuse — délétion de la région différentiellement méthylée ... Je remercie donc Anne, Shau-Ping et Neil pour leurs efforts et pour leur gentillesse.

Je dédie cette thèse à tous ceux que j'aime, au premier rang desquels, bien sûr, ma famille. Ma sœur Géraldine et son mari Franck, qui m'ont fait le bonheur de me rendre tonton, de Lilou, il y a presque un an et demi ; mon frère Clément, qui a partagé mon aventure toulousaine ; ma petite sœur Annabelle, qui, à trois jours de ses quinze ans, aurait pourtant eu beaucoup plus intéressant à faire que venir à Toulouse écouter ma thèse ... Et je remercie mes parents, qui m'ont permis de faire les études qui me plaisaient, et qui ont toujours été à mes côtés ; pour leur soutien permanent, et pour tellement d'autres choses encore, pour tout leur amour, je ne les remercierai jamais assez.

Section M. et Mme.

M. et Mme. Dansleslocisoumisàlempreintegénomiqueparentale ont un fils ; comment s'appelle-t-il ?

Réponse :

Wladimir

M. et Mme. Hervéamislefeuaulabodepatriciaenmettantungelàpré ont un fils ; comment s'appelle-t-il ? (© G. Susbielle)

Réponse :

René

Abréviations

ADN :	acide désoxyribonucléique
ARN :	acide ribonucléique
ARN _m :	ARN messenger
ARN _{pré-r} :	précurseur d'ARN ribosomique
cDNA :	ADN copie (issu de la reverse-transcription d'un ARN)
DMR :	région différentiellement méthylée (voir page 25)
IC :	centre d'empreinte (voir page 23)
kb (ou : kpb) :	kilo-paire de bases
Mb (ou : Mpb) :	méga-paire de bases
nt :	nucléotide
pb :	paire de bases
RdRP :	polymérase à ARN dépendant de l'ARN (voir page 64)
RISC :	complexe effecteur du RNAi (voir page 63)
RITS :	complexe répresseur de la transcription guidé par des siRNA (voir page 83)
RNAi :	interférence à ARN
RNase :	ribonucléase
RT-PCR :	reverse-transcription couplée à une amplification par polymérisation en réaction en chaîne
siRNA :	petit ARN interférent (voir page 62)
snoRNA :	petit ARN nucléolaire (voir page 46)
UTR :	région non traduite d'un ARN messenger

Note

Dans tout le manuscrit, les « ou » sont des « ou » inclusifs, susceptibles d'être remplacés par des « et/ou ».

Table des matières

1	Introduction	19
1.1	Avant-propos	20
1.2	L’empreinte génomique parentale	21
1.2.1	Historique	21
1.2.2	Les loci soumis à l’empreinte génomique parentale	23
1.2.3	Diversité des mécanismes	24
1.2.4	Considérations évolutives	37
1.3	Les ARN C/D	46
1.3.1	Structure des ARN C/D et composition des RNP C/D	46
1.3.2	Biosynthèse des ARN C/D	47
1.3.3	Fonctions des ARN C/D	49
1.3.4	ARN C/D orphelins	53
1.3.5	Gènes d’ARN C/D soumis à l’empreinte génomique parentale	53
1.4	MicroARN, siRNA et leur métabolisme	59
1.4.1	Historique	59
1.4.2	Biogenèse des microARN et siRNA	65
1.4.3	Modes d’action	68
1.4.4	Fonctions biologiques	74
1.4.5	RNAi et guidage de l’hétérochromatinisation	80
1.5	ARN non-codants et contrôle transcriptionnel	91
1.5.1	ARN non-codants et compensation de dose	91
1.5.2	ARN et hétérochromatine péricentromérique chez les Mammifères	97
1.5.3	ARN non-codants et contrôle transcriptionnel spécifique chez les Mammifères	97
2	Résultats	101
2.1	Un deuxième locus de gènes d’ARN C/D répétés	102
2.1.1	Article n°1	103
2.1.2	Recherche de nouvelles répétitions de gènes d’ARN C/D	115
2.2	Le locus 14q32/distal 12 contient des gènes de microARN	118
2.2.1	Article n°2	120
2.2.2	Stabilisation de l’ARN _m Rtl1 en l’absence de miR-127 et miR-136	122
2.2.3	Article n°3	124
2.3	Une concentration de gènes de microARN	130
2.3.1	Article n°4	132
2.3.2	Six gènes de microARN additionnels	140
2.3.3	Le phénotype callipyge du Mouton : implication des microARN ?	140

3 Conclusion	145
3.1 Bilan	146
3.2 Discussion	146
3.2.1 Ces petits ARN non-codants ont-ils une fonction ?	146
3.2.2 Un rôle des petits ARN non-codants dans le développement ?	147
3.2.3 Des ARN-cibles ?	149
3.2.4 Gènes de petits ARN répétés et empreinte génomique parentale : une association étonnante	152
3.2.5 Expression des ARN non-codants du locus 14q32	159
3.2.6 L'ARN, nouveau champ de bataille de la guerre des sexes ?	160
3.3 Perspectives	161
4 Annexe 1	163
4.1 Article n°5	164
5 Annexe 2	173
5.1 Problématique	174
5.1.1 Données présentées	174
5.1.2 Séquences utilisées	177
5.2 Principe des calculs	178
5.2.1 Probabilité de trouver au moins une complémentarité parfaite dans la séquence-cible	178
5.2.2 Nombre le plus probable de complémentarités	179
5.3 Résultats de l'analyse	180
5.3.1 Probabilité de trouver au moins une complémentarité parfaite dans la séquence-cible	180
5.3.2 Nombre le plus probable de complémentarités	186
5.3.3 Complémentarités aux snRNA	191
5.3.4 Complémentarités hors de l'élément antisens	192
5.4 Un rôle peu probable dans la biogenèse des ARN _r et snRNA	194
6 Annexe 3	195
6.1 Article n°6 : revue	196
Bibliographie	203
Index	244

Table des figures

1.1	LE CYCLE DE L'EMPREINTE.	26
1.2	LE LOCUS <i>Igf2r/Air</i>	32
1.3	LE LOCUS <i>Kcnq1</i>	34
1.4	LE LOCUS <i>Igf2/H19</i>	35
1.5	LE LOCUS PWS/AS	36
1.6	LE LOCUS <i>Dlk1/Gtl2</i>	37
1.7	OBTENTION DE SOURIS À DISOMIES UNIPARENTALES.	41
1.8	STRUCTURE DES ARN C/D GUIDES DE MÉTHYLATION.	47
1.9	LA 2'-O-MÉTHYLATION.	52
1.10	LES GÈNES D'ARN C/D DU LOCUS HUMAIN 15q11-13.	56
1.11	COMPLÉMENTARITÉ ENTRE L'ARN _m 5-HT _{2C} ET L'ARN C/D MBII-52/HBII-52.	57
1.12	LIN-4, MEMBRE FONDATEUR DE LA FAMILLE DES MICROARN.	60
1.13	MODÈLE DE MATURATION DES MICROARN DE MAMMIFÈRES.	66
1.14	MODÈLE D'INTERACTION DE DICER AVEC SON SUBSTRAT.	67
1.15	MODÈLES HISTORIQUES DE FONCTIONNEMENT DE L'INTERFÉRENCE À ARN.	70
1.16	ORGANISATION GÉNOMIQUE DES GÈNES HOX.	77
1.17	ORGANISATION DES CENTROMÈRES DE <i>Schizosaccharomyces pombe</i>	81
1.18	RNAi ET GUIDAGE DE L'HÉTÉROCHROMATINISATION DES RÉPÉTITIONS CENTROMÉRIQUES.	84
1.19	LA CONJUGAISON DES CILIÉS.	86
1.20	MODÈLE DE SÉLECTION DES SÉQUENCES ÉLIMINÉES LORS DU DÉVELOPPEMENT MACRONUCLÉAIRE DE <i>Tetrahymena</i>	87
1.21	LA DÉLÉTION α^{-ZF}	98
2.1	RECHERCHE DE GÈNES DE SNORNA C/D DANS LE CHROMOSOME 14 HUMAIN (ORIENTATION SENS).	116
2.2	RECHERCHE DE GÈNES DE SNORNA C/D DANS LE CHROMOSOME 14 HUMAIN (ORIENTATION ANTISENS).	116
2.3	LE GÈNE <i>Rtl1</i> EST UN RÉTROTRANSPOSON CONSERVÉ CHEZ LES MAMMIFÈRES.	118
2.4	DÉLÉTION DE LA DMR DU LOCUS <i>Dlk1/Gtl2</i>	123
2.5	LA MUTATION CALLIPYGE DÉRÈGULE LES GÈNES DU LOCUS 18q OVIN.	141
3.1	MODÈLE D'INVASION DU LOCUS PAR DES RÉTROGÈNES.	153
3.2	EMPREINTE GÉNOMIQUE PARENTALE ET RÉPÉTITIONS DE GÈNES DE PETITS ARN NON-CODANTS.	157
3.3	CONTRIBUTION POST-TRANSCRIPTIONNELLE À L'EMPREINTE ET À LA RÉPRESSION DE <i>Rtl1</i> PAR RNAi.	159
5.1	GUIDAGE DE LA MÉTHYLATION DU 2'-O DU RIBOSE PAR LES ARN C/D.	174

5.2	COMPLÉMENTARITÉ NE FAISANT PAS INTERVENIR LE NUCLÉOTIDE IMMÉDIATEMENT EN AMONT DE LA BOÎTE.	174
5.3	ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D 14q32(II).	175
5.4	ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D HBII-52.	176
5.5	ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D HBII-85.	177

Liste des tableaux

1.1	GÈNES D'ARN NON-CODANTS ET GÈNES À EMPREINTE RÉCIPROQUE CHEZ LA SOURIS.	31
1.2	GÈNES MURINS SOUMIS À L'EMPREINTE GÉNOMIQUE PARENTALE.	45
1.3	LES ARN C/D ORPHELINS DE MAMMIFÈRES.	54
2.1	RÈGLES DE TRANSMISSION DU PHÉNOTYPE CALLIPYGE CHEZ LE MOUTON. . . .	140
5.1	COMPLÉMENTARITÉS ENTRE LES ARN C/D DES LOCI HUMAINS 15q11-13 ET 14q32 ET LE PRÉCURSEUR D'ARN _r HUMAIN DE 45S.	193

Chapitre 1

Introduction

Tout ce qui est petit est gentil.
(devise du journal *La Hulotte*)

1.1 Avant-propos

On appelle « gènes » les unités d'information qui gouvernent les caractères transmissibles à la descendance. Le XX^{ème} siècle a vu l'identification du support physique des gènes, l'ADN chez tous les êtres vivants, et la compréhension progressive des principes de codage de l'information génétique : les phénotypes sont déterminés par les propriétés des molécules qui constituent les cellules, lesquelles dépendent d'activités enzymatiques, portées invariablement par des ARN ou des protéines. Les protéines elles-mêmes sont synthétisées selon l'information portée par des ARN messagers, et, ultimement, les ARN sont issus de la transcription de l'ADN. L'ADN, quant à lui, est répliqué à l'identique par des systèmes spécialisés, et transmis à la descendance.

Il a donc semblé que toute l'information nécessaire au fonctionnement des cellules résidait dans la séquence du génome de l'espèce considérée ; que la simple lecture de cette séquence fournissait des données suffisantes pour prédire la nature des macromolécules (ARN et protéines, et par suite, tous les composants cellulaires, produits de leurs activités enzymatiques) représentés dans les cellules, et donc, simuler le fonctionnement cellulaire. C'est cette perspective de déduction d'une connaissance complète de la biologie d'un organisme, à partir de la lecture de sa séquence génomique, qui a fait écrire à W. Gilbert¹ à propos du séquençage du génome humain : « Trois milliards de bases de séquence peuvent tenir sur un seul disque compact, et l'on pourra sortir un CD de sa poche, et dire : “ Voici un être humain ; c'est moi ! ” » (Gilbert, 1992).

En réalité, d'autres messages que la séquence de l'ADN affectent qualitativement et quantitativement l'expression des gènes ; les modifications responsables de l'altération de l'expression des gènes qui ne sont pas dues à des modifications de la séquence de l'ADN, mais qui sont quand même hérissables par division cellulaire, sont dites « épigénétiques ». Les modifications épigénétiques que l'on connaît sont toutes des modifications chimiques de l'ADN (méthylation des cytosines) ou de protéines qui lui sont associées, les histones (méthylation, acétylation et phosphorylation de résidus bien particuliers). L'information qui contrôle les caractéristiques innées d'un organisme ne se résume donc pas à une séquence de « A », de « C », de « G » et de « T », il existe d'autres codages, moins bien connus pour le moment, qui se superposent à cette information : chez les Eucaryotes, des patrons particuliers de modifications chimiques de la chromatine recrutent des complexes qui altèrent l'expression des gènes du voisinage. L'une des illustrations les plus spectaculaires de ce phénomène est l'empreinte génomique parentale : deux allèles d'un même gène, qui baignent dans le même noyau, se comportent différemment selon leur origine parentale, alors que leurs séquences nucléotidiques peuvent être rigoureusement identiques.

Une autre croyance est en train de s'écrouler : celle selon laquelle les gènes codent des protéines (à l'exception de quelques gènes, dont ceux des ARN ribosomiques et des ARN de transfert, dont la fonction est justement la synthèse des protéines) ; toutes les fonctions régulatrices chez les Eucaryotes, en particulier, semblaient réservées aux protéines. La découverte des microARN, petits ARN régulateurs impliqués dans une très grande variété de phénomènes, et chez la plupart des Eucaryotes, a mis un terme à cette injustice. Tous les ARN ne codent pas des protéines, et les ARN non-codants ne sont plus aujourd'hui d'attendrissantes bizarreries, ce sont des candidats sérieux pour de nombreuses régulations de l'expression de gènes.

Les travaux que je présente dans ce manuscrit, et que j'ai réalisés au cours de ma thèse, ont abouti à la découverte et à une première caractérisation de gènes d'ARN non-codants soumis à l'empreinte génomique parentale. Ces travaux de l'ère post-génomique, où les gènes sont découverts avant leurs fonctions, n'ont pas encore permis d'assigner un rôle certain à ces nouveaux gènes ; cependant, quelques données, présentées également dans le manuscrit, nous permettent de les imaginer.

¹Prix Nobel de chimie 1980 pour avoir contribué à l'élaboration d'un protocole de séquençage de l'ADN.

1.2 L'empreinte génomique parentale

1.2.1 Historique

La différence fonctionnelle entre les deux haplogénomes parentaux chez les Mammifères était pressentie depuis les années 1970, depuis la découverte que les parthénogénètes de Mammifères meurent avant la fin du développement embryonnaire (Kaufman *et al.*, 1977), et depuis que la génétique semblait indiquer que certains allèles s'exprimaient ou non selon leur origine parentale (Johnson, 1974 ; Lyon et Glenister, 1977). C'est en 1984 que trois équipes ont démontré que les deux contributions parentales au génome de la descendance n'étaient pas strictement équivalentes, par des expériences de transplantation de pronuclei d'embryons du stade « 1 cellule » chez la Souris : lorsqu'un pronucleus mâle est transféré à un œuf activé, une fraction notable d'embryons se développe correctement, alors qu'aucun n'y parvient si le pronucleus transféré est un pronucleus femelle (Surani *et al.*, 1984). Une autre équipe est arrivée aux mêmes conclusions, en transférant cette fois dans des œufs énucléés les deux pronuclei d'embryons parthénogénétiques (le développement est alors interrompu à mi-gestation) ou d'embryons issus d'une fécondation (les deux tiers des embryons atteignent alors la naissance) (Mann et Lovell-Badge, 1984) ; l'haplogénome paternel semblait donc requis pour l'ontogenèse. L'haplogénome maternel, quant à lui, est également nécessaire, puisque le remplacement, dans un embryon « 1 cellule », d'un pronucleus femelle par un deuxième pronucleus mâle aboutit, tout comme le remplacement d'un pronucleus mâle par un deuxième pronucleus femelle, à de graves défauts de l'embryogenèse, conduisant à un arrêt du développement et, souvent, à une résorption des embryons (McGrath et Solter, 1984). Les rares embryons qui arrivent à terme ont manifestement hérité d'un pronucleus mâle et d'un pronucleus femelle suite à une confusion lors du prélèvement des pronuclei, comme le démontre leur génotype pour des marqueurs génétiques.

Ces trois groupes ont donc été amenés à proposer que les haplogénomes de chaque gamète portent des marques spécifiques, une « empreinte », apposée, vraisemblablement, au cours de la gamétogenèse. Cette empreinte est nécessairement réversible, puisque chaque individu doit produire des gamètes dont les chromosomes portent une empreinte spécifique de son sexe, alors qu'il a lui-même hérité de chromosomes portant des marques maternelles, et de chromosomes portant des marques paternelles (par exemple, une femelle doit effacer l'empreinte de ses chromosomes paternels pour produire des ovocytes qui ne contiendront que des chromosomes à empreinte maternelle).

L'étude de souris présentant des disomies uniparentales² pour une paire de chromosomes donnée, ou pour une région chromosomique seulement, a montré que ces défauts développementaux pouvaient être attribués à des loci spécifiques (Cattanach et Kirk, 1985). Le procédé d'obtention des souris à disomies uniparentales est présenté en figure 1.7, page 41, sur l'exemple des disomies uniparentales pour la paire n°12. Il a ainsi été possible de dresser une carte des régions du génome soumises à l'empreinte génomique parentale (du moins : celles pour lesquelles une anomalie de la répartition entre copies maternelle et paternelle se traduisait par un phénotype visible) (Cattanach et Beechey, 1990).

Plusieurs techniques ont été développées pour identifier de nouveaux gènes soumis à l'empreinte génomique parentale :

- l'analyse de banques soustractives de cDNA entre embryons normaux et embryons parthénogénétiques (Kaneko-Ishino *et al.*, 1995) ;
- la recherche de méthylations de cytosine différentielles entre embryons parthénogénétiques et androgénétiques, ou entre des souris adultes issues de croisements réciproques entre des

²Pour une paire de chromosomes donnée, ces souris possèdent deux chromosomes provenant du même parent, et aucun provenant de l'autre parent.

souches présentant des polymorphismes de restriction (Hatada *et al.*, 1993 ; Plass *et al.*, 1996 ; Peters *et al.*, 1999 ; Smith *et al.*, 2003) ;

- la comparaison des ARN exprimés par des embryons normaux, des parthénogénètes, et des androgénètes, ou entre des souris adultes issues de croisements réciproques entre des souches présentant des polymorphismes de séquence, par une RT-PCR peu spécifique, le « mRNA differential display » (Kikyo *et al.*, 1997 ; Hagiwara *et al.*, 1997) ;
- l’analyse d’« hybrides monochromosomiques » Homme-Souris (des cellules murines dans lesquelles un chromosome humain, d’origine parentale définie, est transféré ; les marques épigénétiques du chromosome humain sont maintenues dans la cellule murine, et on considère que le patron de transcription du chromosome humain est préservé) (Mitsuya *et al.*, 1999) ;
- l’analyse systématique (par séquençage de cDNA d’hétérozygotes pour des sites de polymorphisme) des gènes voisins de gènes déjà connus pour être soumis à l’empreinte génomique parentale : les gènes soumis à l’empreinte sont en effet souvent regroupés, sur le génome (voir section 1.2.2) (voir par exemple Lee *et al.*, 2000 ; Kim *et al.*, 1999).

Puisque les embryons parthénogénétiques meurent en fin de gestation, la première technique (et la deuxième et la troisième, dans une de leurs versions) se limitent aux gènes exprimés, et soumis à l’empreinte, au cours du développement embryonnaire.

L’empreinte génomique parentale contribue manifestement à l’incapacité des parthénogénètes de Mammifères de se développer à terme, puisqu’ils répriment les deux allèles des gènes à expression paternelle. Au printemps de cette année, l’équipe de T. Kono a montré qu’il était peu probable qu’un phénomène additionnel empêche le développement des parthénogénètes, puisqu’il est possible d’obtenir des souris parthénogénétiques viables en normalisant l’expression de leurs gènes soumis à l’empreinte génomique parentale (Kono *et al.*, 2004). Ces souris sont issues de la fusion de deux ovocytes : le premier est un ovocyte mature, et le deuxième a été prélevé sur une souris femelle peu après la naissance : à ce stade, le génome des ovocytes n’a pas encore acquis d’empreinte maternelle, et il se comporte pratiquement comme un génome paternel (voir aussi 1.2.3, page 27). Outre cet artifice de manipulation (qui normalise partiellement l’expression des gènes soumis à l’empreinte génomique parentale, et allonge la durée de vie des embryons parthénogénétiques, sans toutefois leur permettre d’atteindre la naissance), cette équipe a dû ajouter une complication supplémentaire, pour corriger l’expression d’une paire de gènes soumis à l’empreinte, *Igf2* et *H19*, dont le taux d’expression n’était pas restauré. La délétion d’une région régulatrice de ce locus sur l’un des deux allèles force le monoallélisme de l’expression de ces deux gènes, et permet un développement à terme pour une petite proportion des embryons (chez ces embryons, l’expression de nombreux gènes est alors très similaire à leur expression dans des embryons issus de fécondations ; les gènes soumis à l’empreinte régulent donc, plus ou moins directement, de nombreux aspects du métabolisme). L’utilisation d’ovocytes prélevés à la naissance, et la délétion de la région régulatrice du locus *Igf2/H19*, n’affectent probablement que les gènes soumis à l’empreinte ; les auteurs en concluent donc que seule l’empreinte génomique parentale bloquait le développement des embryons parthénogénétiques.

Chez les Animaux, le phénomène d’empreinte génomique parentale semble restreint aux Mammifères placentaires et aux Marsupiaux (Killian *et al.*, 2000 ; Killian *et al.*, 2001 ; Nolan *et al.*, 2001). Mais il a également été observé chez des Angiospermes (Plantes à fleurs) : ainsi, dans l’endosperme de Maïs, des allèles bien précis de deux gènes ne s’expriment que s’ils sont transmis maternellement (Kermicle, 1970 ; Chaudhuri et Messing, 1994), et des allèles bien précis de deux autres gènes sont méthylés spécifiquement sur l’allèle paternel (Lund *et al.*, 1995) (on ne connaît, chez les Mammifères, aucun gène qui ne soit ainsi soumis à l’empreinte que sous certaines formes alléliques). Chez la Crucifère *Arabidopsis thaliana*, deux autres gènes sont soumis à l’empreinte : *MEDEA* (Vielle-Calzada *et al.*, 1999 ; Kinoshita *et al.*, 1999) et *FWA* (Kinoshita *et al.*, 2004),

alors que des fusions entre trois répresseurs du développement de la graine (*MEDEA* lui-même, *FIE*, qui lui est apparenté, et *FIS2*) et un gène rapporteur ne s'expriment qu'à partir de la copie maternelle, ce qui indique que l'effet maternel observé pour ces gènes est dû, au moins partiellement, à l'empreinte génomique parentale (Luo *et al.*, 2000).

Il est remarquable que tous les gènes d'Angiospermes soumis à l'empreinte génomique parentale sont à expression maternelle (lorsque l'expression a été analysée), ce qui pourrait révéler une contrainte mécanistique. À ce titre, il est intéressant de noter que l'empreinte des gènes *MEDEA* et *FWA* n'est pas due à une méthylation d'ADN *de novo* spécifique d'un allèle (comme cela semble être toujours le cas chez les Mammifères : Reik et Walter, 2001), mais à l'activation spécifique d'un allèle, impliquant une ADN glycosylase soupçonnée de catalyser une déméthylation active ou un remodelage de la chromatine (Choi *et al.*, 2002 ; Kinoshita *et al.*, 2004). L'empreinte génomique parentale des Mammifères et celle des Angiospermes, qui semblent reposer sur des mécanismes différents, ne sont donc pas forcément homologues, et pourraient résulter d'une évolution convergente, sous la pression de contraintes reproductives similaires (voir 1.2.4, p. 40).

Dans toute la suite de cette section, les données présentées concerneront donc exclusivement l'empreinte chez les Mammifères, puisque mon travail de thèse s'est limité aux Mammifères.

1.2.2 Les loci soumis à l'empreinte génomique parentale

On connaît actuellement 67 gènes autosomaux soumis à l'empreinte génomique parentale chez la Souris, auxquels il faut ajouter la plupart des gènes du chromosome X, soumis à l'empreinte dans le placenta des embryons femelles (voir ci-dessous) ; il semble que la plupart soient également soumis à l'empreinte chez les autres Mammifères (voir par exemple Feil *et al.*, 1998 ; Hagemann *et al.*, 1998 ; McLaren et Montgomery, 1999 ; O'Neill *et al.*, 2000 ; Kim *et al.*, 2004). L'expression différentielle des deux allèles de certains gènes soumis à l'empreinte peut dépendre des tissus : ainsi, le gène *Gnas* est à expression biallélique dans de nombreux tissus, et exprimé préférentiellement à partir de l'allèle maternel dans quelques autres (Williamson *et al.*, 2004) ; *Igf2*, principalement exprimé à partir de l'allèle paternel, est exprimé bialléliquement dans un type cellulaire particulier (Charalambous *et al.*, 2004) ; plusieurs gènes d'un locus du chromosome 6 murin sont soumis à l'empreinte dans l'embryon, mais exprimés à partir des deux allèles chez le nouveau-né (Ono *et al.*, 2003). La plupart des gènes soumis à l'empreinte sont regroupés, dans le génome, en domaines chromosomiques dont la taille typique est de l'ordre de 1 Mpb, et où se côtoient gènes à expression maternelle, gènes à expression paternelle, et gènes à expression biallélique : en 2001, Reik et Walter ont estimé que 80 % des gènes soumis à l'empreinte appartenaient à de tels loci (Reik et Walter, 2001). Cette organisation génomique permet probablement une co-régulation de gènes distincts, comme l'illustre la découverte, dans cinq loci, de centres d'empreinte (*imprinting centers*, IC, en anglais), séquences régulatrices qui contrôlent l'empreinte de loci entiers (voir 1.2.3, page 28). On dénombre à l'heure actuelle 18 loci murins autosomaux soumis à l'empreinte génomique parentale ; le chromosome X a un statut particulier : la plupart de ses gènes sont soumis à l'empreinte (et exprimés à partir de l'allèle maternel uniquement) dans le placenta des embryons femelles (voir page 91).

Le tableau 1.2, page 42, donne la liste des gènes et loci connus soumis à l'empreinte.

Les loci soumis à l'empreinte génomique parentale sont caractérisés par plusieurs propriétés remarquables :

- La réplication asynchrone des deux allèles : les deux allèles de certains gènes soumis à l'empreinte ne sont pas répliqués en même temps, à l'interphase (Kitsberg *et al.*, 1993 ; Simon *et al.*, 1999 ; Kagotani *et al.*, 2002) ; cependant, cette propriété n'est pas générale (Shuster *et al.*, 1998). L'asynchronie de la réplication pourrait jouer un rôle dans le contrôle différentiel de la transcription des deux allèles : un complexe répresseur de la transcrip-

tion est spécifiquement recruté sur les fourches de réplication en fin d'interphase (lorsque l'hétérochromatine est répliquée), et pas en début d'interphase (pour la réplication des séquences euchromatiniennes) : la chronologie de la réplication d'une séquence donnée pourrait donc influencer son taux de transcription (Rountree *et al.*, 2000). Mais la réplication asynchrone des deux allèles pourrait aussi simplement refléter une différence de structure chromatinienne (en particulier, elle est abolie par un traitement qui affecte des modifications post-traductionnelles des histones : Kagotani *et al.*, 2002), et serait alors une conséquence des mécanismes aboutissant à la distinction transcriptionnelle des deux allèles : elle n'aurait pas de rôle en tant que telle dans l'expression différentielle des deux allèles (notamment, l'asynchronie de la réplication n'est pas affectée chez un mutant dont l'empreinte du locus *Igf2/H19* est perdue : Cerrato *et al.*, 2003).

- Une localisation particulière des deux allèles dans le noyau : les deux allèles du locus *Igf2/H19* semblent en effet se localiser dans deux régions distinctes du noyau, puisque l'un des deux est le plus souvent proche de l'enveloppe nucléaire, et le deuxième, plus éloigné (et la nature de l'allèle le plus proche de l'enveloppe nucléaire dépend du type cellulaire) ; la localisation des deux allèles dans des régions spécialisées du noyau pourrait être à l'origine de l'asynchronie de leur réplication (Gribnau *et al.*, 2003). Cette différence de localisation subnucléaire pourrait s'expliquer par une interaction différentielle des deux allèles avec la matrice nucléaire : seul l'allèle paternel de ce locus semble pouvoir interagir avec la matrice (Weber *et al.*, 2003).
- Des taux de recombinaison différents entre les méioses des deux lignées germinales : des séquences particulières de plusieurs loci soumis à l'empreinte génomique parentale présentent des taux de recombinaison différents entre la méiose femelle et la méiose mâle, même après correction d'un léger biais observé sur le reste du génome (Paldi *et al.*, 1995 ; Robinson et Lalande, 1995 ; Lercher et Hurst, 2003). Cette différence pourrait découler d'une différence de structure chromatinienne de ces séquences dans les deux lignées germinales, c'est à dire : à la période où l'empreinte se met en place (*cf* page 26).
- Dans plusieurs loci soumis à l'empreinte, des différences de modifications chimiques des histones ont été observées entre les deux allèles (Grandjean *et al.*, 2001 ; Gregory *et al.*, 2001 ; Fournier *et al.*, 2002 ; Perk *et al.*, 2002 ; Yang *et al.*, 2003). Ces modifications des composants protéiques de la chromatine sont connues pour affecter le niveau d'expression des gènes qui les portent : alors que l'euchromatine (conformation peu condensée, riche notamment en histone H3 méthylée sur la lysine n°4 et acétylée sur la lysine n°9) favorise la transcription, l'hétérochromatine (plus compacte, riche notamment en histone H3 méthylée sur la lysine n°9) la réprime. Des combinaisons particulières des différentes modifications post-traductionnelles des histones seraient reconnues par des partenaires spécifiques, aux effets différents sur la transcription : de tels partenaires seraient alors responsables de l'interprétation de ce codage, et traduiraient l'information constituée par le patron de modifications des histones, en différences d'activité transcriptionnelle (voir Jenuwein et Allis, 2001, pour une revue).
- Des méthylations de l'ADN différentielles entre les deux allèles : voir ci-dessous.

Le nombre total de gènes soumis à l'empreinte génomique parentale a été estimé (par une évaluation du nombre de sites différentiellement méthylés dans le génome murin) à une centaine (Hayashizaki *et al.*, 1994).

1.2.3 Diversité des mécanismes

Les mécanismes qui aboutissent à l'expression différentielle des deux allèles des gènes soumis à l'empreinte ne sont pas encore compris dans leur totalité. En particulier, la plupart des caractéristiques de ces gènes (proximité d'un insulateur, transcription d'un ARN antisens,

méthylation de l'allèle inactif, ...) ne sont pas absolument générales, et souffrent des exceptions. Ainsi, l'ARN non-codant *Air*, dont des études poussées suggèrent l'implication dans l'empreinte des gènes *Igf2r*, *Slc22a2* et *Slc22a3* chez la Souris (*cf* page 32), semble incapable de remplir ce rôle dans le placenta humain, puisqu'aucun orthologue humain de *Air* n'a été décrit, alors que IGF2R peut y être soumis à l'empreinte (Oudejans *et al.*, 2001).

Finalement, la propriété la plus générale des gènes soumis à l'empreinte génomique parentale semble être la proximité de régions différentiellement méthylées entre les deux allèles. Des travaux futurs permettront peut-être de dégager d'autres caractères généralisables à tous ces gènes ; mais peut-être révéleront-ils une grande variété de mécanismes, où la régulation de chaque gène apparaîtra comme un cas particulier.

Les régions différentiellement méthylées

La méthylation en position 5 de la cytosine de l'ADN (appelée « méthylation de l'ADN » pour simplifier) semble être la marque, apposée pendant la gamétogenèse, qui distingue les chromosomes provenant de la mère de ceux provenant du père, puisque des régions différentiellement méthylées (appelées « DMR », pour *differentially methylated regions*, en anglais) ont été identifiées à proximité de la plupart des gènes soumis à l'empreinte (voir Reik et Walter, 2001, pour une revue). D'autre part, des embryons de Souris déficients pour la méthyl-transférase de maintien *Dnmt1* perdent l'empreinte de plusieurs gènes (l'expression de *H19* devient biallélique, celle de *Igf2* et *Igf2r* est réprimée sur les deux allèles : Li *et al.*, 1993 ; l'empreinte de *Ascl2/Mash2*, quant à elle, n'est que partiellement levée, ce qui suggère qu'un autre mécanisme — basé sur une autre modification épigénétique ? — renforce l'effet de la méthylation de l'ADN : Tanaka *et al.*, 1999), ce qui montre que, pour ces loci au moins, et à ce stade développemental, le maintien du patron de méthylation de l'ADN est nécessaire au maintien de l'empreinte.

La méthylation de l'ADN présente donc toutes les qualités attendues de l'empreinte responsable des différences d'expression entre allèles maternels et paternels :

1. Elle peut être appliquée différentiellement sur les deux chromosomes parentaux, comme en atteste l'existence des DMR.
2. Elle est réversible (une femelle, par exemple, peut donc effacer les marques de ses chromosomes paternels pour en faire des chromosomes maternels, transmissibles à sa propre descendance).
3. Elle peut influencer l'expression des gènes (voir page 29).

Enfin, le comportement de la méthylation de l'ADN (apparition, maintien, disparition) est synchrone avec celui de l'empreinte (*cf* ci-dessous).

Il est donc couramment admis que l'empreinte est un patron de méthylations de l'ADN (Reik et Walter, 2001), ce qui avait déjà été proposé dès la découverte du phénomène d'empreinte génomique parentale (Mann et Lovell-Badge, 1984).

Le cycle de l'empreinte

On appelle « cycle de l'empreinte » l'enchaînement des événements (effacement, mise en place, maintien) qui affectent l'empreinte, responsable des différences d'expression des deux allèles. La figure 1.1 donne la chronologie des étapes du cycle de l'empreinte.

- a) Effacement de l'empreinte des chromosomes parentaux.

Les cellules germinales primordiales, à l'origine de la lignée germinale, portent un patron de méthylation d'ADN apparemment identique à celui des cellules somatiques environnantes, et ces méthylations sont perdues à un stade très précoce du développement des gonades (entre

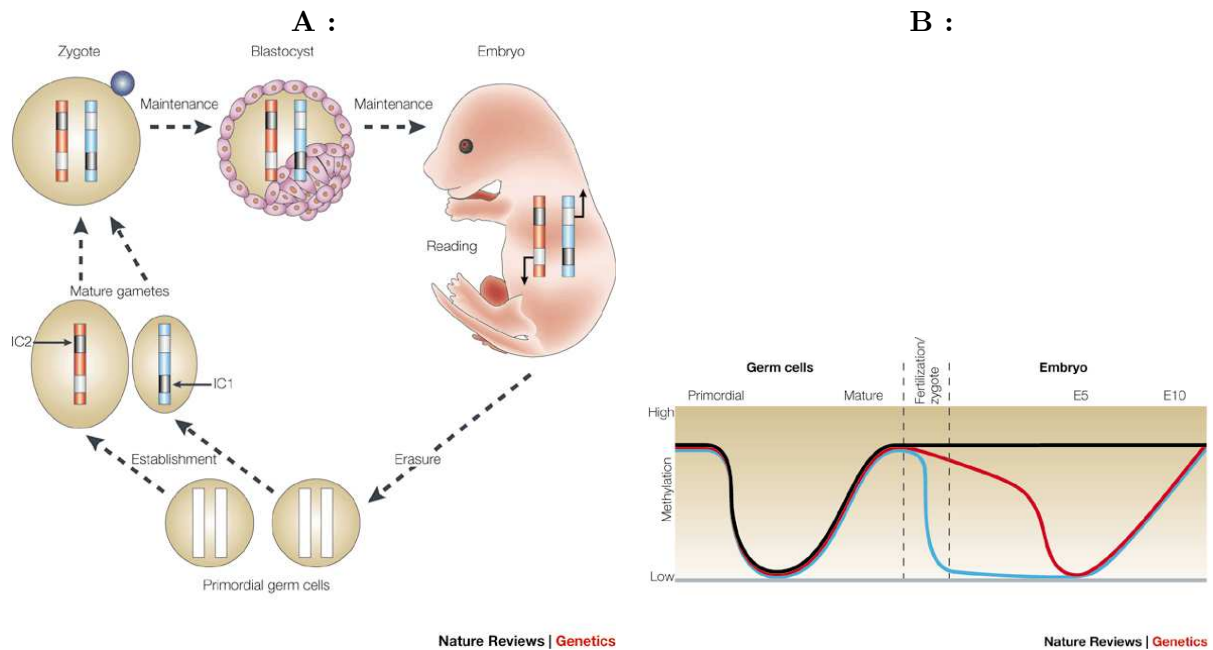


FIG. 1.1 – LE CYCLE DE L'EMPREINTE. Figures tirées de Reik et Walter, 2001. Voir texte pour le détail des étapes. **A :** les marques de méthylation différentielle de l'ADN sont effacées dans les cellules germinales primordiales (« Primordial germ cells »), dans les futures gonades de l'embryon. De nouvelles marques, spécifiques du sexe de l'embryon, sont établies au cours de la gamétogenèse : les centres d'empreinte (IC, voir page 28) acquièrent leur statut de méthylation à cette étape. Après fécondation, les chromosomes du zygote portent donc des marques différentielles, et ce patron différentiel est maintenu au cours du développement. **B :** l'ordonnée représente le niveau de méthylation, et l'abscisse représente le déroulement chronologique, de la maturation des cellules germinales primordiales, au développement embryonnaire (E5, E10 : après 5 et 10 jours de développement embryonnaire, respectivement). Courbe grise : allèle non méthylé d'un gène soumis à l'empreinte ; courbe noire : allèle méthylé d'un gène soumis à l'empreinte ; courbe rouge : gène méthylé non soumis à l'empreinte, transmis maternellement ; courbe bleue : gène méthylé non soumis à l'empreinte, transmis paternellement.

$t = 10,5 j$ et $t = 11,5 j$, ou entre $t = 11,5 j$ et $t = 12,5 j$ du développement embryonnaire, selon les auteurs) (Lee *et al.*, 2002a ; Hajkova *et al.*, 2002). Cette déméthylation est si rapide (elle se déroule sur une période de 24 h au maximum, alors que les cellules germinales primordiales se divisent en 16 h) qu'elle est certainement due à un processus actif, plus qu'à une « déméthylation passive » (dilution des brins d'ADN méthylés dans des brins néo-synthétisés non méthylés, au fil des duplications). Cette déméthylation, qui semble affecter tous les loci (soumis à l'empreinte, et non soumis à l'empreinte) à l'exception des séquences répétées, fait de chaque chromosome un substrat « naïf », apte à recevoir une nouvelle empreinte. En particulier, l'empreinte des chromosomes hérités du parent du même sexe est effacée, pour être re-appliquée ensuite (*cf* ci-dessous), alors qu'il semblait possible que, par exemple, l'empreinte des chromosomes maternels des futurs ovocytes ne soit pas effacée.

b) Mise en place de l'empreinte.

Après l'effacement de l'empreinte, les cellules germinales primordiales migrent vers les crêtes génitales, et se multiplient par mitose. Les cellules de la lignée germinale mâle (appelées, à ce stade, « spermatogonies ») entrent alors en phase G0. Les cellules de la lignée germinale femelle

(appelées, à ce stade, « ovogonies »), quant à elles, entament la première division de méiose, et s'arrêtent en prophase ; ces cellules sortent elles aussi du cycle cellulaire, et subissent la « croissance ovocytaire », phase longue de plusieurs mois ou années, qui s'achève par la maturation de l'ovocyte, et l'ovulation. C'est au cours de ces phases G0 que les chromosomes des cellules des deux lignées germinales acquièrent les empreintes paternelle et maternelle.

Chez la Souris, les gènes *Peg1/Mest*, *Peg3*, *Snrpn*, *Igf2r* et *p57^{KIP2}/Cdkn1c* acquièrent en effet une empreinte maternelle au cours de la croissance de l'ovocyte, puisque le transfert de noyaux d'ovocytes prélevés avant la croissance dans des ovocytes ayant terminé la croissance permet le développement d'embryons parthénogénétiques qui expriment ces gènes à des taux comparables à ceux des embryons biparentaux (Obata *et al.*, 1998) ; ces embryons parthénogénétiques se développent d'ailleurs jusqu'à un stade plus avancé que les embryons parthénogénétiques classiques, issus du transfert d'un noyau d'ovocyte prélevé après la croissance, dans un autre ovocyte prélevé après la croissance (Kono *et al.*, 1996). La correction de l'expression de ces gènes (les trois premiers sont habituellement à expression paternelle, et les deux derniers, à expression maternelle) semble due à la contribution de l'ovocyte prélevé avant croissance qui, pour ces gènes, se comporte comme une contribution paternelle : *Peg3* est en effet exprimé uniquement à partir de l'allèle *ng* (pour : *non-growing oocyte*, ovocyte prélevé avant la croissance), alors que *Igf2r* et *p57^{KIP2}/Cdkn1c* sont exprimés uniquement à partir de l'allèle *fg* (pour : *fully grown oocyte*, ovocyte prélevé après la croissance), et on estime que *Peg1/Mest* et *Snrpn* sont exprimés à partir de l'allèle *ng* uniquement, puisque ces gènes ne sont pas exprimés dans des embryons parthénogénétiques à deux allèles *fg* (Obata *et al.*, 1998). Ainsi, ces cinq gènes acquièrent l'empreinte maternelle au cours de la croissance de l'ovocyte.

De manière comparable, une autre équipe a transféré des noyaux de cellules germinales primordiales mâles (prélevées peu après la déméthylation présentée en a) ci-dessus) dans des ovocytes, énucléés cette fois (Kato *et al.*, 1999). L'analyse de l'expression de plusieurs gènes soumis à l'empreinte, dans des embryons issus de cette manipulation, montre là aussi que l'empreinte n'a pas encore été appliquée sur le génome de ces cellules germinales primordiales. En revanche, des noyaux de spermatocytes primaires (des cellules pré-méiotiques, mais plus tardives), introduits dans des ovocytes permettent le développement de souriceaux viables, ce qui indique que l'empreinte paternelle est établie sur les chromosomes des spermatocytes primaires (Ogura *et al.*, 1998 ; Kimura *et al.*, 1998). L'acquisition de la méthylation spécifiquement paternelle de cette région est approximativement synchrone à l'établissement de l'empreinte (Ueda *et al.*, 2000 ; Davis *et al.*, 2000).

L'établissement de l'empreinte, au moins dans la lignée germinale femelle, implique Dnmt3l, une protéine apparentée aux méthylases *de novo* Dnmt3a et Dnmt3b, mais dépourvue de site catalytique : la progéniture de femelles homozygotes mutantes pour *Dnmt3l* présente des défauts de méthylation et d'empreinte de la plupart des gènes réprimés maternellement, à l'exception notable de *H19*, dont l'empreinte semble justement être une méthylation paternelle. Quant aux mâles homozygotes mutants pour *Dnmt3l*, ils ne produisent pas de spermatozoïde : le rôle de Dnmt3l dans l'établissement de l'empreinte dans la lignée germinale mâle n'a donc pas pu être analysé (Bourc'his *et al.*, 2001). Dnmt3l, qui semble dépourvue d'activité méthyl-transférase (Bourc'his *et al.*, 2001 ; Hata *et al.*, 2002), serait donc requise pour la méthylation *de novo* des séquences méthylées maternellement, catalysée par une autre protéine. La délétion conditionnelle de *Dnmt3a* et *Dnmt3b* chez la Souris a montré que la méthyl-transférase est Dnmt3a, autant dans la lignée germinale mâle que dans la lignée germinale femelle (Kaneda *et al.*, 2004), ce que corroborent des résultats indépendants (Hata *et al.*, 2002 ; Chédin *et al.*, 2002).

Si Dnmt3a, assistée de Dnmt3l, est responsable de la méthylation des séquences méthylées maternellement ou paternellement, il faut encore expliquer sa spécificité, puisque chacune de ces deux protéines est exprimée dans chacune des deux lignées germinales : comment se fait-il, donc,

que les séquences méthylées maternellement ne soient pas méthylées dans les spermatogonies, et que les séquences méthylées paternellement échappent à la méthylation dans les ovogonies ? Il est possible que des protéines additionnelles, spécifiques de l'une ou l'autre des lignées germinales, guident le complexe Dnmt3a/Dnmt3l vers ses séquences-cibles (Chédin *et al.*, 2002). Ou encore, des protéines, spécifiques de l'ovogenèse, et spécifiques de la spermatogenèse, pourraient protéger sélectivement certaines séquences contre la méthylation, dans chacune des deux lignées germinales. Il a été proposé que la protéine à doigts de zinc CTCF pouvait jouer ce rôle de protection de la DMR dérivée de la lignée germinale du locus *Igf2/H19* pendant l'ovogenèse, mais des mutations de ses sites de fixation n'ont aucun effet sur l'établissement de l'empreinte maternelle de cette séquence (Schoenherr *et al.*, 2003 ; Szabó *et al.*, 2004b)³. En revanche, les testicules de souriceaux mâles nouveaux-nés (contenant les gonocytes, où l'empreinte paternelle se met en place) expriment une protéine qui pourrait aussi se fixer spécifiquement sur la même DMR du locus *Igf2/H19* (Bowman *et al.*, 2003) ; cette protéine pourrait être un ou des récepteur(s) à hormones nucléaire(s) (RXR α ou ER β), dont des sites de fixation ont également été identifiés dans la DMR, et qui semble(nt) s'y fixer spécifiquement dans la lignée germinale mâle ; dans ce cas, la fixation de ce(s) récepteur(s) favoriserait la méthylation de la DMR (Szabó *et al.*, 2004a).

Une séquence peut également être différenciellement méthylée sous le contrôle d'une autre DMR du même locus : l'état de méthylation de cette autre DMR va déterminer la méthylation, ou non, de la DMR contrôlée. Cette méthylation secondaire pourrait se dérouler pendant la gamétogenèse, ou après la fécondation (voir Reik et Walter, 2001, pour une revue). Bien sûr, cette possibilité ne fait que repousser le problème de la spécificité parentale de la méthylation, puisqu'il faut encore expliquer la méthylation différentielle de la DMR de contrôle (qui, quant à elle, peut mettre en jeu l'un des mécanismes discutés dans le paragraphe ci-dessus). De telles séquences, qui peuvent contrôler l'épigénotype différentiel de très grandes régions (de l'ordre de 1 Mpb) sont appelées « centres d'empreinte » (ou : IC, pour : *imprinting centers*, en anglais). On a identifié, pour l'heure, des IC dans cinq des 18 loci murins soumis à l'empreinte génomique parentale (pour des revues, voir Spahn et Barlow, 2003 ; Ferguson-Smith *et al.*, 2004) ; lorsque de telles séquences sont délétées, le patron de méthylation et d'expression de leur locus entier est perturbé. Le mécanisme de la méthylation différentielle d'une DMR (alors appelée « DMR secondaire ») sous le contrôle de la méthylation d'une autre DMR, n'est pas encore connu. Il pourrait impliquer le recrutement d'une méthyl-transférase par l'un des deux allèles de la DMR de contrôle, en fonction de son état de méthylation ; il pourrait aussi passer par l'intermédiaire d'un ARN non-codant, dont la transcription dépend de la méthylation du promoteur, et qui pourrait provoquer la méthylation de séquences adjacentes, sur l'allèle à partir duquel il est transcrit (l'interaction entre les deux DMR du locus *Igf2r/Air* pourrait s'expliquer de la sorte : voir page 32 pour la description des données disponibles sur le mécanisme de l'empreinte de ce locus ; voir la section 1.5, page 91, pour l'implication des ARN non-codants dans les modifications épigénétiques).

Enfin, il reste aussi à expliquer la spécificité de site de cette méthylation *de novo*, qui prend place pendant les gamétogenèses : pourquoi seules les DMR sont-elles méthylées, et pas tous les dinucléotides CpG du génome (voir par exemple les données de Monk *et al.*, 1987) ? Une première possibilité implique des facteurs en *trans*, qui pourraient cibler la méthyl-transférase spécifiquement sur les DMR (et ces facteurs pourraient être spécifiques de l'ovogenèse, ou de la

³Un autre groupe a également étudié le rôle de CTCF vis à vis de l'empreinte maternelle de ce locus, en réprimant CTCF par RNAi (Fedoriw *et al.*, 2004). Leurs travaux montrent que l'expression de siRNA dirigés contre l'ARN_m CTCF pendant l'ovogenèse provoque, de façon assez hétérogène, la méthylation de la séquence de la DMR dans les ovocytes (cet effet est attribué au défaut de CTCF qui, dans des conditions normales, protégerait la DMR contre la méthylation, dans les ovocytes). Ces résultats paraissent difficilement réconciliables avec ceux de Schoenherr *et al.*, 2003 et Szabó *et al.*, 2004b, et peuvent être dus à un effet non spécifique des siRNA, ou à un effet indirect de CTCF (par « indirect », lire : indépendant de sa capacité à se fixer sur les sites identifiés dans la DMR de *H19*) (Szabó *et al.*, 2004b).

spermatogénèse, ce qui expliquerait aussi la spécificité parentale de la méthylation : *cf* paragraphe ci-dessus). Une deuxième possibilité ferait intervenir des séquences en *cis*, du voisinage de la DMR (Perk *et al.*, 2002 ; Yoon *et al.*, 2002 ; Herman *et al.*, 2003) ; notamment, de nombreuses DMR jouxtent des séquences répétées, qui pourraient attirer des méthyl-transférases. Une troisième possibilité repose sur l'intervention de la machinerie du RNAi, qui pourrait, comme dans d'autres espèces, nucléer l'hétérochromatinisation à partir de sites qui exprimeraient des ARN maturés en petits ARN d'une vingtaine de nucléotides (voir Delaval et Feil, 2004, pour une revue) (voir aussi section 1.4.5, page 80, et discussion page 154).

c) Maintien de l'empreinte.

Après l'implantation de l'embryon, puis tout au long de la vie de l'organisme, les méthylations des DMR sont vraisemblablement maintenues, comme toutes les autres méthylations de l'ADN, par une méthyl-transférase de maintien, qui méthyle les chromatides hémiméthylées après la réplication. En revanche, les DMR ont ceci de particulier qu'elles restent différenciellement méthylées malgré la vague de déméthylation globale, puis la vague de méthylation *de novo*, qui se produisent respectivement après la fécondation, et après l'implantation (Monk *et al.*, 1987).

La résistance des DMR à ces phénomènes globaux n'est pas encore comprise. Notamment, la méthyl-transférase de maintien semble exclue du noyau jusqu'au stade « 8 cellules » dans les embryons de Souris (Carlson *et al.*, 1992) ; on ne sait pas encore pourquoi les DMR échappent, alors, à une déméthylation passive (Reik et Walter, 2001).

d) Lecture de l'empreinte.

On appelle « lecture de l'empreinte » le phénomène qui interpète les différences de méthylation entre allèles, et les traduit en différences d'expression des allèles. On a identifié quatre principaux mécanismes de lecture de l'empreinte :

- la répression transcriptionnelle, par méthylation d'un promoteur
- l'activation transcriptionnelle, par méthylation d'éléments répresseurs
- l'insulation par des éléments-frontières sensibles à la méthylation
- la répression par un ARN non-codant.

Les trois premiers points impliquent des méthylations différentielles sur des séquences régulatrices. Ces séquences peuvent être méthylées différenciellement dans les lignées germinales mâle et femelle (voir b) Mise en place de l'empreinte, ci-dessus) et sont alors appelées « DMR dérivées de la lignée germinale » (Stoger *et al.*, 1993 ; Warnecke *et al.*, 1998 ; Takada *et al.*, 2002). Elles peuvent également être méthylées différenciellement après la fécondation (et sont appelées : « DMR secondaires »), sous le contrôle d'un IC (Brandeis *et al.*, 1993 ; Ferguson-Smith *et al.*, 1993 ; Feil *et al.*, 1994 ; Bielinska *et al.*, 2000).

La répression transcriptionnelle par méthylation du promoteur (qui met à profit la richesse en dinucléotides CpG de nombreux gènes soumis à l'empreinte génomique parentale) s'accompagne d'autres modifications épigénétiques (Bartolomei *et al.*, 1993 ; Ferguson-Smith *et al.*, 1993 ; Feil *et al.*, 1997). La répression transcriptionnelle est en effet souvent associée à la méthylation des cytosines du promoteur (voir Jost et Bruhat, 1997 pour une revue) et, dans plusieurs cas, la méthylation de l'ADN semble être une cause de la répression (voir par exemple Cameron *et al.*, 1999 ; Di Croce *et al.*, 2002 ; Fahrner *et al.*, 2002) : des complexes répresseurs de la transcription (contenant en particulier des déacétylases d'histones) peuvent être recrutés par les séquences d'ADN méthylées (voir par exemple Nan *et al.*, 1998 ; Ng *et al.*, 1999 ; Wade *et al.*, 1999 ; Fuks *et al.*, 2003b), ou plus directement, par la méthyl-transférase à ADN elle-même (voir par exemple Fuks *et al.*, 2000 ; Robertson *et al.*, 2000 ; Fuks *et al.*, 2003a).

L'activation transcriptionnelle par méthylation d'éléments répresseurs associe quant à elle la méthylation d'un allèle et son activité. Un tel répresseur contrôle l'empreinte génomique parentale de *Igf2* (Constância *et al.*, 2000 ; Ainscough *et al.*, 2000 ; Drewell *et al.*, 2000).

L'insulation par un élément-frontière sensible à la méthylation est responsable de l'empreinte génomique parentale⁴ du gène *Igf2* : déméthylée (sur le chromosome maternel), cette séquence est reconnue par la protéine CTCF, et la fixation de cette protéine isole le promoteur de *Igf2* d'éléments activateurs situés en aval : *Igf2* est alors réprimé. Sur le chromosome paternel, l'insulateur méthylé ne fixe pas CTCF, ce qui permet l'activation du promoteur de *Igf2* par les éléments activateurs (Bell et Felsenfeld, 2000 ; Hark *et al.*, 2000 ; Kanduri *et al.*, 2000 ; Szabó *et al.*, 2000). D'autres gènes pourraient connaître le même contrôle, par des insulateurs sensibles à la méthylation (Hikichi *et al.*, 2003 ; Kim *et al.*, 2003).

Enfin, la régulation de plusieurs gènes soumis à l'empreinte génomique parentale pourrait faire intervenir un ARN non-codant, éventuellement transcrit en orientation antisens. Plusieurs loci soumis à l'empreinte contiennent en effet des gènes non-codants ; cette particularité est parfois présentée comme une spécificité (Reik et Walter, 2001 ; Sleutels et Barlow, 2002), mais le nombre de gènes non-codants soumis à l'empreinte ne reflète peut-être que la grande attention qui est portée sur ces loci, puisque le reste des génomes semble également très riches en séquences transcrites non-codantes, notamment dans l'orientation antisens d'autres séquences transcrites (Numata *et al.*, 2003 ; Yelin *et al.*, 2003 ; Cawley *et al.*, 2004). Plusieurs gènes d'ARN non-codants et des gènes en *cis* sont soumis à une empreinte réciproque : ils sont rassemblés dans le tableau 1.1, page 31. Dans quelques cas bien étudiés (voir ci-dessous, « Régulation de l'empreinte génomique parentale sur quelques loci »), des mutations qui aboutissent à la répression de gènes non-codants s'accompagnent aussi de l'expression biallélique de gènes à expression réciproque du même locus. Plusieurs hypothèses ont donc été proposées, selon lesquelles des ARN non-codants pourraient réprimer en *cis* d'autres gènes du locus, restreignant ainsi leur expression à l'autre allèle ; les ARN non-codants seraient donc responsables de l'empreinte de ces gènes. Les mécanismes proposés sont variés :

- la compétition transcriptionnelle : le promoteur du gène non-codant et celui des gènes à expression réciproque sont en compétition pour des facteurs de transcription diffusibles ; le promoteur du gène non-codant, sur son allèle actif, titrerait ces facteurs, réduisant les gènes du voisinage au silence sur cet allèle (Bartolomei *et al.*, 1993 ; Barlow, 1997 ; Wutz *et al.*, 1997 ; Wroe *et al.*, 2000) ;
- la répression post-transcriptionnelle en *cis* : dans les couples de gènes à empreinte réciproque, et transcrits en orientation antisens, l'hybridation de l'ARN non-codant sur son ARN complémentaire transcrit à partir du même chromosome pourrait provoquer la dégradation de cet ARN complémentaire, localement ; l'ARN du gène antisens du gène non-codant serait donc transcrit à partir des deux allèles, mais l'ARN d'un seul allèle s'accumulerait (Sleutels *et al.*, 2002) ;
- la répression transcriptionnelle par l'ARN non-codant : l'ARN non-codant, sous une forme non diffusible, pourrait réprimer la transcription de gènes voisins (par exemple, en association avec des protéines de remodelage de la chromatine, en provoquant l'hétérochromatinisation de séquences-cibles ; un ARN non-codant, Xist, possède cette propriété : voir page 91) (Sleutels *et al.*, 2003).

L'implication des ARN non-codants dans le mécanisme de l'empreinte génomique parentale n'est toutefois pas une généralité ; ainsi, la substitution du gène *H19* par un gène quelconque n'a

⁴Ce phénomène prend place dans les tissus dérivés de l'endoderme, alors que le phénomène exposé précédemment (activation par la méthylation de répresseurs) prend place dans des tissus dérivés du mésoderme. Ainsi, selon les tissus, l'empreinte du gène *Igf2* n'est pas due aux mêmes mécanismes.

Gène non-codant	Expr.	Gènes à empreinte récip.	Expr.	Sens/antisens	Autres
<i>Air</i>	pat.	<i>Igf2r, Slc22a2, Slc22a3</i>	mat.	Air et Igf2r	non
<i>Kcnq1ot1/Lit1</i>	pat.	<i>Tssc3, Slc22a1l, Cdkn1c/p57^{KIP2}, Kcnq1/K_VLqt1, Tssc4, Ascl2/Mash2</i>	mat.	Kcnq1ot1/Lit1 et Kcnq1/K _V Lqt1	non
<i>Ube3a-as</i>	pat.	<i>Ube3a, Atp10c</i>	mat.	Ube3a-as et Ube3a	oui
<i>H19</i> ^a	mat.	<i>Ins2, Igf2</i>	pat.	∅	non
<i>Gtl2</i> ^b	mat.	<i>Dlk1/Pref-1, Dio3</i> ^c	pat.	∅	oui
<i>A19</i>	pat.			∅	oui
<i>Nespas</i>	pat.	<i>Gnas, Nesp</i>	mat.	Nespas et Nesp	oui
<i>Copg2as</i>	pat.	<i>Copg2</i>	mat.	Copg2as et Copg2	oui
<i>Zim3</i>	mat.	<i>Peg3/Pw1, Ocat/Us29, Zfp264</i>	pat.	Zim3 et Ocat/Us29	oui
<i>Xist</i> ^d	pat.	<i>Tsix</i>	mat.	Xist et Tsix	non

TAB. 1.1 – GÈNES D'ARN NON-CODANTS ET GÈNES À EMPREINTE RÉCIPROQUE CHEZ LA SOURIS. Ce tableau récapitule tous les couples de gènes à empreinte réciproque dont un des gènes est non-codant. Colonnes « Gène non-codant » et « Expr. » : nom du gène non-codant, et allèle exprimé ; « Gènes à empreinte récip. » et « Expr. » : gènes du même locus à empreinte réciproque avec le gène non-codant, et allèle exprimé ; « Sens/antisens » : couples d'ARN chevauchants en orientation antisens (∅ désigne les loci sans couple d'ARN sens/antisens à empreinte réciproque identifié) ; « Autres » : présence dans le locus d'autres gènes soumis à l'empreinte, de même empreinte que le gène non-codant (pour le détail, voir tableau 1.2, page 42). ^a : Un autre ARN non-codant est exprimé à partir de ce locus, l'ARN Igf2as, exprimé à partir de l'allèle paternel uniquement ; c'est un ARN antisens à *Igf2* (Moore *et al.*, 1997). ^b : D'autres gènes non-codants ont été découverts dans ce locus pendant ma thèse : voir Résultats, pages 102 et 130. ^c : Un autre gène à expression paternelle a été découvert dans ce locus pendant ma thèse : voir Résultats, page 118. ^d : les ARN Xist et Tsix sont tous les deux non-codants, j'ai choisi arbitrairement de citer *Xist* dans la colonne « Gène non-codant » et *Tsix* dans la colonne « Gènes à empreinte récip. » ; ces deux gènes sont soumis à l'empreinte génomique parentale dans les embryons femelles de Souris avant l'implantation, et dans les tissus extraembryonnaires des embryons femelles de Souris (voir page 91).

aucune incidence sur l'empreinte du gène voisin *Igf2* : l'ARN H19 ne semble donc pas participer au contrôle de *Igf2* (Jones *et al.*, 1998).

Outre ces quatre principaux mécanismes de lecture de l'empreinte, un cinquième mécanisme pourrait être à l'œuvre dans le locus *Igf2/H19*. Un processus post-transcriptionnel spécifique de l'allèle paternel semble en effet contribuer à l'empreinte génomique parentale de *H19* dans le foie embryonnaire de Souris : dans ce système, la transcription de l'allèle paternel de *H19* n'est pas complètement réprimée, puisque l'ARN de cet allèle est détecté dans le noyau ; en revanche, il est presque indétectable dans un échantillon d'ARN total, ce qui indique que la fraction cytoplasmique est très appauvrie en ARN H19 de l'allèle paternel (Jouvenot *et al.*, 1999). Les auteurs ont proposé que l'épigénotype de l'allèle paternel pouvait, d'une certaine manière, se transmettre à l'ARN transcrit à partir de cet allèle (par exemple, la chromatine de l'allèle paternel pourrait recruter des protéines qui s'associeraient spécifiquement à l'ARN transcrit sur place, et ces protéines pourraient déstabiliser l'ARN H19).

Régulation de l’empreinte génomique parentale sur quelques loci

Cette partie résume les connaissances actuelles des mécanismes moléculaires qui contrôlent l’expression des gènes des cinq loci soumis à l’empreinte génomique parentale les mieux caractérisés.

a) Le locus *Igf2r/Air*.

Une carte de ce locus est présentée en figure 1.2.

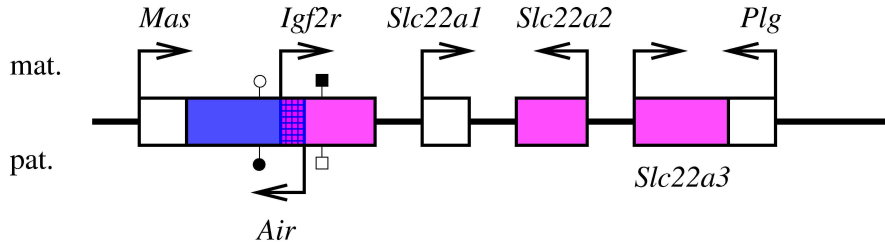


FIG. 1.2 – LE LOCUS *Igf2r/Air*. Les gènes à expression maternelle sont représentés en rose, le gène à expression paternelle est représenté en bleu, et les gènes à expression biallélique, en blanc. L’intersection entre le gène à expression paternelle et un gène à expression maternelle est quadrillée en bleu et rose. Les flèches indiquent les sens de transcription. Les carrés noir et blanc représentent la DMR dérivée de la lignée germinale (noir : hyperméthylé ; blanc : hypométhylé) ; les disques noir et blanc représentent la DMR secondaire (noir : hyperméthylé ; blanc : hypométhylé). Les disques et carrés dessinés au-dessus de la ligne horizontale représentent l’état de méthylation sur l’allèle maternel, et les disques et carrés dessinés en-dessous de la ligne horizontale représentent l’état de méthylation sur l’allèle paternel. Les promoteurs des gènes *Mas* et *Plg* sont éloignés d’environ 580 kb chez la Souris (le schéma n’est pas à l’échelle).

Ce locus contient trois gènes exprimés uniquement à partir de l’allèle maternel : *Igf2r* code le récepteur à IGF2 (*insulin-like growth factor 2*) et à mannose-6-phosphate ; *Slc22a2* et *Slc22a3* codent deux transporteurs de cations transmembranaires (le gène *Slc22a1*, qui n’est pas soumis à l’empreinte, code un autre transporteur de cations). Le gène non-codant *Air* (pour : *antisense Igf2r RNA*) est exprimé à partir de l’allèle paternel uniquement, chez la Souris. Deux régions différenciellement méthylées (DMR) ont été identifiées dans le locus. La première est une DMR dérivée de la lignée germinale : elle est méthylée dans l’ovocyte, et non méthylée dans le spermatozoïde (Stoger *et al.*, 1993). Elle est située dans le deuxième intron de *Igf2r*, et coïncide avec le promoteur de *Air*. La deuxième DMR (qui acquiert son patron de méthylation — elle est méthylée sur l’allèle paternel — pendant le développement embryonnaire : Stoger *et al.*, 1993) est le promoteur de *Igf2r*. Dans une souris mutante pour la méthyl-transférase de maintien (le promoteur de *Air* est alors hypométhylé sur les deux allèles), le gène *Igf2r* est réprimé bialléliquement (Li *et al.*, 1993). Réciproquement, des délétions du promoteur de *Air* sur le chromosome paternel, qui abolissent l’expression de l’ARN non-codant, dérèpriment les allèles paternels des gènes *Igf2r*, *Slc22a2* et *Slc22a3*, alors qu’elles n’ont pas d’effet sur l’expression des gènes du locus si elles sont portées par le chromosome maternel (Wutz *et al.*, 1997 ; Zwart *et al.*, 2001). La DMR du deuxième intron de *Igf2r* est donc un centre d’empreinte (voir p. 28).

Il semble donc que l’activité du gène *Air*, qui corrèlerait avec l’hypométhylation de son promoteur, serait responsable de la répression des allèles paternels de *Igf2r*, *Slc22a2* et *Slc22a3*. Effectivement, si l’on supprime l’expression de l’ARN *Air* sans affecter la DMR de son promoteur (en insérant un site de polyadénylation en aval de la DMR ; l’ARN *Air* est alors tronqué ≈ 4 kb en aval du site d’initiation de la transcription, alors que *Air* sauvage est long de 108 kb), les

allèles paternels des gènes *Igf2r*, *Slc22a2* et *Slc22a3* sont dérèprimés (Sleutels *et al.*, 2002). Cette dérèpression s'accompagne de la perte de méthylation de la DMR du promoteur de *Igf2*.

En outre, lorsque le promoteur de *Igf2r* est remplacé par un promoteur quelconque, pas nécessairement riche en dinucléotides CpG, l'empreinte des gènes du locus n'est pas affectée (Sleutels *et al.*, 2003). Elle ne l'est pas non plus si le promoteur de *Igf2r* est tout simplement délété : alors que *Igf2r* n'est évidemment plus exprimé, les gènes *Slc22a2* et *Slc22a3* restent à expression maternelle, et *Air* reste à expression paternelle (Sleutels *et al.*, 2003). La répression des gènes à expression maternelle du locus par *Air* ne nécessite donc pas la transcription simultanée de deux ARN antisens, *Igf2r* et *Air*.

Alors que l'origine de l'empreinte de *Air* peut s'interpréter simplement, par la méthylation différentielle de son promoteur, héritée de la lignée germinale, le mécanisme de la répression des allèles paternels de *Igf2r*, *Slc22a2* et *Slc22a3* par l'activité du gène *Air* reste mystérieuse. L'ARN *Air* semble jouer un rôle dans cette répression, comme le suggère l'expérience de sa troncation par un site de polyadénylation (Sleutels *et al.*, 2002) ; cependant, le site de polyadénylation a aussi un effet sur la transcription du gène *Air* : la transcription s'interrompt peu après la polymérisation du signal de polyadénylation (Zhao *et al.*, 1999), donc l'insertion du site de polyadénylation, outre son effet sur la nature de l'ARN produit, modifie également la fréquence du passage des polymérases à ARN sur le segment de chromatine situé en aval de l'insertion du site de polyadénylation (donc notamment, dans le gène *Igf2r*), qui pourrait avoir une influence sur la conformation de la chromatine de *Igf2r*.

Sleutels et ses collaborateurs font remarquer que la répartition des gènes dans le locus peut influencer le nombre de tissus dans lesquels ils sont soumis à l'empreinte génomique parentale (Sleutels *et al.*, 2003) : le promoteur de *Igf2r* est situé à seulement 29 kb du promoteur de *Air*, et *Igf2r* est soumis à l'empreinte à la fois dans l'embryon et chez l'adulte ; le promoteur de *Slc22a2* est situé à 170 kb du promoteur de *Air*, et son gène est soumis à l'empreinte dans le placenta et le rein adulte ; le promoteur de *Slc22a3* est situé à 215 kb du promoteur de *Air*, et son gène n'est soumis à l'empreinte qu'en début de gestation. Si l'ARN *Air* est responsable de la répression monoallélique des gènes voisins, son action s'estompe donc peut-être avec la distance ; cependant, deux gènes proches du promoteur de *Air* ne sont pas soumis à l'empreinte génomique parentale (*Mas* et *Slc22a1*, voir fig. 1.2).

Il faut noter toutefois que l'empreinte de *Igf2r* ne corrèle pas systématiquement avec l'expression de l'ARN *Air*, puisque dans le cerveau de Souris (qui exprime l'ARN *Air*), *Igf2r* est exprimé à partir des deux allèles, et dans certains tissus chez l'Homme, *Igf2r* est soumis à l'empreinte en l'absence de l'ARN *Air* (Vu *et al.*, 2004).

b) Le locus *Kcnq1*.

Une carte de ce locus est présentée en figure 1.3.

Ce locus constitue, avec le locus *Igf2/H19* (présenté ci-dessous), le domaine du syndrome de Beckwith-Wiedemann (abrégé : BWS, en anglais), qui doit son nom à un syndrome humain causé par des dérégulations de l'expression de gènes du domaine.

Le gène non-codant *Kcnq1ot1* (aussi appelé : *Lit1*) est transcrit à partir d'un promoteur situé dans le gène *Kcnq1* (qui est aussi appelé *K_VLqt1*), dans un îlot CpG qui se trouve être différentiellement méthylé (il est hyperméthylé sur l'allèle maternel, et hypométhylé sur l'allèle paternel). Cette DMR est dérivée de la lignée germinale. Lorsqu'elle est délétée sur le chromosome maternel, l'empreinte des gènes du locus n'est pas affectée ; en revanche, si elle est délétée sur le chromosome paternel, l'expression de *Kcnq1ot1/Lit1* est abolie, et les allèles paternels des gènes *Tssc3* (également appelé : *Ipl* et *Phlda2*), *Slc22a1l*, *Cdkn1c* (également appelé : *p57^{KIP2}*), *Kcnq1*, *Tssc4* et *Ascl2* (également appelé : *Mash2*) sont dérèprimés (Horike *et al.*, 2000 ; Fitzpatrick *et al.*, 2002). Des translocations humaines, à l'origine du syndrome de Beckwith-Wiedemann, provoquent l'hypométhylation de l'allèle maternel de la DMR si elles

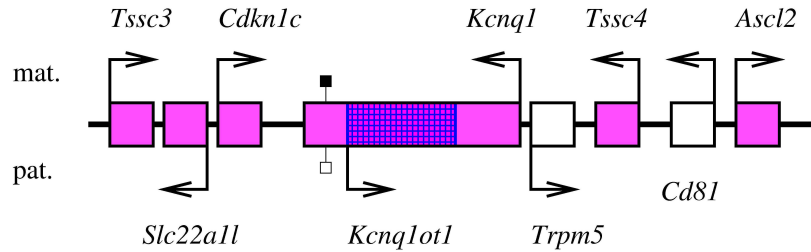


FIG. 1.3 – LE LOCUS *Kcnq1*. Les gènes à expression maternelle sont représentés en rose, les gènes à expression biallélique, en blanc, et l'intersection entre le gène à expression paternelle et un gène à expression maternelle est quadrillée en bleu et rose. Les flèches indiquent les sens de transcription. Les carrés noir et blanc représentent la DMR dérivée de la lignée germinale (noir : hyperméthylé ; blanc : hypométhylé). Le carré dessiné au-dessus de la ligne horizontale représente l'état de méthylation sur l'allèle maternel, et le carré dessiné en-dessous de la ligne horizontale représente l'état de méthylation sur l'allèle paternel. Les promoteurs de *Tssc3* et *Ascl2* sont éloignés d'environ 530 kb chez la Souris (le schéma n'est pas à l'échelle).

sont héritées maternellement, et cette déméthylation s'accompagne d'une dérégulation de l'allèle maternel de *Kcnq1ot1/Lit1*, et de la répression biallélique des six gènes à expression maternelle (Lee *et al.*, 1999c) ; cette DMR est donc un centre d'empreinte pour le locus.

La régulation de l'empreinte de ce locus partage donc de nombreux points communs avec celle du locus *Igf2r/Air* : un ARN non-codant, à expression paternelle, semble réprimer *in cis*, sur le chromosome paternel, plusieurs gènes du voisinage, leur imposant leur statut de gènes à expression maternelle.

c) Le locus *Igf2/H19*.

Ce locus, probablement le plus étudié des loci soumis à l'empreinte génomique parentale, est situé dans le domaine distal 7, à une centaine de kilobases du locus *Kcnq1* (voir ci-dessus). Une carte du locus *Igf2/H19* est donnée en figure 1.4.

Ce locus contient un gène à expression maternelle, *H19*, transcrit en un ARN non-codant, épissé, d'environ 2,5 kb (Bartolomei *et al.*, 1991), et deux gènes à expression paternelle, *Ins2* (gène de la prépro-insuline 2) et *Igf2* (pour : *insulin-like growth factor 2*). Une DMR est située en amont de *H19* ; elle est dérivée de la lignée germinale : hyperméthylée sur l'ADN du spermatozoïde, elle est hypométhylée dans l'ovocyte (Warnecke *et al.*, 1998). L'allèle non méthylé de cette DMR peut fixer la protéine CTCF, qui isole alors le gène *Igf2* des éléments activateurs spécifiques des tissus endodermiques, partagés avec *H19*, qui se situent en aval de *H19*⁵ ; sur l'allèle paternel, la DMR méthylée n'est pas reconnue par CTCF, ce qui permet l'activation de *Igf2* par les éléments activateurs (Bell et Felsenfeld, 2000 ; Hark *et al.*, 2000 ; Kanduri *et al.*, 2000 ; Szabó *et al.*, 2000). En plus de ce rôle, la DMR contrôle directement l'activité du promoteur de *H19* : la méthylation sur l'allèle paternel de la DMR dérivée de la lignée germinale se propage au promoteur de *H19*, et inactive la transcription de *H19*. À partir du moment où la méthylation a atteint le promoteur de *H19*, la DMR est dispensable pour la répression de l'allèle paternel du gène (Srivastava *et al.*, 2000). En revanche, elle reste nécessaire pour la répression de *Igf2* par insulation des éléments activateurs (Srivastava *et al.*, 2000).

Ce double rôle de la DMR conduit à une différence remarquable entre les effets de sa délétion sur chacun des deux chromosomes parentaux : lorsque la DMR est délétée sur le chromosome

⁵Plus précisément, des travaux récents montrent que la fixation de CTCF sur l'allèle non méthylé de cette DMR éloignent physiquement le gène *Igf2* des éléments activateurs, et relocalise ce gène dans un domaine chromatien inactif (Murrell *et al.*, 2004).

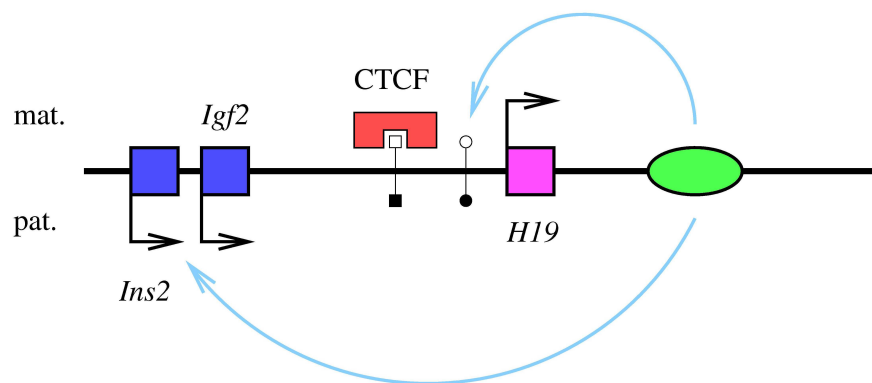


FIG. 1.4 – LE LOCUS *Igf2/H19*. Le gène à expression maternelle est représenté en rose, et les gènes à expression paternelle sont représentés en bleu. Les flèches noires indiquent les sens de transcription. L'ellipse verte représente des éléments activateurs endodermiques, le polygone rouge représente la protéine CTCF, et les flèches bleues, l'activation de la transcription. Les carrés noir et blanc représentent la DMR dérivée de la lignée germinale (noir : hyperméthylé ; blanc : hypométhylé) ; les disques noir et blanc représentent la DMR secondaire (noir : hyperméthylé ; blanc : hypométhylé). Les disques et carrés dessinés au-dessus de la ligne horizontale représentent l'état de méthylation sur l'allèle maternel, et les disques et carrés dessinés en-dessous de la ligne horizontale représentent l'état de méthylation sur l'allèle paternel. Les promoteurs de *Ins2* et *H19* sont éloignés d'environ 100 kb chez la Souris (le schéma n'est pas à l'échelle).

maternel, l'expression de *Igf2* devient biallélique (l'allèle maternel de *Igf2* a accès aux éléments activateurs, en l'absence de l'insulateur), et *H19* est partiellement réprimé, probablement par la compétition de *Igf2* pour les éléments activateurs (Kaffer *et al.*, 2000 ; Thorvaldsen *et al.*, 1998 ; Thorvaldsen *et al.*, 2002) ; lorsque la DMR est délétée sur le chromosome paternel, l'expression de *H19* est activée (son promoteur n'est plus méthylé sur l'allèle paternel), et *Igf2* est partiellement réprimé, probablement, là encore, par compétition pour des activateurs communs avec *H19* (Drewell *et al.*, 2000 ; Kaffer *et al.*, 2000 ; Thorvaldsen *et al.*, 1998 ; Thorvaldsen *et al.*, 2002). Cette DMR, qui contrôle l'empreinte des trois gènes du locus, est donc un centre d'empreinte.

Il est à noter que la plupart des auteurs se sont intéressés au contrôle de l'empreinte de *Igf2* et de *H19*, mais qu'ils ont souvent négligé de généraliser leurs résultats à *Ins2* (Arney, 2003).

Enfin, d'autres séquences participent à la répression des allèles silencieux de *Igf2* (Constância *et al.*, 2000) et *H19* (Drewell *et al.*, 2000) (voir aussi p. 30, pour le mécanisme de l'empreinte de *Igf2* dans les tissus mésodermiques).

d) Le locus PWS/AS.

Des dérégulations des gènes de ce locus sont associées à deux syndromes humains : le syndrome de Prader-Willi (abrégié : PWS, en anglais), et le syndrome d'Angelman (abrégié : AS, en anglais), qui lui ont donné son nom. La figure 1.5 présente l'organisation de ce locus.

Le locus (voir Nicholls et Knepper, 2001 pour revue) contient de nombreux gènes à expression paternelle : *Mkrn3*, *Magel2* et *Ndn*, en amont du centre d'empreinte, le bicistron *Snurf-Snrpn*, de nombreux gènes d'ARN C/D (voir page 53), et un gène antisens au gène *Ube3a*, noté : *Ube3a-as*. *Snurf-Snrpn*, les gènes d'ARN C/D et l'antisens de *Ube3a* semblent cotranscrits (*cf* p. 53). *Snurf-Snrpn* possède plusieurs premiers exons alternatifs, notés U1A, U1B, U1D, et 1 (voir figure 1.5 ; l'exon 1 constitue l'extrémité 5' du transcrit *Snurf-Snrpn*-ARN C/D-*Ube3a-as* représenté). Le locus contient aussi deux gènes à expression maternelle : *Ube3a* et *Atp10c* (également appelé : *Atp10a*).

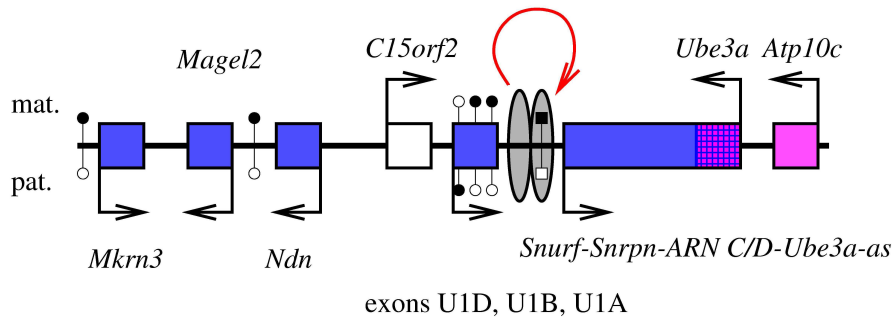


FIG. 1.5 – LE LOCUS PWS/AS. Les gènes à expression maternelle sont représentés en rose, les gènes à expression paternelle sont représentés en bleu, et le gène à expression biallélique, en blanc. L'intersection entre un gène à expression paternelle et un gène à expression maternelle est quadrillée en bleu et rose. Les flèches noires indiquent les sens de transcription. Les carrés noir et blanc représentent la DMR dérivée de la lignée germinale (noir : hyperméthylé ; blanc : hypométhylé) ; les disques noirs et blancs représentent les DMR secondaires (noir : hyperméthylé ; blanc : hypométhylé). Les disques et carré dessinés au-dessus de la ligne horizontale représentent l'état de méthylation sur l'allèle maternel, et les disques et carré dessinés en-dessous de la ligne horizontale représentent l'état de méthylation sur l'allèle paternel. Le centre d'empreinte bipartite est représenté par deux ellipses grisées. La flèche rouge représente le contrôle de l'épigénotype du PWS-IC par celui de l'AS-IC (voir texte). Les promoteurs de *Mkrn3* et *Atp10c* sont éloignés d'environ 3130 kb chez la Souris (le schéma n'est pas à l'échelle).

Le centre d'empreinte (IC, pour : *imprinting center*), identifié entre les exons U1A et 1 de *Snurf-Snrpn*, est bipartite : deux séquences de cet IC, appelées PWS-IC et AS-IC, remplissent des fonctions distinctes. Le PWS-IC (représenté par l'ellipse grisée de droite, sur la figure 1.5) est une DMR (dérivée de la lignée germinale chez la Souris, mais pas chez l'Homme), hyperméthylée sur l'allèle maternel et hypométhylée sur l'allèle paternel. Elle coïncide avec le promoteur principal de *Snurf-Snrpn*, et sa délétion sur le chromosome paternel abolit la transcription de *Snrpn* et *Ube3a-as*, et déréprime l'allèle paternel de *Ube3a* (Dittrich *et al.*, 1996 ; Buiting *et al.*, 2001 ; Chamberlain et Brannan, 2001). L'AS-IC (représenté par l'ellipse grisée de gauche, sur la figure 1.5), est située à 35 kb du PWS-IC. Cette séquence n'est pas différenciellement méthylée, mais présente des différences entre les deux allèles pour d'autres modifications épigénétiques (Perk *et al.*, 2002). L'AS-IC du chromosome maternel est nécessaire à l'acquisition, par le PWS-IC sur ce même chromosome, de son hyperméthylation : si l'AS-IC maternel est délété, le PWS-IC maternel est hypométhylé, et dirige la transcription des gènes à expression paternelle sur le chromosome maternel (Perk *et al.*, 2002). Dans ce contexte génétique, les deux gènes à expression maternelle, *Ube3a* et *Atp10c*, sont réprimés bialléliquement.

Ainsi, l'AS-IC semble contrôler l'épigénotype du PWS-IC, qui contrôle l'expression des gènes du locus. Puisque l'expression de *Ube3a* et celle de *Ube3a-as* corrélaient inversement, sur les deux allèles, l'ARN antisens *Ube3a-as* pourrait être responsable de la répression en *cis* du gène *Ube3a* (Chamberlain et Brannan, 2001).

e) Le locus *Dlk1/Gtl2*.

Une carte de ce locus est donnée en figure 1.6.

Ce locus contient de nombreux gènes non-codants à expression maternelle : *Gtl2* (aussi appelé : *Meg3*) (Miyoshi *et al.*, 2000), des gènes d'ARN C/D (voir Résultats, page 102) et des gènes de microARN (voir Résultats, page 130). Il contient aussi trois gènes codants à expression paternelle : *Dlk1* (Schmidt *et al.*, 2000 ; Takada *et al.*, 2000) (également appelé : *Pref-1*,

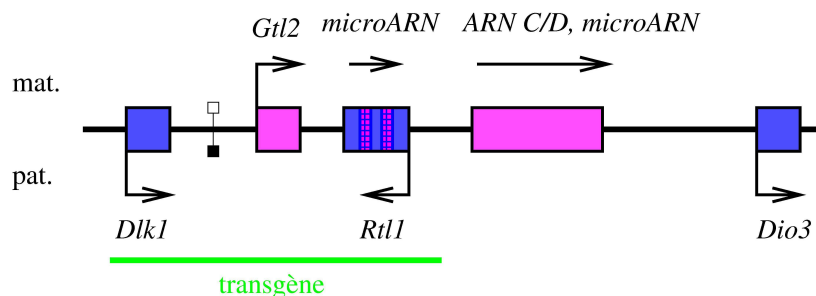


FIG. 1.6 – LE LOCUS *Dlk1/Gtl2*. Les gènes à expression maternelle sont représentés en rose et les gènes à expression paternelle sont représentés en bleu. L'intersection entre un gène à expression paternelle et des gènes à expression maternelle est quadrillée en bleu et rose. Les flèches indiquent les sens de transcription. Les carrés noir et blanc représentent la DMR dérivée de la lignée germinale (noir : hyperméthylé ; blanc : hypométhylé). Le carré dessiné au-dessus de la ligne horizontale représente l'état de méthylation sur l'allèle maternel, et le carré dessiné en-dessous de la ligne horizontale représente l'état de méthylation sur l'allèle paternel. La barre verte horizontale représente la séquence couverte par un transgène dont le gène *Gtl2* est soumis à l'empreinte (voir texte). Les promoteurs de *Dlk1* et *Dio3* sont éloignés d'environ 820 kb chez la Souris (le schéma n'est pas à l'échelle).

FA1, *Zog*, *pG2* et *SCP-1*) code une protéine transmembranaire qui inhibe la différenciation des préadipocytes en adipocytes, *ex vivo* (Smas et Sul, 1993) et *in vivo* (Moon *et al.*, 2002) ; un rétrotransposon, *Rtl1* (également appelé : *PEG11* et *Shr1* chez le Mouton (Charlier *et al.*, 2001b ; Lynch et Tristram, 2003), *Hur1* chez l'Homme, *Rar1* chez le Rat, et *Mor1* chez la Souris : Lynch et Tristram, 2003) (voir Résultats, page 118) ; et *Dio3*, qui code une enzyme qui inactive les hormones thyroïdiennes T_3 et T_4 (Tsai *et al.*, 2002 ; Yevtodiyenko *et al.*, 2002).

Une DMR dérivée de la lignée germinale a été localisée entre les gènes *Dlk1* et *Gtl2* : elle est hypométhylée sur l'allèle maternel, et hyperméthylée sur l'allèle paternel. Le laboratoire d'Anne Ferguson-Smith, en Angleterre, a réalisé une délétion de cette séquence chez la Souris ; l'analyse de l'expression des gènes du locus chez des embryons de Souris délétés de la DMR sur l'un ou l'autre de leurs chromosomes parentaux, est présentée dans les résultats (voir p. 122). Des souris transgéniques pour la séquence génomique couvrant de 3,5 kb en amont de *Dlk1* à 69 kb en aval de *Gtl2* (qui contient, donc, la DMR ; la séquence couverte par ce transgène est représentée par la barre verte, sur la figure 1.6) soumettent le gène *Gtl2* du transgène à l'empreinte génomique parentale : ce transgène contient donc tous les éléments *in cis* requis pour le contrôle de l'empreinte de *Gtl2* (Yevtodiyenko *et al.*, 2004).

1.2.4 Considérations évolutives

Les gènes soumis à l'empreinte génomique parentale sont fonctionnellement haploïdes, donc ce phénomène rend une espèce plus vulnérable aux mutations, ce qui est *a priori* un inconvénient. Et même, à considérer que, pour des raisons inconnues, il serait bénéfique d'être fonctionnellement haploïde, il faudrait justifier que la plupart des gènes ne sont pas soumis à l'empreinte génomique parentale (Wilkins et Haig, 2003). Plusieurs théories ont été proposées pour justifier l'apparition de ce phénomène.

L'hypothèse de la bombe à retardement ovarienne

Cette hypothèse, proposée en 1994 (Varmuza et Mann, 1994), propose un avantage sélectif de l'empreinte génomique parentale pour les espèces vivipares : puisque l'empreinte génomique

parentale rend indispensable une contribution paternelle pour le développement d'un embryon et un placenta, ce phénomène empêche l'apparition intempestive de ces structures invasives, en l'absence de fécondation. En particulier, le trophoblaste (à l'origine du placenta, il envahit l'épithélium utérin, chez la Souris et l'Homme en particulier) est très réduit chez les embryons parthénogénétiques. Puisque les ovocytes s'activent spontanément, sans fécondation, à une haute fréquence, ces tumeurs ovariennes sont particulièrement nombreuses (elles affectent au moins une fois quelques pour-cent des femmes, et jusqu'à plus de 80 % des souris femelles de certaines lignées, au cours de leur vie), et l'empreinte génomique parentale, qui limite leur développement, serait responsable de leur bénignité. Elle représenterait donc un avantage sélectif certain.

Quant aux gènes exprimés à partir de l'allèle maternel uniquement, dont l'empreinte ne protège en rien les femelles contre le développement de parthénogénètes, ils seraient des « victimes innocentes » du processus d'empreinte génomique parentale, et n'auraient pas contribué à son succès évolutif.

Cet argument *ad hoc* n'a pas convaincu tous les lecteurs (Solter, 1994 ; Haig, 1994 ; Moore, 1994). Il est signalé de surcroît qu'un seul locus à expression uniquement paternelle suffirait à protéger les femelles du développement des parthénogénètes, alors que de nombreux loci soumis à l'empreinte ont été identifiés ; que le raisonnement de Varmuza et Mann n'est pas généralisable aux Angiospermes ni à certains Mammifères, qui ne sont pas menacés par un organe prompt à envahir l'organisme maternel, comme le trophoblaste murin ou humain, mais qui connaissent malgré tout le phénomène d'empreinte génomique parentale ; enfin, tous les auteurs ne s'accordent pas sur l'incidence réelle du risque des tumeurs ovariennes (Solter, 1994 ; Haig, 1994 ; Moore, 1994 ; Wilkins et Haig, 2003 ; Mann et Varmuza, 1994).

La facilitation de l'évolution

L'empreinte génomique parentale, en réprimant un allèle indépendamment de sa séquence, puis en le réactivant dans les générations suivantes au gré de son mode de transmission, maternel ou paternel, soustrait les allèles des gènes concernés à la sélection naturelle. Un allèle peu adapté aux conditions environnementales a donc la possibilité de se propager dans une population jusqu'à ce que les conditions lui soient plus favorables, ou jusqu'à ce qu'il ait lui-même muté jusqu'à devenir avantageux dans les mêmes conditions. L'empreinte génomique parentale permet donc le maintien dans une population d'une plus grande variété d'allèles, ce qui peut présenter un avantage évolutif (Beaudet et Jiang, 2002). D'autre part, puisque les auteurs considèrent que les mécanismes moléculaires impliqués dans l'empreinte génomique parentale (modifications épigénétiques, régulations par des ARN non-codants, compétition pour des éléments communs *in cis*) permettent une plus grande variabilité du taux d'expression des gènes soumis à l'empreinte, le phénomène permettrait un ajustement fin de l'expression de ces gènes, et une répartition des phénotypes sur un continuum, ce qui faciliterait l'émergence du génotype le plus adapté.

Cette théorie ne rend pas compte de la répartition de l'empreinte génomique parentale dans le règne vivant (ses avantages s'appliqueraient à tous, or seuls les Mammifères et les Angiospermes connaissent l'empreinte), et elle explique mal pourquoi le phénomène n'affecte que certains gènes en particulier (Wilkins et Haig, 2003).

Victimes innocentes de l'extinction de parasites génomiques

La comparaison de transgènes qui, chez la Souris, sont spontanément soumis à l'empreinte génomique parentale, avec des gènes endogènes qui subissent le phénomène, a amené D. Barlow à proposer que l'empreinte génomique parentale dérive du système de défense contre les parasites génomiques (endovirus et transposons) par méthylation de l'ADN de leurs gènes (Barlow, 1993). Effectivement, les gènes soumis à l'empreinte partagent de nombreuses caractéristiques avec

les parasites génomiques : ils sont pauvres en introns, contiennent des séquences répétées en tandem, et plusieurs gènes soumis à l'empreinte sont visiblement issus de rétrotransposition ; réciproquement, les rétrotransposons sont souvent réprimés différemment dans les deux lignées germinales (Yoder *et al.*, 1997).

L'empreinte génomique parentale serait donc une conséquence fortuite de l'extinction de séquences parasites, et ne présenterait donc pas nécessairement d'avantage évolutif *per se*. Les gènes soumis à l'empreinte seraient donc des victimes innocentes de la répression des endovirus et transposons insérés à proximité : l'apparition d'une séquence parasite, et son extinction (inefficace, elle n'affecterait qu'un allèle), aboutiraient à la répression d'un allèle du locus entier. Les mécanismes de l'empreinte génomique parentale et ceux de l'extinction des rétrotransposons sont similaires (Yoder *et al.*, 1997), et partagent des effecteurs communs (Bourc'his et Bestor, 2004 ; Lippman *et al.*, 2004). La délétion de séquences répétées d'un locus soumis à l'empreinte génomique parentale a même abouti à la levée de l'empreinte du locus (Yoon *et al.*, 2002) ; ce résultat n'est toutefois pas généralisable à tous les loci (Lewis *et al.*, 2003).

Modulation de l'expression des gènes

Le développement des Mammifères semble très sensible au dosage des copies des gènes, puisque chez la Souris, toutes les monosomies et toutes les trisomies s'accompagnent de défauts développementaux aboutissant à la létalité. L'activation ou l'inactivation d'un des deux allèles à une étape particulière du développement permettrait donc de moduler l'expression de certains gènes (Solter, 1988).

Cette hypothèse s'accorde mal avec plusieurs observations : elle n'explique pas pourquoi l'empreinte génomique parentale serait restreinte aux Mammifères et aux Angiospermes, pourquoi l'empreinte de certains gènes est polymorphe⁶, ni pourquoi les gènes à expression paternelle favorisent souvent la croissance, alors que les gènes à expression maternelle tendent à l'inhiber. Enfin, des simulations montrent que, selon cette hypothèse, l'empreinte génomique parentale ne devrait pas se maintenir stablement dans une population (Weisstein et Spencer, 2003).

La théorie du conflit

Cette théorie est la plus populaire, bien qu'elle soit elle aussi mise en difficulté sur certains points (*cf* plus bas). Elle a été proposée en 1991, et justifie l'empreinte génomique parentale par un conflit entre les intérêts des allèles hérités de la mère, et du père (Moore et Haig, 1991) : les allèles du père ont de meilleures chances de se transmettre si l'enfant se porte bien, quitte à ce que ce soit au détriment de la mère et de ses gestations futures ; les allèles de la mère auront de meilleures chances de se transmettre si la mère est épargnée par sa gestation (ce qui lui permettra d'autres gestations ultérieures). Et en effet, il est fréquemment observé que les gènes à expression paternelle favorisent la croissance, alors que les gènes à expression maternelle la répriment (*cf* tableau 1.2, page 42).

Cette théorie a ensuite été généralisée : les allèles d'un individu ne profitent pas uniquement de sa santé, mais aussi de celle de sa famille, puisque l'individu partage (en moyenne) 50 % de ses allèles avec ses frères et sœurs, 25 % avec ses demi-frères et demi-sœurs, *etc.* Le conflit entre l'allèle maternel et l'allèle paternel peut donc éclater pour le partage des ressources trophiques avec tous ceux qui ne possèdent pas les allèles maternels et paternels dans les mêmes proportions que l'individu en question (Wilkins et Haig, 2003).

Elle explique pourquoi l'empreinte est limitée aux Mammifères et Angiospermes : dans ces deux phyla, l'embryon se développe dans une structure coûteuse pour la mère (l'embryon, et éventuellement le placenta, pour les Mammifères ; l'embryon dans la graine, elle-même contenue

⁶Comme par exemple IGF2R (Xu *et al.*, 1993) et WT1 (Jinno *et al.*, 1994) dans la population humaine.

dans un fruit, pour les Angiospermes)⁷. L'allocation des ressources entre les différents descendants, qui n'ont pas forcément le même père, est donc un enjeu important de la guerre que se livrent les allèles. L'empreinte serait donc apparue deux fois au cours de l'évolution, chez des Animaux, et chez des Plantes, alors que leur dernier ancêtre commun en était certainement dépourvu ; ce serait un exemple de convergence fonctionnelle (Haig et Westoby, 1991).

Cependant, tous les gènes soumis à l'empreinte génomique parentale n'ont pas un rôle évident dans la croissance (*cf* tableau 1.2, page 42), et, étant donné que plusieurs protocoles couramment utilisés pour détecter les gènes soumis à l'empreinte ne peuvent, par construction, révéler que des gènes exprimés pendant le développement embryonnaire (voir page 21, les protocoles de détection de gènes soumis à l'empreinte), il est prévisible que la liste dont nous disposons soit biaisée en faveur des gènes impliqués dans le développement. D'autre part, des gènes restent soumis à l'empreinte après le sevrage, alors que l'individu n'y est plus en compétition spécifiquement avec ses frères et sœurs (et demi-frères et demi-sœurs maternels) (Wilkins et Haig, 2003), et des gènes dont l'effet est évident sur la croissance, *Igf1* chez la Souris (Liu *et al.*, 1993) et *IGF2R* chez l'Homme (du moins, dans la majorité de la population) (Kalscheuer *et al.*, 1993 ; Xu *et al.*, 1993), ne sont pas soumis à l'empreinte.

Avantage sélectif de la distinction des chromosomes parentaux

Chez de nombreuses espèces animales, végétales et chez des Champignons, et pas uniquement chez les Mammifères et les Angiospermes, les deux chromosomes parentaux peuvent se comporter différemment : outre les différences d'expression entre allèles parentaux, chez les Mammifères et les Angiospermes, les deux chromosomes homologues peuvent se répartir, à la méiose, en fonction de leur origine parentale ; ils peuvent adopter des structures chromatiniennes distinctes, qui influenceront différemment l'expression de transgènes, ou subir des événements de destruction sélective de l'un des deux chromosomes parentaux (Pardo-Manuel de Villena *et al.*, 2000 ; de la Casa-Esperón et Sapienza, 2003). L'empreinte génomique parentale, observée chez les Mammifères et les Angiospermes, pourrait ainsi n'être qu'un aspect d'un phénomène beaucoup plus général, de différence épigénétique entre les chromosomes homologues. Ces différences épigénétiques présenteraient un avantage sélectif : par exemple, elles pourraient faciliter la reconnaissance des chromosomes homologues, à la méiose ; les différences d'expression des allèles parentaux auraient pu, en plus, conférer des avantages spécifiques aux mâles ou aux femelles dans les espèces où se manifeste l'empreinte génomique parentale (comme le postule la théorie du conflit, voir ci-dessus), mais ces différences d'expression ne constitueraient pas le seul avantage sélectif de la différence épigénétique des deux chromosomes parentaux.

L'une des qualités de cette hypothèse est qu'elle réconcilie des observations très variées, provenant d'espèces très éloignées, dans un unique cadre théorique ; elle justifie de surcroît les différences épigénétiques entre allèles, chez les Mammifères, dans des tissus où le gène concerné n'est pas exprimé. Cependant, l'avantage sélectif que conférerait, chez les espèces à reproduction sexuée, la distinction des deux chromosomes homologues de chaque paire n'est pas évident, et pour le moment, aucune donnée expérimentale ne soutient l'hypothèse selon laquelle des modifications épigénétiques différentielles favoriseraient la reconnaissance mutuelle des chromosomes homologues à la méiose.

⁷Cependant, chez les Gymnospermes, l'embryon se développe aussi aux dépens de sa mère (certes hors d'un fruit), donc la théorie du conflit devrait prédire une empreinte génomique parentale chez les Gymnospermes, or elle n'a jamais été rapportée ; toutefois, à ma connaissance, elle n'a jamais été recherchée non plus.

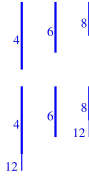
					
					
	Femelle double hétérozygote Rb(8.12)5Bnr/Rb(6.12)3Sic	X	Mâle double hétérozygote Rb(8.12)5Bnr/Rb(4.12)9Bnr		
Gamètes					
	 [†]	 [†]	 [†]	matUPD12	 [†]
	 [†]	 équil.	 équil.	 [†]	 [†]
	 [†]	 équil.	 équil.	 [†]	 [†]
	 patUPD12	 [†]	 [†]	 [†]	 [†]
	 [†]	 [†]	 [†]	 [†]	 [†]
	 [†]	 [†]	 [†]	 [†]	 [†]

FIG. 1.7 – OBTENTION DE SOURIS À DISOMIES UNIPARENTALES. Des souris doubles hétérozygotes pour deux translocations robertsoniennes impliquant la paire n°12 sont croisées ; parmi les descendants possibles, les 5/6 seront aneuploïdes, donc non viables [†] ; 1/9 auront un caryotype anormal (en raison de la translocation) mais équilibré, et seront viables (« équil. ») ; et 1/18 auront un caryotype équilibré, mais dont les deux chromosomes n°12 proviendront du même parent (« matUPD12 » héritent de deux chromosomes 12 maternels, et « patUPD12 », de deux chromosomes 12 paternels).

Locus	Gène	Allèle exprimé	Fonction
prox. 2	<i>Gatm</i>	maternel (Sandell <i>et al.</i> , 2003)	synthèse de la créatine (Wyss et Kaddurah-Daouk, 2000) (sous le nom du gène humain, AGAT)
distal 2	<i>Nnat</i>	paternel (Kagitani <i>et al.</i> , 1997) (Kikyo <i>et al.</i> , 1997)	développement du système nerveux, protection contre le stress chimique (Kagitani <i>et al.</i> , 1997)
distal 2	<i>Gnas</i>	maternel (Yu <i>et al.</i> , 1998)	réprime la croissance (Weinstein <i>et al.</i> , 2004) (Yu <i>et al.</i> , 2000)
	<i>Gnasxl</i>	paternel (Peters <i>et al.</i> , 1999)	active la croissance (Yu <i>et al.</i> , 2000)
	<i>Nesp</i>	maternel (Peters <i>et al.</i> , 1999)	chromogranine antagoniste d'un récepteur à la sérotonine (Ischia <i>et al.</i> , 1997)
	<i>Nespas</i>	paternel (Wroe <i>et al.</i> , 2000)	répresseur de <i>Nesp</i> (Wroe <i>et al.</i> , 2000)
prox. 6	<i>Calcr</i>	maternel (Hoshiya <i>et al.</i> , 2003)	récepteur transmembranaire (Pondel, 2000)
	<i>Sgce</i>	paternel (Piras <i>et al.</i> , 2000)	adhérence cellulaire (McNally <i>et al.</i> , 1998)
	<i>Peg10</i>	paternel (Ono <i>et al.</i> , 2003)	
	<i>Neurabin</i>	maternel (Ono <i>et al.</i> , 2003)	formation des neurites (Nakanishi <i>et al.</i> , 1997)
	<i>Pon2</i>	maternel (Ono <i>et al.</i> , 2003)	
	<i>Pon3</i>	maternel (Ono <i>et al.</i> , 2003)	
	<i>Asb4</i>	maternel (Mizuno <i>et al.</i> , 2002)	
sub-prox. 6	<i>Peg1/Mest</i>	paternel (Kaneko-Ishino <i>et al.</i> , 1995)	croissance embryonnaire, comportement (Lefebvre <i>et al.</i> , 1998)
	<i>Copg2</i>	maternel (Lee <i>et al.</i> , 2000)	
	<i>Copg2as</i>	paternel (Lee <i>et al.</i> , 2000)	
	<i>Mit1/Lb90</i>	paternel (Lee <i>et al.</i> , 2000)	
	<i>Nap1l5</i>	paternel (Smith <i>et al.</i> , 2003)	
prox. 7	<i>Zim1</i>	maternel (Kim <i>et al.</i> , 1999)	facteur de transcription (Kim <i>et al.</i> , 1999)
	<i>Peg3/Pw1</i>	paternel (Kaneko-Ishino <i>et al.</i> , 1995)	active la croissance, comportement (Li <i>et al.</i> , 1999)
	<i>Ocat/Usp29</i>	paternel	protéase de la voie

Locus	Gène	Allèle exprimé	Fonction
prox. 7		(Szeto <i>et al.</i> , 2000) (Kim <i>et al.</i> , 2000)	<i>ubiquitine</i> (Kim <i>et al.</i> , 2000)
	<i>Zim3</i>	maternel (Kim <i>et al.</i> , 2001)	<i>ARN non-codant (gène de protéine à doigts de zinc dégénéré), antisens à Ocat/Usp29</i> (Kim <i>et al.</i> , 2001)
	<i>Zfp264</i>	paternel (Kim <i>et al.</i> , 2001)	<i>ARN non-codant (gène de protéine à doigts de zinc dégénéré)</i> (Kim <i>et al.</i> , 2001)
central 7	<i>Snurf-Snrpn</i>	paternel (Leff <i>et al.</i> , 1992)	bicistron (protéine inconnue, et protéine du spliceosome) (Gray <i>et al.</i> , 1999)
	<i>Pwcr1, Ipw</i> <i>ARN C/D</i>	paternel (Wevrick et Francke, 1997) (de los Santos <i>et al.</i> , 2000) (Cavaillé <i>et al.</i> , 2000) (Meguro <i>et al.</i> , 2001b)	gène-hôte d'ARN C/D répétés (Cavaillé <i>et al.</i> , 2000)
	<i>Magel2</i>	paternel (Boccaccio <i>et al.</i> , 1999)	
	<i>Ndn</i>	paternel (MacDonald et Wevrick, 1997)	<i>développement du système nerveux</i> (Uetsuki <i>et al.</i> , 1996)
	<i>Zfp127/Mkrn3</i>	paternel (Jong <i>et al.</i> , 1999) (Hershko <i>et al.</i> , 1999)	<i>interaction avec des ARN</i> Jong <i>et al.</i> , 1999
	<i>Frat3</i>	paternel (Chai <i>et al.</i> , 2001)	développement embryonnaire (Jonkers <i>et al.</i> , 1999)
	<i>Atp10c</i>	maternel (Kashiwagi <i>et al.</i> , 2003), mais discuté (Kayashima <i>et al.</i> , 2003)	composition de la membrane plasmique (Meguro <i>et al.</i> , 2001a)
	<i>Ube3a</i>	maternel (Albrecht <i>et al.</i> , 1997)	ubiquitine-protéine ligase (Huibregtse <i>et al.</i> , 1993)
	<i>Ube3a-as</i>	paternel (Chamberlain et Brannan, 2001)	
distal 7	<i>H19</i>	maternel (Bartolomei <i>et al.</i> , 1991)	croissance cellulaire (Ayesh <i>et al.</i> , 2002)
	<i>Igf2</i>	paternel (DeChiara <i>et al.</i> , 1991)	croissance embryonnaire (Jones et Clemmons, 1995)
	<i>Igf2as</i>	paternel (Moore <i>et al.</i> , 1997)	
	<i>Ins2</i>	paternel (Giddings <i>et al.</i> , 1994)	croissance, métabolisme (Wentworth <i>et al.</i> , 1986)
	<i>Ascl2/Mash2</i>	maternel (Guillemot <i>et al.</i> , 1995)	développement du placenta (Guillemot <i>et al.</i> , 1994)
	<i>Kvlqt1</i>	maternel (Gould et Pfeifer, 1998) (Jiang <i>et al.</i> , 1998)	sous-unité d'un canal à K ⁺ , contrôle du rythme cardiaque (Wang <i>et al.</i> , 1996)
	<i>Kvlqt1-AS/</i> <i>Lit1/Kcnlot1</i>	paternel (Smilnich <i>et al.</i> , 1999)	antisens à <i>Kvlqt1</i> (Smilnich <i>et al.</i> , 1999)

Locus	Gène	Allèle exprimé	Fonction
distal 7	<i>p57^{KIP2}/Cdkn1c</i>	maternel (Hatada et Mukai, 1995)	régulation du cycle cellulaire (Lee <i>et al.</i> , 1995)
	<i>Msuit/Lit1/Kcnlot1</i>	maternel (Onyango <i>et al.</i> , 2000)	
	<i>Slc22a1l/Itm/Impt1/Orctl2</i>	maternel (Cooper <i>et al.</i> , 1998) (Dao <i>et al.</i> , 1998) (Morisaki <i>et al.</i> , 1998)	transporteur membranaire (Dao <i>et al.</i> , 1998)
	<i>Ipl/Tssc3</i>	maternel (Qian <i>et al.</i> , 1997)	transduction du signal (Qian <i>et al.</i> , 1997)
	<i>Tssc4</i>	maternel (Paulsen <i>et al.</i> , 2000)	
	<i>Opbh1</i>	maternel (Engemann <i>et al.</i> , 2000)	interaction avec l'oxystérol (Engemann <i>et al.</i> , 2000)
distal 9	<i>A19</i>	paternel (de la Puente <i>et al.</i> , 2002)	
	<i>Rasgrf1</i>	paternel (Plass <i>et al.</i> , 1996)	contrôle du cycle cellulaire (Shou <i>et al.</i> , 1992)
prox. 10	<i>Zac1/Lot1</i>	paternel (Piras <i>et al.</i> , 2000)	anti-prolifération (Varrault <i>et al.</i> , 1998)
distal 10	<i>Dcn</i>	maternel (Mizuno <i>et al.</i> , 2002)	protéine de la matrice extracellulaire (Scholzen <i>et al.</i> , 1994)
prox. 11	<i>Meg1/Grb10</i>	maternel (Miyoshi <i>et al.</i> , 1998)	inhibiteur du récepteur à l'insuline (Liu et Roth, 1995)
	<i>U2af1-rs1</i>	paternel (Hatada <i>et al.</i> , 1993)	associée au snRNA U2 (Hatada <i>et al.</i> , 1993)
	<i>Murr1</i>	maternel (Wang <i>et al.</i> , 2004b)	homéostasie du cuivre (van De Sluis <i>et al.</i> , 2002)
distal 12	<i>Dlk1/Pref-1</i>	paternel (Takada <i>et al.</i> , 2000)	répresseur de la différenciation des adipocytes (Moon <i>et al.</i> , 2002)
	<i>Rtl1/PEG11</i>	paternel (Seitz <i>et al.</i> , 2003)	
	<i>Gtl2/Meg3</i>	maternel (Miyoshi <i>et al.</i> , 2000)	
	<i>Rian,ARN C/D</i>	maternel (Cavaillé <i>et al.</i> , 2002)	gène-hôte d'ARN C/D
	<i>Mirg, microARN</i>	maternel (Seitz <i>et al.</i> , 2003) (Seitz <i>et al.</i> , 2004)	répression de Rtl1
	<i>Dio3</i>	paternel (Tsai <i>et al.</i> , 2002) (Yevtodiyenko <i>et al.</i> , 2002)	inactivation des hormones thyroïdiennes T3 et T4 (Hernandez <i>et al.</i> , 1999)
distal 14	<i>Htr2a</i>	maternel (Kato <i>et al.</i> , 1998)	récepteur à la sérotonine (Guillet-Deniau <i>et al.</i> , 1997)
distal 15	<i>Slc38a4/Ata3</i>	paternel	transport des acides aminés

Locus	Gène	Allèle exprimé	Fonction
		(Mizuno <i>et al.</i> , 2002)	(Sugawara <i>et al.</i> , 2000)
	<i>Peg13</i>	paternel (Smith <i>et al.</i> , 2003)	
prox. 17	<i>Slc22a2/Orct2</i>	maternel (Zwart <i>et al.</i> , 2001)	transporteur transmembranaire (Koepsell, 1998)
prox. 17	<i>Slc22a3/Orct3/EMT</i>	maternel (Zwart <i>et al.</i> , 2001)	transporteur transmembranaire (Koepsell, 1998)
	<i>Igf2r</i>	maternel (Barlow <i>et al.</i> , 1991)	inhibition de la croissance (Hawkes et Kar, 2004)
	<i>Air</i>	paternel (Wutz <i>et al.</i> , 1997)	empreinte du locus <i>Igf2r/Air</i> (Sleutels <i>et al.</i> , 2002)
prox. 18	<i>Impact</i>	paternel (Hagiwara <i>et al.</i> , 1997)	
prox. 19	<i>Ins1</i>	paternel (Giddings <i>et al.</i> , 1994)	facteur de croissance (Wentworth <i>et al.</i> , 1986)
chr. X	quasi-totalité du chromosome	maternel (voir p. 91)	

TAB. 1.2: GÈNES MURINS SOUMIS À L'EMPREINTE GÉNOMIQUE PARENTALE. La première colonne donne la position du locus ; la deuxième colonne donne le nom des gènes (avec éventuellement les noms alternatifs, séparés par « / ») ; la troisième colonne donne l'allèle préférentiellement exprimé, dans au moins un tissu de la Souris (et renvoie aux articles qui ont démontré cette empreinte) ; la quatrième colonne donne les fonctions connues de ces gènes (les fonctions supposées, mais non démontrées, sont indiquées en italique).

1.3 Les ARN C/D

Les ARN C/D doivent leur nom aux motifs de séquence appelés « boîte C » et « boîte D » (voir 1.3.1, ci-dessous). Jusqu'en 2000, les seuls ARN C/D connus étaient des ARN nucléolaires d'Eucaryotes, appelés « petits ARN nucléolaires à boîtes C et D », ou : snoRNA C/D (pour : *C/D box small nucleolar RNAs*, en anglais) (voir page 50). Mais la découverte de petits ARN aux propriétés très similaires chez les Archæ (Gaspin *et al.*, 2000 ; Omer *et al.*, 2000), dépourvues de noyau donc de nucléole, puis la mise en évidence, chez des Eucaryotes, d'ARN C/D accumulés dans une autre structure subnucléaire, les corpuscules de Cajal (Darzacq *et al.*, 2002), ont élargi la famille des ARN C/D.

1.3.1 Structure des ARN C/D et composition des RNP C/D

Les ARN C/D sont des ARN stables (Nicoloso *et al.*, 1994 ; Nicoloso *et al.*, 1996 ; Qu *et al.*, 1994 ; Qu *et al.*, 1995), et relativement abondants (Hüttenhofer *et al.*, 2001). La structure des ARN C/D guides de méthylation de Vertébrés (qui constituent la majorité des ARN C/D, voir 1.3.3, page 49) est présentée en figure 1.8, page 47 (deux ARN C/D de Vertébrés, U3 et U8, sont plus longs que les ARN C/D guides de méthylation, et présentent des extensions en 5' de la structure présentée en figure 1.8). Ces petits ARN (leur longueur est généralement comprise entre 70 et 100 nt) portent la « boîte C » (consensus : **RUGAUGA**) à proximité de leur extrémité 5', et la « boîte D » (consensus : **CUGA**) à proximité de leur extrémité 3'. Dans la plupart des ARN C/D, la courte séquence en 5' du nucléotide immédiatement en amont de la boîte C, et la courte séquence en 3' de la boîte D, sont complémentaires (voir Bachellerie *et al.*, 2000, pour revue). La structure en tige (issue de l'appariement de ces deux séquences terminales), associée aux boîtes C et D, est connue sous le nom de « motif C/D » (voir figure 1.8). Dans ce motif, les deux brins d'ARN adoptent une conformation particulière (Watkins *et al.*, 2000), appelée « kink-turn ». Le kink-turn est un motif structural impliqué dans plusieurs types d'interactions ARN / protéine (Vidovic *et al.*, 2000 ; Klein *et al.*, 2001). Enfin, les ARN C/D possèdent des copies dégénérées des boîtes C et D, appelées respectivement C' (Kiss-László *et al.*, 1998) et D' (Pellizzoni *et al.*, 1994 ; Tycowski *et al.*, 1996a), plus internes dans la séquence de l'ARN (voir figure 1.8 ; la boîte D' est systématiquement en 5' de la boîte C').

L'appariement des séquences 5' et 3' terminales de l'ARN C/D, rapproche les boîtes C et D, en permettant la formation du motif C/D. Il semble que les boîtes C' et D' adoptent une structure similaire, puisque dans la plupart des ARN C/D dont les boîtes C' et D' sont fonctionnelles, celles-ci sont très souvent proches dans la séquence primaire (elles ne sont séparées que par 3 à 9 nucléotides) ; dans les quelques exemples où elles sont plus éloignées, des appariements intramoléculaires dans la séquence séparant les boîtes C' et D' semblent pouvoir les rapprocher (Kiss-László *et al.*, 1998).

Le motif C/D interagit avec des protéines spécifiques ; chez les Eucaryotes, quatre protéines forment, avec l'ARN C/D, une ribonucléoparticule (ou : RNP) active. La protéine NhpX (également appelée 15.5 kDa, et Snul3p chez la Levure) interagit directement avec le kink-turn du motif C/D (Vidovic *et al.*, 2000 ; Watkins *et al.*, 2000). Les protéines Nop56 et Nop58, et la fibrillarine (appelée « Nop1p » chez la Levure) complètent la particule, et leur fixation requiert la présence de NhpX (Watkins *et al.*, 2002). Nop56, Nop58 et la fibrillarine interagissent elles aussi avec l'ARN : Nop58 préférentiellement avec la boîte C, Nop56 préférentiellement avec la boîte C', et la fibrillarine, avec les boîtes D et D' (Cahill *et al.*, 2002).

La fibrillarine présente des motifs de séquence partagés par un grand nombre de méthyltransférases (Niewmierzycka et Clarke, 1999), et la structure tridimensionnelle de l'extrémité C-terminale d'une fibrillarine d'Archæ s'apparente à celle du domaine catalytique des méthyl-

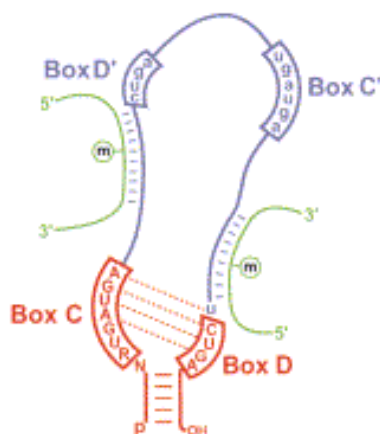


FIG. 1.8 – STRUCTURE DES ARN C/D GUIDES DE MÉTHYLATION. Figure adaptée de Kiss, 2002. Le motif C/D est dessiné en rouge. Le repliement du motif C/D en kink-turn implique des appariements, de type Watson-Crick et non-Watson-Crick, entre les nucléotides des boîtes (pointillés rouges). Les boîtes (C, D, C' et D') sont encadrées. La plupart des ARN C/D possèdent un ou deux élément(s) antisens (voir texte, p. 49), situé(s) immédiatement en amont des boîtes D ou D', ou un nucléotide en amont, capable(s) de s'apparier aux précurseurs d'ARN ribosomiques ou de petits ARN nucléaires, dessinés en vert. Le nucléotide apparié au 5ème nucléotide en amont de la boîte D (ou D') est méthylé sur son oxygène 2' (lettre « m » dans un disque).

transférases (Wang *et al.*, 2000). La fibrillarine est donc probablement une méthyl-transférase, ce qui expliquerait l'activité de méthylation de la plupart des RNP C/D (voir 1.3.3, p. 49).

Enfin, deux protéines semblent capables de s'associer aux ARN C/D, mais sont absentes de la particule mature. Ces protéines, Tip48 et Tip49 (également appelées, entre autres noms, p50 et p55, respectivement), sont connues pour interagir avec la machinerie transcriptionnelle, et pourraient donc être impliquées dans les étapes très précoces de la maturation des ARN C/D, sur leur site de transcription (Newman *et al.*, 2000).

1.3.2 Biosynthèse des ARN C/D

L'organisation génomique des gènes d'ARN C/D est très variable, et dépend des groupes phylétiques.

- Chez les Vertébrés, la plupart des ARN C/D connus sont issus de la maturation d'introns⁸ (voir Bachellerie *et al.*, 2000, pour revue) : après débranchement du lasso, les ARN C/D sont libérés par des clivages endo- puis exonucléolytiques. Les gènes-hôtes des ARN C/D introniques sont très souvent impliqués dans la traduction (gènes de protéines ribosomiques, de facteurs d'initiation et d'élongation de la traduction) (Bachellerie *et al.*, 2000 ; Nicoloso *et al.*, 1996), mais ils peuvent aussi être sans lien manifeste avec la traduction (gène de l'ATP synthase β), ou même, ne pas avoir de phase ouverte de lecture (Tycowski *et al.*, 1996a ; Smith et Steitz, 1998). Ainsi, les exons du gène UHG (qui héberge les ARN

⁸Les gènes de certains ARN C/D de Vertébrés ne sont pas introniques, mais transcrits indépendamment. Ces quelques ARN C/D transcrits indépendamment (c'est le cas des ARN U3, U8 et U13) ont la particularité de ne pas guider de méthylation sur des ARN-cibles, comme la plupart des ARN C/D connus (*cf* 1.3.3). L'extrémité 5' des ARN C/D transcrits indépendamment porte une coiffe (une guanosine triméthylée sur la base, et reliée au nucléotide suivant par une liaison triphosphate 5'-5'), alors que les ARN C/D maturés à partir d'introns portent un nucléotide phosphorylé en 5' à leur extrémité 5'.

C/D U22 et U25 à 31), et du gène *gas5* (qui héberge les ARN C/D U44, U47 et U74 à 81) sont nettement moins conservés phylogénétiquement que les introns, et leurs ARN épissés ont une courte durée de vie : les exons de ces gènes ne semblent avoir aucune fonctionnalité (Tycowski *et al.*, 1996a ; Smith et Steitz, 1998). De nombreux gènes-hôtes d'ARN C/D chez les Vertébrés appartiennent à la classe des 5' TOP-gènes (gènes à expression ubiquitaire, riches en pyrimidine autour du site de démarrage de la transcription, et à 5' UTR courte ; les ARN de ces gènes ont la particularité de se dissocier des ribosomes lorsque la croissance cellulaire est arrêtée, et de s'accumuler dans des particules de taille plus modeste) (Smith et Steitz, 1998), ce qui pourrait coordonner les synthèses des ARN C/D.

- Chez un Insecte, *Drosophila melanogaster*, le clonage de petits ARN (Yuan *et al.*, 2003) et la recherche bio-informatique de gènes d'ARN C/D dans le génome (Accardo *et al.*, 2004) ont récemment enrichi la collection d'ARN C/D connus chez cet organisme. Alors qu'il semblait que la presque totalité des gènes d'ARN C/D de la Drosophile étaient introniques (Renalier *et al.*, 1996 ; Tycowski et Steitz, 2001 ; Yuan *et al.*, 2003) (en particulier, comme chez les Vertébrés, tous les ARN C/D guides de méthylation de l'ARNpré-ribosomique étaient introniques), l'analyse bio-informatique a mis en évidence de nombreux gènes d'ARN C/D susceptibles de guider des 2'-O-méthylations de l'ARNpré-ribosomique (noté : ARNpré-r) dans des exons, et dans des régions présumées intergéniques (Accardo *et al.*, 2004).
- Chez la Levure, seule une minorité des ARN C/D est intronique : la plupart des ARN C/D sont transcrits par l'ARN-polymérase II à partir de gènes indépendants (Maxwell et Fournier, 1995) ; certains gènes d'ARN C/D sont polycistroniques, et la maturation du transcrit fait intervenir des clivages endonucléolytiques entre les ARN C/D (dans le cas des ARN C/D à gènes polycistroniques), et une dégradation exonucléolytique des séquences flanquantes aux ARN C/D (Qu *et al.*, 1999 ; Fatica *et al.*, 2000 ; Lee *et al.*, 2003a). On connaît aussi quelques exemples d'ARN C/D de Levure issus de la maturation d'introns de gènes codants impliqués dans la traduction (par exemple, des gènes de protéines ribosomiques) (Qu *et al.*, 1995 ; Lowe et Eddy, 1999).
- Chez les Plantes, la plupart des ARN C/D connus sont polycistroniques. Ces polycistrons peuvent être transcrits indépendamment, ou bien être hébergés dans des introns (Brown *et al.*, 2003, pour une revue).
- Chez les Archæ : une recherche bioinformatique a permis d'identifier des homologues des ARN C/D dans trois génomes entièrement séquencés (Gaspin *et al.*, 2000) ; il n'existe pour l'heure aucune donnée sur le mode de biosynthèse de ces petits ARN, mais il est intéressant de noter que les loci de plusieurs de ces ARN recouvrent des phases ouvertes de lecture, suggérant une relation entre les modes d'expression des gènes de ces petits ARN et des autres gènes. En outre, l'un des ARN C/D est hébergé dans un intron de son propre substrat, un pré-ARN_t du tryptophane (Clouet d'Orval *et al.*, 2001).

La voie principale de maturation des ARN C/D introniques nécessite un débranchement du lasso, mais, chez la Levure (Ooi *et al.*, 1998) et les Vertébrés (Caffarelli *et al.*, 1994) au moins, un mécanisme alternatif faisant intervenir des clivages endonucléolytiques (chez les Vertébrés, ces clivages prennent place sur l'ARNpré-m ; cette voie de maturation des ARN C/D est donc alternative à l'épissage) puis une dégradation exonucléolytique des séquences flanquantes, peut libérer l'ARN C/D mature.

Les boîtes C et D sont nécessaires à la maturation des ARN C/D introniques, probablement en fixant des protéines qui protègent l'ARN C/D de la dégradation exonucléolytique (Caffarelli *et al.*, 1996) ; ces deux séquences sont d'ailleurs impliquées également dans la stabilité des ARN C/D matures, qu'ils soient d'origine intronique (Caffarelli *et al.*, 1996) ou non-intronique (Huang *et al.*, 1992).

1.3.3 Fonctions des ARN C/D

ARN C/D et clivage des ARN pré-ribosomiques

Les ARN C/D U3 et U22 (ainsi que U8 chez les Vertébrés et U14 chez la Levure) participent à la maturation des ARN pré-ribosomiques : U3 et U22 (et U14 chez la Levure) sont requis pour des réactions de clivage nécessaires à la libération de l'ARN ribosomique de 18S mature (Kass *et al.*, 1990 ; Hughes et Ares, 1991 ; Tycowski *et al.*, 1994 ; Li *et al.*, 1990), alors que U8 est requis pour des clivages de la voie de maturation des ARN ribosomiques de 5,8S et de 28S chez les Vertébrés (Peculis et Steitz, 1993).

Le mode d'action de ces ARN n'est pas encore bien compris, mais dans le cas de U3, U8 et U14, il semble faire intervenir des appariements entre des segments de ces ARN C/D, et l'ARN pré-ribosomique (ARNpré-r). Plus précisément, plusieurs expériences de mutagenèse dirigée sur l'ARN C/D U3, suivies de mutagenèse compensatoire sur l'ARNpré-r, ont montré que U3 s'apparie à l'ARN pré-ribosomique (Beltrame et Tollervey, 1995 ; Sharma et Tollervey, 1999 ; Borovjagin et Gerbi, 2004) ; le même type d'approche a montré que U14 s'y appariait également (Liang et Fournier, 1995) ; des expériences de compétition (par des oligonucléotides complémentaires de l'ARN C/D U8) indiquent que la fonction de l'ARN U8 des Vertébrés nécessite l'appariement de son extrémité 5' avec l'ARNpré-r (Peculis, 1997)⁹.

Ces segments de complémentarité à l'ARNpré-r ne sont pas situés immédiatement en amont, ou 1 nt en amont, de la boîte D (ou D'), à la différence des « éléments antisens » présentés ci-dessous.

ARN C/D et le guidage de la réaction de 2'-O-méthylation

De nombreux ARN C/D présentent une longue complémentarité (de 10 à 21 nt de long) à l'ARNpré-r ou à un précurseur de petit ARN nucléaire, située immédiatement en amont de leur boîte D, ou D' (ou 1 nt en amont). Cette séquence est appelée « élément antisens » (voir figure 1.8, page 47) (voir Bachellerie *et al.*, 2002, pour une revue).

La comparaison des éléments antisens des différents ARN C/D a permis d'établir une propriété remarquable : les séquences d'ARN_r reconnues par ces éléments¹⁰ portent toujours une fonction 2'-O-méthyl, et systématiquement, l'un des nucléotides 2'-O-méthylés s'apparie avec le 5ème nucléotide en amont de la boîte D, ou D', dans l'élément antisens de l'ARN C/D (cf figure 1.8) (Kiss-László *et al.*, 1996 ; Nicoloso *et al.*, 1996).

Il a donc été suggéré (Kiss-László *et al.*, 1996 ; Nicoloso *et al.*, 1996) que la reconnaissance des régions à méthyliser sur un ARNpré-r par ces ARN C/D guidait la machinerie de méthylation vers ces sites ; l'appariement entre l'ARN C/D et l'ARNpré-r (étant donnée la cinétique de la 2'-O-méthylation, ce précurseur est le transcrit primaire) serait donc responsable de la spécificité de la réaction de méthylation.

De fait, il a bien été montré que la reconnaissance de l'ARNpré-r par les ARN C/D était nécessaire et suffisante au guidage de la réaction de méthylation :

- la déplétion d'un ARN C/D abolit la 2'-O-méthylation du nucléotide de l'ARN_r apparié au 5ème nucléotide en amont de la boîte D ou D' (Kiss-László *et al.*, 1996 ; Lowe et Eddy, 1999 ; Qu *et al.*, 1999 ; Tycowski *et al.*, 1996b) ; les autres 2'-O-méthylations ne sont alors pas affectées¹¹ (Kiss-László *et al.*, 1996 ; Lowe et Eddy, 1999) ;

⁹Chez la Levure, dépourvue d'ARN U8, cette fonction serait assurée, en *cis*, par l'ARNpré-r lui-même (Côté *et al.*, 2002).

¹⁰À l'époque, le rôle des ARN C/D dans la maturation des snRNA n'était pas connu.

¹¹À l'exception des méthylations de C₆₄₈, G₁₄₄₈, A₂₂₇₉ et G₂₉₁₉ (numérotation chez la Levure) de l'ARN de la grande sous-unité ribosomique : ces quatre nucléotides 2'-O-méthylés sont tous immédiatement en 3' d'un autre nucléotide 2'-O-méthylé (ce sont les seuls dans ce cas), et, alors qu'on ne leur a trouvé aucun ARN C/D guide de méthylation par une recherche bio-informatique (Lowe et Eddy, 1999), leur méthylation est abolie si l'on déplete

- la délétion d'un nucléotide en amont de la boîte D, qui ne perturbe pas l'appariement entre le guide et sa cible, mais qui décale la boîte D par rapport à l'élément antisens, déplace, comme prévu, le site méthylé sur l'ARN_r d'un nucléotide en direction du 3' (Cavaillé *et al.*, 1996 ; Kiss-László *et al.*, 1996) ;
- la transfection de cellules de Mammifère par un guide de méthylation artificiel, dirigé contre une séquence d'ARN_r dépourvue de 2'-O-méthylation, provoque une méthylation massive à la position prédite ; et la cotransfection d'un minigène dépourvu de séquence d'ARN_r, placé sous le contrôle d'un promoteur de l'ARN-polymérase I, avec un guide de méthylation artificiel dirigé contre ce minigène, provoque la méthylation d'environ 90 % des ARN dérivés du minigène (Cavaillé *et al.*, 1996).

On connaît actuellement des ARN C/D guides de méthylation, potentiels ou avérés, pour 51 des 55 sites 2'-O-méthylés des ARN_r de *Saccharomyces cerevisiae* (Lowe et Eddy, 1999) et 93 des 105 sites 2'-O-méthylés des ARN_r de Mammifères (Hüttenhofer *et al.*, 2001).

Ces ARN C/D, qui s'accumulent tous dans le nucléole, sont appelés « snoRNA C/D » (pour : *C/D box small nucleolar RNAs*). Il a été démontré que quelques snoRNA C/D, qui ne présentent pas de complémentarité à l'ARN_r, mais au petit ARN nucléaire (snRNA, pour : *small nuclear RNA*) U6, remplissent le même rôle de guide de méthylation, sur U6 (Kiss-László *et al.*, 1998 ; Tycowski *et al.*, 1998 ; Ganot *et al.*, 1999).

Une autre classe d'ARN C/D s'accumule dans les corpuscules de Cajal ; ils ont été nommés « scaRNA », pour *small Cajal body-specific RNAs* (Darzacq *et al.*, 2002). Ce sont également des guides de méthylation : leurs éléments antisens sont complémentaires de séquences des snRNA U1, U4 et U5 (Jády et Kiss, 2001 ; Darzacq *et al.*, 2002), et la réaction de 2'-O-méthylation de leurs ARN-cibles a lieu dans les corpuscules de Cajal, par un mécanisme qui semble identique à celui de la réaction de 2'-O-méthylation nucléolaire des ARN ribosomiques (Jády *et al.*, 2003). Alors qu'il avait été montré que le motif C/D suffisait à l'adressage nucléolaire des snoRNA C/D (Samarsky *et al.*, 1998 ; Narayanan *et al.*, 1999b), les scaRNA C/D, pourtant munis d'un motif C/D, s'accumulent dans les corpuscules de Cajal. Cette localisation particulière pourrait être due à un motif de séquence spécifique des scaRNA C/D, à l'image de celui qui a été identifié dans une autre classe de scaRNA, les scaRNA H/ACA¹², dont l'effet sur la localisation est épistatique à celui du motif H/ACA (Richard *et al.*, 2003).

Ainsi, les ARN C/D peuvent guider des 2'-O-méthylations d'ARN transcrits par les ARN-polymérase II et III, en addition des ARN ribosomiques, transcrits par l'ARN-polymérase I.

Les ARN C/D sont donc responsables du guidage de la réaction de 2'-O-méthylation ; en revanche, ils ne semblent pas porter l'activité catalytique de transfert de méthyl. La protéine fibrillarine, associée à tous les ARN C/D (y compris ceux qui ne guident pas de méthylation, mais qui participent à des clivages de l'ARN_{pré-r}), est très certainement responsable de cette activité catalytique, étant donnée sa similarité avec des méthyl-transférases (voir page 46).

La formule développée d'un nucléotide 2'-O-méthylé est donnée en figure 1.9, page 52. Il est tentant d'imaginer que cette modification de nucléotide joue un rôle dans le fonctionnement

le guide de la méthylation du nucléotide immédiatement en 5'.

¹²Les scaRNA H/ACA contiennent les motifs de séquence appelés « H » (consensus : ANANNA) et « ACA » (consensus : ACA). De même que pour les ARN C/D, la majorité des ARN H/ACA connus sont nucléolaires (voir Kiss, 2002, pour une revue), où ils guident une autre modification post-transcriptionnelle des ARN_{pré-r}, la pseudo-uridylation ; les boîtes H et ACA sont suffisantes à la localisation nucléolaire des snoRNA H/ACA (Narayanan *et al.*, 1999a). Les scaRNA H/ACA (qui guident des pseudo-uridylation sur les snRNA U2 et U5 : Darzacq *et al.*, 2002 ; Kiss *et al.*, 2002) possèdent un motif UGAG, souvent dégénéré sur l'une des deux premières positions, et localisé sur des boucles apicales de tiges des ARN H/ACA ; ce motif, appelé « CAB-box » (pour : *Cajal body box*) est nécessaire à la localisation des scaRNA H/ACA dans les corpuscules de Cajal, plutôt que dans le nucléole (Richard *et al.*, 2003).

ou le repliement des ARN_r ; curieusement cependant, aucune 2'-O-méthylation testée n'était nécessaire à la fonctionnalité du ribosome, et même la suppression simultanée de sept 2'-O-méthylations n'affecte pas la viabilité de *Saccharomyces cerevisiae* (Kiss-László *et al.*, 1996 ; Qu *et al.*, 1999 ; Lowe et Eddy, 1999) (dans chacune de ces études, des mutations d'ARN C/D bien précis ont permis d'abolir spécifiquement certaines 2'-O-méthylations). En revanche, une mutation ponctuelle de la fibrillarine de Levure, Nop1p, qui affecte le site putatif de liaison à la S-adénosyl méthionine (le donneur de groupement méthyl pour toutes les méthyltransférases connues), et annule l'activité de 2'-O-méthylation des ARN ribosomiques, est létale à température non permissive (cette mutation est appelée *nop1.3*) (Tollervey *et al.*, 1993). La létalité ne semble pas due à un défaut de maturation des ARN_r, puisque l'accumulation des ARN_r correctement maturés (mais non 2'-O-méthylés) n'est pas affectée ; il semble donc que la 2'-O-méthylation des ARN ribosomiques, dans sa globalité, soit nécessaire à la survie de la Levure. Alternativement, la létalité du mutant *nop1.3* à température non permissive pourrait être due au défaut de 2'-O-méthylation de quelques nucléotides bien spécifiques, dont la modification serait guidée par des ARN C/D qui n'ont pas été délétés dans les trois études présentées ci-dessus.

D'autre part, une analyse plus fine de la souche de Levure délétée pour sept guides de méthylation (Qu *et al.*, 1999) met en évidence une augmentation significative du taux de translecture des codons stop (Henras, 2002). Dans cette étude, un système rapporteur permet la quantification des événements de translecture (*i.e.* : poursuite de la traduction en aval d'un codon stop) de codons stop placés dans un contexte de séquence qui augmente la fréquence de translecture (Stahl *et al.*, 1995). Il apparaît que l'abolition des 2'-O-méthylations guidées par ces sept guides double et triple la fréquence de translecture des codons stop UAG et UAA (respectivement), sans affecter notablement la translecture du codon stop UGA. Cet effet pourrait être dû à un défaut subtil de structuration du ribosome, ou à l'altération de la séquence d'événements qui suit l'entrée d'un codon stop dans le site A du ribosome, en l'absence de ces sept 2'-O-méthylations (Henras, 2002).

Outre les ARN ribosomiques, d'autres classes d'ARN portent également des 2'-O-méthylations, mais leur fonction n'est jamais clairement établie : les ARN_t (elles participeraient à la structuration des ARN_t) (Agris, 1996), différents petits ARN nucléaires (elles influenceraient l'appariement des snRNA aux ARN_{pré-m}, et le repliement des snRNA ; un spliceosome contenant le snRNA U2 transcrit *in vitro*, donc dépourvu de modifications post-transcriptionnelles — dont les 2'-O-méthylations — est inactif) (Massenet *et al.*, 1998), et les ARN_m, dont la coiffe, qui protège l'extrémité 5' de l'ARN, peut être 2'-O-méthylée sur les deux premiers nucléotides en aval de la guanosine méthylée sur la base (elles participeraient à la protection de l'ARN_m) (Banerjee, 1980 ; Narayan et Rottman, 1992).

Au niveau moléculaire, la 2'-O-méthylation favorise la conformation 3'-endo du nucléotide modifié, qui éloigne le groupement méthyl de l'oxygène 2' de la base (voir figure 1.9, page 52). Comme la conformation 3'-endo stabilise l'interaction d'empilement des bases (appelée *stacking* en anglais), elle rigidifie les structures en hélice ; cette propriété expliquerait la grande abondance en nucléotides 2'-O-méthylés des organismes thermophiles : la stabilisation des structures de ces ARN présenterait un avantage aux hautes températures (Davis, 1998).

Des chaperones ribonucléiques ?

Les snoRNA C/D, guides de méthylation ou impliqués dans des réactions de clivage, s'apparient très précocément à l'ARN_{pré-r}, alors que sa transcription n'est pas encore terminée (Mougey *et al.*, 1993 ; Maden, 1998). De plus, ces appariements sont relativement longs (les appariements des guides de méthylation, par exemple, sont longs de 10 à 21 nt). Étant donné le

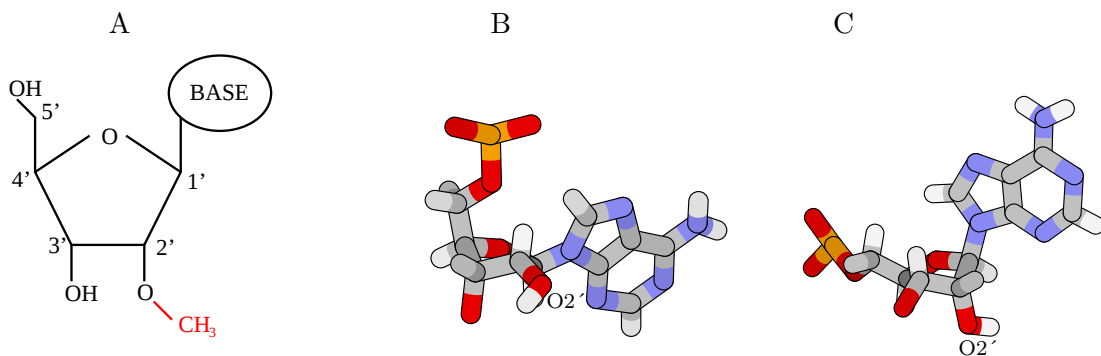


FIG. 1.9 – LA 2'-O-MÉTHYLATION. A : Formule développée d'un nucléotide 2'-O-méthylé (le groupement méthyl de l'oxygène 2' est représenté en rouge). B : Conformation 2'-endo (sur l'exemple d'une adénosine phosphorylée en 5'); la liaison C3'-C2' est au premier plan, et l'oxygène 2' est signalé. C : Conformation 3'-endo (sur l'exemple d'une adénosine phosphorylée en 5'); la liaison C3'-C2' est au premier plan, et l'oxygène 2' est signalé. Les structures des figures B et C sont tirées du fichier UH0005 de la Nucleic Acid Database (NDB).

grand nombre de guides de méthylation (une centaine chez les Vertébrés, une cinquantaine chez la Levure) et de snoRNA H/ACA (voir note de bas de page, p. 50) (en nombre équivalent), les séquences de l'ARN_{pré-r} de Vertébrés impliquées dans ces appariements représentent environ un tiers de la séquence totale des ARN_r matures (Yu *et al.*, 1999). L'hybridation de cette nuée de petits ARN sur l'ARN_{pré-r} n'est certainement pas sans conséquence sur le repliement des futurs ARN ribosomiques : en rigidifiant leurs séquences-cibles, ils peuvent faciliter ou empêcher des interactions entre séquences de l'ARN_{pré-r}; en s'appariant à une séquence donnée, ils interdisent son repliement sur une autre région de l'ARN_{pré-r}. Finalement, la structuration du transcrit naissant sera contrainte par toutes ces hybridations, et l'ARN_{pré-r} adoptera une structure spatiale qu'il n'aurait pas forcément prise en l'absence des snoRNA. Ensuite, lorsque les snoRNA se détacheront de leur ARN-cible, celui-ci aura adopté une structure suffisamment stable (éventuellement stabilisée par des interactions avec des protéines ribosomiques, associées entretemps) pour ne plus avoir la possibilité de se transconformer : la structure spatiale des ARN_r sera tombée dans un piège cinétique.

Cet effet probable de l'appariement des snoRNA est peut-être leur fonction principale : en contraignant la conformation des ARN_r, ils joueraient un rôle de chaperones (Steitz et Tycowski, 1995 ; Bachelierie et Cavaillé, 1998). Cette idée est renforcée par l'observation que les snoRNA à deux éléments antisens reconnaissent souvent des séquences spatialement proches, dans la structure de l'ARN_r mature (ils pourraient les avoir rapprochées en s'appariant simultanément à chacune d'elles) (Bachelierie *et al.*, 2002). Et dans le cadre de cette hypothèse, la méthylation (ou la pseudo-uridylation, dans le cas des appariements avec les snoRNA H/ACA) d'un nucléotide de l'ARN_{pré-r} n'est peut-être qu'un signal qui permet le décrochage du snoRNA, une fois sa fonction accomplie, et elle n'aurait pas de rôle particulier en tant que telle (Bachelierie et Cavaillé, 1998). Il faut noter toutefois que dans le cas des 2'-O-méthylations guidées sur les snRNA, ceux-ci ont déjà été transcrits, exportés dans le cytoplasme, et se sont associés à des protéines, avant de rentrer dans le noyau et y être méthylés et pseudo-uridylés (dans le nucléole, par des snoRNA, pour le snRNA U6, et dans les corpuscules de Cajal, par des scaRNA, pour les snRNA U1, U2, U4 et U5) : il est donc probable que ces ARN aient déjà acquis une structure stable, au moment de leur interaction avec les ARN guides de modifications, et que cet appariement ne perturbera pas excessivement leur conformation.

1.3.4 ARN C/D orphelins

Ces dernières années, des travaux de clonage de petits ARN ont permis l'identification de nombreux nouveaux ARN C/D chez différents organismes (Jády et Kiss, 2000 ; Cavaillé *et al.*, 2001 ; Hüttenhofer *et al.*, 2001 ; Gogolevskaya *et al.*, 2002 ; Marker *et al.*, 2002 ; Tang *et al.*, 2002 ; Yuan *et al.*, 2003 ; Zhou *et al.*, 2004). Parmi ces nouveaux ARN C/D, une proportion significative ne présentent pas de complémentarité évidente aux ARN ribosomiques ou aux petits ARN nucléaires en amont de leurs boîtes D ou D' ; ces ARN C/D, sans cible triviale, sont dits « orphelins ». Leur fonction éventuelle reste donc mystérieuse : ils pourraient participer à la biogenèse des ARN_r ou des snRNA, par un mécanisme qui n'impliquerait pas d'appariement des séquences en amont des boîtes D ou D' (par exemple, en s'appariant par des séquences quelconques, ils pourraient jouer un rôle de chaperones : voir page 51). Ils pourraient également guider une réaction de 2'-O-méthylation sur un autre type d'ARN : en particulier, l'un des ARN C/D clonés à partir du cerveau de Souris présente une complémentarité de 18 nt à un ARN_m (voir page 57) (Cavaillé *et al.*, 2000 ; Hüttenhofer *et al.*, 2001).

Le tableau 1.3 recense les ARN C/D de Mammifères sans cible évidente.

1.3.5 Gènes d'ARN C/D soumis à l'empreinte génomique parentale

L'une des études présentées en 1.3.4 a mis en évidence 5 ARN C/D de Souris spécifiques du cerveau (Hüttenhofer *et al.*, 2001). L'identification des gènes de trois d'entre eux (MBII-13, MBII-52 et MBII-85¹³) a révélé qu'ils étaient physiquement très proches : ces trois gènes sont situés dans le locus 7C chez la Souris (Cavaillé *et al.*, 2000). Ces ARN ont des orthologues humains (appelés respectivement HBII-13, HBII-52 et HBII-85), dont les gènes se trouvent dans le locus humain de synténie conservée, le locus 15q11-13. Les trois ARN C/D humains s'expriment aussi préférentiellement dans le cerveau, mais HBII-13 et HBII-85 sont aussi clairement détectés dans un muscle squelettique, dans le poumon et dans le rein.

Outre leur expression spécifique de tissu, inédite jusqu'alors pour des ARN C/D, ils présentent deux autres particularités (Cavaillé *et al.*, 2000) :

- Les gènes de MBII-52, MBII-85, HBII-52 et HBII-85 sont répétés plusieurs dizaines de fois. Chaque copie est intronique (comme la plupart des gènes d'ARN C/D de Vertébrés : voir page 47), mais l'intron qui contient la séquence de l'ARN C/D, et l'exon adjacent, sont répétés plusieurs dizaines de fois (ainsi, le locus humain 15q11-13 contient 27 copies du gène HBII-85 (Runte *et al.*, 2001), et 47 copies de HBII-52).
- Ce locus (également appelé « locus PWS/AS » : voir page 35) est connu pour être soumis à l'empreinte génomique parentale, chez la Souris et chez l'Homme (voir page 35). L'analyse de l'expression des ARN HBII-13, HBII-52 et HBII-85 chez des sujets humains présentant une délétion sur l'un des deux chromosomes 15 parentaux a montré que ces trois gènes sont exprimés uniquement à partir de l'allèle paternel. De la même manière, l'expression de MBII-13, MBII-52 et MBII-85 chez des souris présentant des défauts similaires démontre que ces gènes sont également soumis à l'empreinte chez la Souris, et à expression paternelle.

Les exons des gènes-hôtes de MBII-52/HBII-52 et MBII-85/HBII-85 sont non-codants, et peu conservés entre l'Homme et la Souris. Une étude ultérieure suggère que ces nombreuses copies de gènes d'ARN C/D sont co-transcrites, sous la forme d'un long précurseur, puisque des ARN issus du raboutage de chaque exon avec ses voisins peuvent être détectés, pour tous les exons compris entre le promoteur du transcrit SNURF-SNRPN, et l'ARN antisens de UBE3A,

¹³Leurs noms sont hérités de la méthode qui a permis leur découverte : ils ont été clonés à partir d'une fraction de petite taille (appelée « fraction II ») d'ARN de cerveau de Souris. Les clones obtenus à partir de cette fraction ont donc été nommés : MBII-*N*, pour : *mouse brain, fraction II, clone n°N*.

Nom	Identifié chez	Référence	Remarques
U83	Homme	Jády et Kiss, 2000	non associé à l'ARNpré-r ; guide la 2'-O-méthylation d'un substrat artificiel
U84	Homme	Jády et Kiss, 2000	non associé à l'ARNpré-r ; guide la 2'-O-méthylation d'un substrat artificiel
RBII-36	Rat	Cavaillé <i>et al.</i> , 2001	spécifique du cerveau ; propre au Rat
MBI-46	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-52	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-106	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-4	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-115	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-163	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-170	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-244	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-289	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-295	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-343	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-361	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-366	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-419	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-426	Souris	Hüttenhofer <i>et al.</i> , 2001	
MBI-13	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
MBI-48	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
MBI-49	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
MBI-52	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
MBI-78	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
MBI-85	Souris	Hüttenhofer <i>et al.</i> , 2001	spécifique du cerveau
HBII-436	Homme	Runte <i>et al.</i> , 2001	exprimé principalement dans le cerveau, un muscle squelettique, le poumon et le rein
HBII-438	Homme	Runte <i>et al.</i> , 2001	exprimé principalement dans le cerveau et le rein

TAB. 1.3 – LES ARN C/D ORPHELINS DE MAMMIFÈRES.

UBE3A-AS¹⁴ (Runte *et al.*, 2001). Cette étude a de plus identifié quatre nouveaux gènes d'ARN C/D dans le locus : les gènes HBII-436, HBII-437, HBII-438A et HBII-438B (les ARN matures prédits, exprimés à partir de ces deux derniers gènes, seraient rigoureusement identiques ; l'ARN HBII-437 serait muté sur la boîte D, ce qui devrait le déstabiliser ; et effectivement, il n'est pas détecté). Les ARN C/D HBII-436 et HBII-438 sont exprimés uniquement à partir de l'allèle paternel.

La figure 1.10 résume l'organisation du locus humain 15q11-13.

¹⁴Cependant, aucun des produits détectés n'est un produit d'épissage complet, qui contiendrait l'exon le plus en 5' et l'exon le plus en 3' ; il reste donc possible qu'aucun ARN ne couvre toute la longueur du transcrit annoncé, et qu'il existe des promoteurs disséminés dans la séquence. En particulier, puisque les ARN HBII-13 et HBII-85 d'une part, et HBII-52 d'autre part, n'ont pas exactement le même patron d'expression (les deux premiers sont aussi détectés dans un muscle squelettique, le poumon et le rein, à la différence de HBII-52), il est envisageable que les 47 copies de HBII-52 aient un promoteur qui leur est propre. Ainsi, la région 3' du long transcrit proposé (la région antisens à *Ube3a*) semble transcrite à partir d'un promoteur distinct de celui de SNURF-SNRPN (Landers *et al.*, 2004). Alternativement, si un unique transcrit joue le rôle de précurseur commun pour HBII-13, HBII-85 et HBII-52, des maturations post-transcriptionnelles spécifiques de tissu, ou un arrêt de la transcription spécifique de tissu (puisque les copies de HBII-52 sont les plus en 3') pourraient justifier de la différence du patron d'expression de ces ARN C/D.

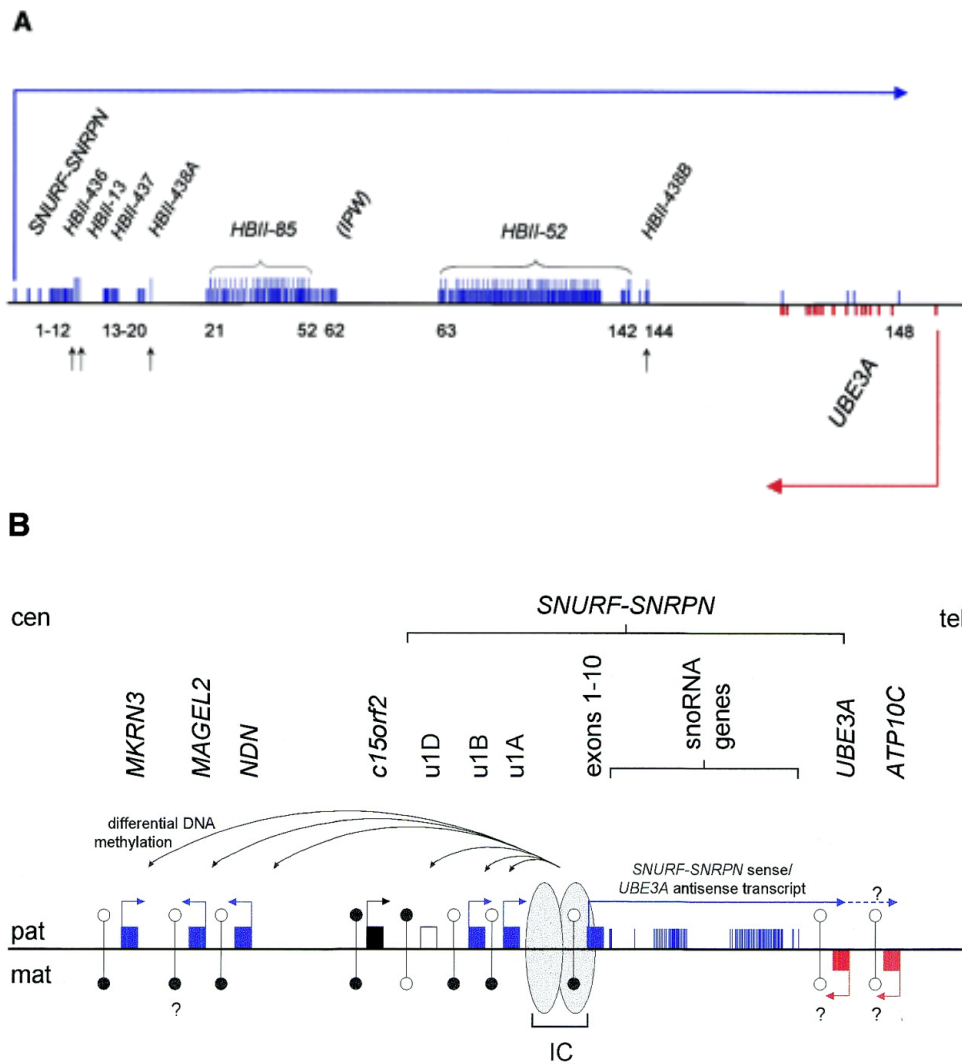


FIG. 1.10 – LES GÈNES D'ARN C/D DU LOCUS HUMAIN 15q11-13. Figures tirées de Runte *et al.*, 2001. Les séquences exprimées à partir de l'allèle paternel sont en bleu, et les séquences exprimées à partir de l'allèle maternel, en rouge. **A** : Les exons sont indiqués par des barres verticales courtes, et les gènes d'ARN C/D, par des barres verticales longues. Les numéros sous la ligne horizontale sont les numéros d'exons particuliers. Les quatre nouveaux gènes d'ARN C/D identifiés par Runte *et al.*, 2001 sont signalés par des flèches verticales. Les flèches bleue et rouge donnent les sens de transcription des séquences à expression paternelle (en bleu) et à expression maternelle (en rouge). **B** : Le centre d'empreinte bipartite (voir page 35), représenté par deux ellipses grisées, contrôle l'expression des gènes à expression paternelle, en particulier du précurseur des ARN C/D (appelés « snoRNA » sur le schéma), qui pourrait être cotranscrit avec l'ARN antisens à UBE3A. Celui-ci pourrait être responsable de la répression de l'allèle paternel de UBE3A (éventuellement aussi de ATP10C) : voir également la page 35, pour plus de détails sur l'empreinte de ce locus, et la partie 1.2.3, pour les rôles que pourraient jouer les ARN antisens dans le mécanisme de l'empreinte. « pat » : allèle paternel ; « mat » : allèle maternel. Disque blanc : site de méthylation de l'ADN hypométhylé ; disque noir : site de méthylation de l'ADN hyperméthylé. Point d'interrogation : statut de méthylation inconnu ; point d'interrogation et flèche en pointillés : statut d'expression inconnu.

MBII-52/HBII-52 et l'ARN_m 5-HT_{2C} : contrôle de l'édition ?

Les ARN MBII-52 et MBII-85 s'accumulent dans le nucléole (Jérôme Cavaillé, communication personnelle) : ce sont donc des snoRNA, petits ARN nucléolaires. Aucun des ARN C/D exprimés à partir des gènes de ce locus ne présente de complémentarité notable aux précurseurs d'ARN_r, ni aux précurseurs de snRNA, en amont de leurs boîtes D ou D'. En revanche, l'un d'entre eux (l'ARN MBII-52/HBII-52) possède, immédiatement en amont de sa boîte D, une complémentarité de 18 nt à un ARN_m spécifique du cerveau, l'ARN_m du récepteur 5-HT_{2C} à la sérotonine (Cavaillé *et al.*, 2000). Cette complémentarité est parfaitement conservée entre l'Homme et la Souris.

Le récepteur 5-HT_{2C} est un récepteur à sept domaines transmembranaires, couplé à la protéine G. La séquence de son ARN_m codant la deuxième boucle intracellulaire (domaine protéique qui interagit avec la protéine G) est susceptible d'être éditée, sur cinq bases particulières (cf figure 1.11) : ces cinq adénines sont éditées en inosines, différemment selon les régions du cerveau (Burns *et al.*, 1997 ; Niswender *et al.*, 1998). Les ARN_m édités ne codent pas la même séquence protéique que la séquence génomique de 5-HT_{2C} (en effet, l'inosine s'apparie préférentiellement avec la cytosine ; lors de la traduction, elle est donc lue comme une guanine, là où la forme génomique porte une adénine). Finalement, les différentes formes éditées de l'ARN_m 5-HT_{2C} (issues de combinaisons variables de l'édition des cinq sites) sont traduites en protéines dont les propriétés de couplage à la protéine G diffèrent nettement (Burns *et al.*, 1997). Ainsi, l'édition de cet ARN semble jouer un rôle dans le contrôle de la transduction du signal de la sérotonine.

Or, le 5ème nucléotide en amont de la boîte D de l'ARN MBII-52/HBII-52 s'apparie avec l'un de ces sites d'édition : le site C (voir figure 1.11), ce qui suggère qu'une 2'-O-méthylation de l'ARN_m de 5HT_{2C}, guidée par MBII-52/HBII-52, pourrait jouer un rôle dans la régulation de l'édition de cet ARN. Il a en effet été montré *in vitro* que la 2'-O-méthylation inhibait l'édition d'ARN synthétiques par l'adénosine déaminase ADAR 2 (l'une des deux enzymes qui édite le site C) (Yi-Brunozzi *et al.*, 1999).

D'autre part, un site donneur d'épissage alternatif se trouve à proximité de la séquence complémentaire à l'élément antisens de l'ARN MBII-52/HBII-52 (voir figure 1.11) ; l'utilisation de ce site d'épissage fait apparaître un codon stop prématuré, et la traduction de ce produit d'épissage donne une protéine tronquée, dépourvue d'activité physiologique (Canton *et al.*, 1996).

Une 2'-O-méthylation guidée par MBII-52/HBII-52, à proximité de ce site d'épissage alternatif, ou tout simplement l'appariement de l'ARN C/D, pourrait peut-être influencer le choix du site donneur d'épissage (Cavaillé *et al.*, 2000).

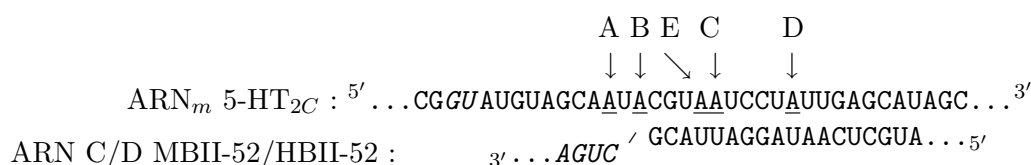


FIG. 1.11 – COMPLÉMENTARITÉ ENTRE L'ARN_m 5-HT_{2C} ET L'ARN C/D MBII-52/HBII-52. Le 5ème exon de l'ARN_m 5-HT_{2C} (en haut) peut être édité sur cinq sites (soulignés, sur la figure), appelés « A », « B », « C », « D » et « E ». Le site E avait d'abord été nommé « C' » (Niswender *et al.*, 1998). Le dinucléotide GU écrit en italique est un site alternatif donneur d'épissage. L'ARN C/D MBII-52/HBII-52 (en bas) présente une complémentarité de 18 nt à cet ARN_m, immédiatement en amont de sa boîte D (en italique).

MBII-85/HBII-85 et le syndrome de Prader-Willi

Hormis MBII-52/HBII-52 (voir ci-dessus), aucun autre ARN C/D de ce locus n'a d'ARN-cible évident. Au cours de ma thèse, j'ai recherché les complémentarités (en amont des boîtes D et D') entre ces ARN C/D et d'autres ARN C/D caractérisés au cours de ma thèse (voir Résultats, p. 102), et les précurseurs d'ARN_r et de snRNA : les résultats de cette analyse sont présentés en Annexe 2, page 173.

Une observation suggère que leurs gènes ont une fonction biologique, et qu'elle n'est pas forcément en relation avec la biogenèse des ribosomes. Le locus humain 15q11-13 est en effet impliqué dans une maladie neurodéveloppementale, le syndrome de Prader-Willi : lorsque des délétions sur le chromosome paternel abolissent l'expression de gènes à expression paternelle de ce locus, le sujet présente des troubles du développement (il est de taille réduite, et les gonades sont peu développées) et du comportement (il est hypotonique, retardé mentalement, et ne ressent pas la satiété ; il est donc hyperphagique, ce qui conduit invariablement à une obésité). L'analyse d'un modèle murin qui mime partiellement ce phénotype, suite à une délétion sur le locus orthologue (Tsai *et al.*, 1999 ; Runte *et al.*, 2001), et l'étude d'une famille humaine présentant un réarrangement chromosomique dans ce locus (Gallagher *et al.*, 2002), montrent que les seuls gènes connus dans la région critique pour le syndrome (*i.e.* : ceux dont le défaut d'expression corrèle systématiquement avec le syndrome) sont les gènes de MBII-85/HBII-85 et HBII-438A. Il semble donc que la répression de l'un au moins de ces gènes (peut-être même : d'une copie particulière de MBII-85/HBII-85, puisque toutes ne sont pas absolument identiques) soit responsable de l'apparition du syndrome. Ces ARN C/D seraient donc impliqués dans le contrôle du développement et du comportement.

1.4 MicroARN, siRNA et leur métabolisme

1.4.1 Historique

Les microARN, des régulateurs endogènes

Chez le Nématode *Caenorhabditis elegans*, la séquence des étapes du développement larvaire peut être affectée chez certains mutants : les gènes pour lesquels ils sont mutés sont appelés « hétérochroniques ». Ainsi, des mutants peuvent exécuter un programme de divisions et de différenciations spécifiques d'un stade larvaire donné à un stade antérieur ou postérieur. C'est notamment le cas du mutant *lin-4(e912)*, qui reproduit, à des stades tardifs, des programmes spécifiques des étapes précédentes (Chalfie *et al.*, 1981), dans un grand nombre de lignages cellulaires au cours des différents stades larvaires. Or un mutant nul pour un autre gène, *lin-14*, présente un phénotype inverse (Ambros et Horvitz, 1987), et la mutation *lin-14(0)* est épistatique à la mutation *lin-4(e912)* (Ambros, 1989), ce qui indique que *lin-4* serait un répresseur de *lin-14*. De plus, *lin-14* est réprimé au niveau traductionnel ou post-traductionnel (le niveau de l'ARN_m de *lin-14* est sensiblement le même aux stades larvaires L1 d'une part, et L2 et L3 d'autre part, alors que la quantité de protéine Lin-14p est environ dix fois plus faible en L2/L3 qu'en L1), et cette répression est abolie chez le mutant *lin-4(e912)* (Wightman *et al.*, 1993). En outre, des mutations de la 3'UTR de l'ARN_m *lin-14* (qui confèrent un phénotype gain-de-fonction pour *lin-14*) réduisent également cette répression (Wightman *et al.*, 1991) : il semblait donc que la répression de *lin-14* prenait place au niveau traductionnel, par une interaction entre la 3'UTR de l'ARN_m *lin-14* et le produit du gène *lin-4*. Le produit du gène *lin-4*, quant à lui, est un petit ARN retrouvé sous deux formes : *lin-4S* (d'environ 22 nt de long), et *lin-4L* (environ 61 nt)¹⁵; ce dernier est imparfaitement autoccomplémentaire, et peut donc se refermer en une structure en tige-boucle (voir figure 1.12) (Lee *et al.*, 1993). Ces ARN *lin-4* sont partiellement complémentaires à sept sites de la 3'UTR de l'ARN_m-cible *lin-14* (voir figure 1.12) (Wightman *et al.*, 1993).

Ainsi, l'ARN *lin-4S*, le plus petit ARN fonctionnel connu à cette époque, semble responsable de la répression traductionnelle de l'ARN_m *lin-14*, en s'appariant de façon imparfaite à des sites multiples de sa 3'UTR.

Quatre ans plus tard, une deuxième cible de *lin-4* était identifiée : l'ARN_m de *lin-28* (un autre gène dont l'expression est finement régulée au cours du développement larvaire de *C. elegans*), dans la 3'UTR duquel un seul site de complémentarité imparfaite à l'ARN *lin-4* était identifié; ce site de complémentarité est nécessaire à la répression traductionnelle régulée temporellement de *lin-28* (Moss *et al.*, 1997). Puis, encore trois ans plus tard était décrit un deuxième petit ARN (de 21 nt de long), appelé *let-7*, et identifié à l'occasion d'un crible destiné à découvrir de nouveaux gènes hétérochroniques chez *C. elegans* (Reinhart *et al.*, 2000). Des sites de complémentarité imparfaite à l'ARN *let-7* ont été découverts dans les 3'UTR des ARN_m de cinq gènes hétérochroniques : *daf-12*, *lin-14*, *lin-28*, *lin-41* et *lin-42*. La fusion de la 3'UTR de l'ARN_m *lin-41* à un gène rapporteur confère à ce gène rapporteur un contrôle temporel similaire à celui du gène *lin-41* endogène, et cette régulation requiert les sites de complémentarité à *let-7*. En revanche, les 3'UTR des ARN_m *lin-14* et *lin-28* (celles de *daf-12* et *lin-42* n'ont pas été testées) ne semblent pas conférer de sensibilité à *let-7* : les sites de complémentarité des 3'UTR de ces deux gènes ne sont donc probablement pas fonctionnels (Reinhart *et al.*, 2000). De la même manière, un site de complémentarité imparfaite à *lin-4* a également été identifié dans la 3'UTR de l'ARN_m *lin-41*, mais *lin-4* ne semble pas réguler l'expression de *lin-41* (Slack *et al.*, 2000).

La même année, des orthologues convaincants de l'ARN *let-7* étaient détectés chez de nom-

¹⁵La séquence de *lin-4S* est celle des 22 nucléotides les plus en 5' de *lin-4L*; à la lumière des connaissances actuelles, *lin-4L* est certainement un précurseur de *lin-4S*, ce qui avait déjà été proposé par Lee *et al.*, 1993.



FIG. 1.12 – LIN-4, MEMBRE FONDATEUR DE LA FAMILLE DES MICROARN. Le précurseur lin-4L de lin-4 est un ARN susceptible de se replier en tige-boucle. Sa maturation libère le petit ARN lin-4, qui inhibe la traduction de gènes spécifiques, dont *lin-14*, sa première cible identifiée. En bas à droite : carte de la 3'UTR de *lin-14*, figure tirée de Wightman *et al.*, 1993. Les régions grisées sont les séquences de la 3'UTR de *lin-14* de *C. elegans* d'au moins 10 pb de long parfaitement conservées chez *C. briggsae* ; les régions colorées en noir bordé de gris sont les sept séquences conservées complémentaires à lin-4 (le détail en est donné en-dessous) ; les mutations qui empêchent, ou limitent, la répression traductionnelle de *lin-14* (ce sont des mutations gain-de-fonction de *lin-14*) sont indiquées (« n355 breakpoint », « n536 deletion »).

breuses espèces animales. Plus précisément : outre celui de *C. elegans*, le génome de *Drosophila melanogaster* contient un gène candidat, et celui de *Homo sapiens* en contient plusieurs, et tous sont imparfaitement autoccomplémentaires (des précurseurs plus longs que let-7 mature devraient donc se replier en tige-boucle, comme l'ARN lin-4L) ; et des ARN let-7 d'environ 21 nt sont exprimés dans de nombreux tissus adultes chez l'Homme, et à certains stades développementaux chez *D. melanogaster*, chez des Mollusques, une Annélide, le Poisson *Danio Rerio*, chez un Urochordé, un Hémichordé, et un Échinoderme (Pasquinelli *et al.*, 2000). Finalement, parmi tous les Métazoaires analysés, seuls les Cnidaires et les Porifères ne semblent pas exprimer let-7 ; les Plantes et les organismes unicellulaires ne possèdent pas non plus d'orthologue de let-7. Ce petit ARN serait donc apparu très tôt dans l'évolution (à la divergence entre Diploblastiques et Triploblastiques), et aurait conservé un rôle de contrôle du développement chez les Triploblastiques. En revanche, lin-4 semble restreint aux Nématodes.

La découverte du petit ARN let-7 dans une telle variété d'espèces, ainsi que la découverte accidentelle de molécules similaires chez la Drosophile, a incité plusieurs laboratoires à rechercher d'autres ARN du même type chez la Drosophile et dans des cellules humaines HeLa (Lagos-Quintana *et al.*, 2001), et chez *C. elegans* (Lau *et al.*, 2001 ; Lee et Ambros, 2001), par des approches de clonage extensif (le troisième article présente aussi l'identification de trois gènes de microARN — c'est le nom qui leur a été donné — par recherche bio-informatique de séquences conservées entre *C. elegans* et *C. briggsae*, et susceptibles de se replier en tiges-boucles de taille et de structure semblables à celles des précurseurs de lin-4 et let-7). Les trois groupes ont identifié, en tout, une centaine de microARN, dont certains étaient relativement conservés entre des espèces distantes, dont l'expression pouvait être régulée au cours du développement, et dont les gènes pouvaient être très proches les uns des autres, suggérant la co-transcription des précurseurs de ces microARN sous forme d'un transcrit primaire commun. Les précurseurs supposés de chacun de ces nouveaux microARN sont susceptibles de se replier en tiges-boucles.

Par la suite, d'autres travaux de clonage (Reinhart *et al.*, 2002 ; Lagos-Quintana *et al.*, 2002 ; Llave *et al.*, 2002a ; Mourelatos *et al.*, 2002 ; Ambros *et al.*, 2003 ; Aravin *et al.*, 2003 ; Dostie *et al.*, 2003 ; Lagos-Quintana *et al.*, 2003 ; Pfeffer *et al.*, 2004 ; Suh *et al.*, 2004 ; Wang *et al.*, 2004a) ou de prédiction *in silico* suivie de validation expérimentale (Grad *et al.*, 2003 ; Lai *et al.*, 2003 ; Lim *et al.*, 2003b ; Jones-Rhoades et Bartel, 2004), ont considérablement enrichi la liste de microARN, chez les Métazoaires comme chez des Métaphytes. Le site web du « miRNA registry » (URL : <http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml>) rassemble les séquences des microARN identifiés et de leurs précurseurs, leur assigne des noms officiels, et donne les références bibliographiques qui s'y rapportent (Griffiths-Jones, 2004). Il recense actuellement (sous la version 4.0) environ 1200 précurseurs différents (certains libèrent un microARN mature identique), dont par exemple 214 précurseurs de microARN murins.

Tous les microARN sont issus de la maturation de précurseurs de plus grande taille, capables de se replier en tiges-boucles imparfaites. Mais les déterminants précis des extrémités du microARN mature sont encore peu connus : le microARN peut provenir du brin 5', ou bien du brin 3' de la tige-boucle (et, dans de rares exemples, chaque brin libère un microARN mature), et la position du microARN mature, sur le précurseur, ne semble pas dépendre de la position des éventuels mésappariements de la tige (Lagos-Quintana *et al.*, 2001 ; Lau *et al.*, 2001 ; Lee et Ambros, 2001 ; voir Bartel, 2004, pour une revue). Il n'y a donc, à l'heure actuelle, aucun moyen de prédire la séquence du microARN issu de la maturation d'une tige-boucle donnée, ni même si cette tige-boucle est un authentique précurseur de microARN.

Plusieurs programmes informatiques ont été développés pour contourner cette limitation, et tâcher de détecter les gènes de microARN dans les séquences génomiques. Tous ces programmes limitent leurs recherches aux séquences phylogénétiquement conservées (en s'appuyant sur le principe que les séquences fonctionnellement importantes sont les plus conservées au cours de

l'évolution) : ils recensent les séquences susceptibles d'être transcrites en ARN qui se replieraient en tige-boucle de structure et de stabilité comparables à celles des précurseurs connus de microARN, et qui soient conservées phylogénétiquement (Ambros *et al.*, 2003 ; Grad *et al.*, 2003 ; Lai *et al.*, 2003 ; Lim *et al.*, 2003a ; Lim *et al.*, 2003b). Très récemment, une courte séquence consensus a été découverte, environ 200 pb en amont de la plupart des séquences de microARN du Nématode : cet octamère, riche en cytosines, joue probablement un rôle dans la transcription ou les étapes précoces de la maturation des microARN ; un programme de prédiction qui inclut cette nouvelle caractéristique des gènes de microARN s'avère plus efficace que les programmes précédents (Ohler *et al.*, 2004).

Les siRNA, des guides de clivage

En 1998, Andrew Fire et ses collaborateurs ont mis en évidence un phénomène de répression de gènes spécifiques suite à l'injection, dans le ver *Caenorhabditis elegans*, d'ARN double brin homologue (Fire *et al.*, 1998). Ils ont baptisé ce phénomène « interférence à ARN », ou : *RNA interference*, en anglais (abrégé en « RNAi »). Il est rapidement apparu que ce nouveau mode de répression de l'expression des gènes était très répandu chez les Eucaryotes, et qu'il expliquait de nombreuses observations, collectées chez les Animaux, les Végétaux et les Champignons (Cogoni et Macino, 2000). En particulier chez les Plantes, il avait été proposé dès 1994 que des ARN antisens seraient à l'origine de la répression post-transcriptionnelle de gènes spécifiques (Flavell, 1994 ; Baulcombe, 1996b), et des produits de clivage des ARN-cibles avaient été observés (Smith *et al.*, 1990 ; Goodwin *et al.*, 1996). Le phénomène de RNAi semble restreint aux Eucaryotes : un phénomène d'apparence similaire a été décrit chez l'Eubactérie *Escherichia coli* (dans laquelle de petits ARN guident le clivage d'ARN-cibles complémentaires), mais le mécanisme semble totalement différent (en particulier, le petit ARN est dégradé en même temps que son ARN-cible) (Massé *et al.*, 2003).

De petits ARN, de 21 à 25 nt, et dont la séquence est portée par l'un ou l'autre des deux brins de l'ARN déclencheur, sont systématiquement retrouvés dans le tissu ou l'échantillon affecté par le RNAi, chez les Plantes (Hamilton et Baulcombe, 1999), dans des embryons de Drosophile (Yang *et al.*, 2000) et dans un lysat d'embryons de Drosophile (Zamore *et al.*, 2000), dans des cellules de Drosophile en culture (Hammond *et al.*, 2000), et chez le Nématode (Parrish *et al.*, 2000). Ces petits ARN sont issus de la digestion endonucléolytique de l'ARN double brin déclencheur (Yang *et al.*, 2000 ; Zamore *et al.*, 2000 ; Parrish *et al.*, 2000) par une enzyme de type RNase III (Elbashir *et al.*, 2001a) : ils sont donc produits sous la forme de petits ARN double brin de 21 à 25 pb. Ces petits duplexes, baptisés « siRNA » (pour : *small interfering RNAs*, petits ARN interférents, en anglais), sont également fonctionnels s'ils sont synthétisés, et ajoutés au lysat d'embryon de Drosophile qui récapitule la réaction (au lieu d'être fournis sous forme de longs ARN double brin) (Elbashir *et al.*, 2001a). L'enzyme de type RNase III à l'origine de la production des siRNA a été identifiée en 2001, et baptisée « Dicer » (Bernstein *et al.*, 2001 ; Knight et Bass, 2001).

Depuis, de nombreuses études de génétique et de biochimie ont précisé la description de ce phénomène (voir « L'interférence à ARN », page 68). Parallèlement, des protocoles de répression de gènes spécifiques par des siRNA ont été mis au point, et le RNAi est devenu un outil très répandu, en particulier dans les systèmes mammifères, qui manquaient d'une méthode simple de répression de gènes spécifiques. De très nombreuses études ont utilisé des petits ARN double brin synthétiques dirigés contre un gène d'intérêt (voir Dorsett et Tuschl, 2004, pour revue).

Similitudes entre microARN et siRNA

Les siRNA ont d'abord été décrits comme des guides de clivage d'ARN-cibles, dans les processus d'interférence à ARN. Cette définition est donc fonctionnelle, alors que la définition

des microARN (*cf* page 59) est une description structurale (les microARN sont de petits ARN simple brin de 21 à 23 nt issus de la maturation de précurseurs repliés en tiges-boucles). La découverte que des microARN pouvaient guider des clivages par interférence à ARN, et que des molécules *a priori* destinées à se comporter en siRNA pouvaient réprimer la traduction d'ARN_m-cibles a mis en évidence cette difficulté : certaines molécules, maturées comme des microARN à partir de précurseurs endogènes, sont capables de guider l'interférence à ARN, comme des siRNA. Ces deux familles ne sont donc pas disjointes.

La ressemblance structurale entre les microARN (qui, à l'époque, n'étaient qu'au nombre de deux, *lin-4* et *let-7*) et les petits ARN simple brin qui constituent les siRNA (longueur similaire, de 21 à 23 nt, extrémité 5' phosphorylée, extrémité 3' non phosphorylée) a amené trois équipes à se demander si l'enzyme Dicer, responsable de la maturation des siRNA, pouvait être aussi impliquée dans celle des microARN. Leurs résultats ont montré que chez le Nématode *C. elegans* et dans les cellules humaines HeLa, Dicer est l'endonucléase qui libère *let-7* (et, chez le Nématode, *lin-4*) de son précurseur immédiat, un ARN d'environ 70 nt replié en tige-boucle (Grishok *et al.*, 2001 ; Hutvagner *et al.*, 2001 ; Ketting *et al.*, 2001). La protéine Dicer de *Drosophila* est elle-même capable de maturer correctement *let-7* (Ketting *et al.*, 2001). La maturation de précurseurs d'environ 70 nt par Dicer semble être une généralité, puisque plusieurs autres précurseurs de microARN s'accumulent en l'absence de Dicer (Lee et Ambros, 2001 ; Grad *et al.*, 2003).

Ainsi, microARN et siRNA sont issus de la digestion d'ARN bicaténaires par la même enzyme, Dicer ; alors que les précurseurs de microARN sont monomoléculaires (la structure bicaténaire est due à un repliement de la molécule de précurseur), la plupart des précurseurs de siRNA sont bimoléculaires.

La découverte, en 2002, que le microARN *let-7* était capable de guider le clivage d'un ARN-cible, pourvu que ce dernier lui soit parfaitement complémentaire (Hutvagner et Zamore, 2002), a montré que les microARN pouvaient se comporter en siRNA ; notamment, *in vivo* dans les cellules humaines HeLa, le microARN *let-7* est incorporé dans le complexe RISC (pour : *RNA induced silencing complex*, le complexe effecteur du RNAi : Hammond *et al.*, 2000), puisqu'il suffit d'incuber un ARN présentant une complémentarité parfaite à *let-7* dans un extrait cytoplasmique S100 de cellules HeLa pour que l'ARN incubé soit clivé. Ainsi, *in vivo*, le microARN *let-7* s'engage dans la voie du RNAi, et il ne semble diriger le clivage d'aucun ARN-cible, simplement parce qu'il n'existe pas dans la cellule d'ARN qui lui présente une complémentarité parfaite. Chez les Plantes, en revanche, de nombreux exemples de clivages d'ARN endogènes dirigés par des microARN ont été relevés (Llave *et al.*, 2002b ; Kasschau *et al.*, 2003 ; Palatnik *et al.*, 2003 ; Floyd et Bowman, 2004 ; Mallory *et al.*, 2004 ; Vazquez *et al.*, 2004). Une autre équipe a montré que la cotransfection (dans des cellules humaines en culture) d'un gène de microARN et d'un gène rapporteur présentant une complémentarité parfaite à ce microARN conduisait à la dégradation de l'ARN messager du gène rapporteur (Zeng *et al.*, 2003).

Réciproquement, les siRNA guident la répression traductionnelle d'ARN-cibles qui présentent une complémentarité imparfaite à l'un des deux brins du siRNA (Doench *et al.*, 2003 ; Zeng *et al.*, 2003). Finalement, les microARN et les brins de siRNA semblent pouvoir jouer les mêmes rôles, et seule la qualité de l'appariement entre le petit ARN et sa cible déterminerait le mode de répression post-transcriptionnelle : si la complémentarité est parfaite, l'ARN-cible est clivé par RNAi, et si elle est imparfaite, sa traduction est inhibée, sans que sa stabilité soit affectée (voir page 72).

D'autre part, les complexes effecteurs impliqués dans les deux modes de répression (le complexe RISC, responsable du RNAi, et le complexe répresseur de la traduction¹⁶) partagent plusieurs points communs : ils sont tous deux associés aux polyribosomes (Ishizuka *et al.*, 2002 ;

¹⁶Bien qu'il n'ait pas encore été identifié avec certitude (par exemple, aucun système *in vitro* ne reproduit

Djikeng *et al.*, 2003 ; Pham *et al.*, 2004 ; Nelson *et al.*, 2004), contiennent au moins trois constituants communs chez l'Homme, eIF2C2, Gemin3 et Gemin4 (Hutvagner et Zamore, 2002 ; Martinez *et al.*, 2002 ; Mourelatos *et al.*, 2002), et sont tous les deux assemblés autour d'un petit ARN simple brin d'environ 22 nt de long (Martinez *et al.*, 2002 ; Mourelatos *et al.*, 2002). De plus, puisque le microARN let-7, connu pour participer à un guidage de répression traductionnelle (chez le Nématode), est incorporé dans RISC *in vivo* dans des cellules de Mammifères (Hutvagner et Zamore, 2002), alors qu'il ne semble impliqué dans aucun guidage de clivage *in vivo*, peut-être son association à RISC signifie-t-elle que le complexe médiateur de la répression traductionnelle guidée par let-7 est RISC lui-même.

Il est donc possible qu'il n'existe aucune différence biochimique entre les complexes effecteurs de ces deux voies. Cependant, dans les cellules S2 de Drosophile, la plupart des microARN sont retrouvés dans un complexe plus lourd que le complexe qui contient les siRNA (Caudy *et al.*, 2002), et des protéines spécifiques semblent associées à ces deux types de complexe : AGO1 est principalement retrouvée dans le complexe léger, contenant la plupart des microARN, alors que AGO2 est principalement retrouvée dans le complexe lourd, en association avec les siRNA. Mais surtout, il a été montré récemment que la protéine AGO2 de Drosophile est nécessaire au clivage guidé par les siRNA, alors qu'elle est dispensable pour le clivage guidé par des microARN (Okamura *et al.*, 2004). Chez les Mammifères, de lourdes présomptions laissent penser que la protéine appelée Ago2 porte l'activité endonucléasique du complexe RISC assemblé autour des siRNA (voir « L'interférence à ARN », page 68).

Ainsi, il semblerait que les microARN et les siRNA parfaitement complémentaires d'un ARN-cible en guident le clivage à la même position (la liaison phosphodiester clivée est celle qui relie les nucléotides appariés aux 10ème et 11ème nucléotides du petit ARN, à partir du 5' du petit ARN) (Elbashir *et al.*, 2001a ; Elbashir *et al.*, 2001b ; Xie *et al.*, 2003 ; Yekta *et al.*, 2004), mais que ce clivage est catalysé par des enzymes différentes chez la Drosophile.

Dans la suite de cette section, les données présentées ne concerneront que les Mammifères, sauf mention contraire. Les Mammifères ne possèdent qu'une enzyme Dicer et ne semblent pas avoir de RdRP (polymérase à ARN dépendant de l'ARN). Cette dernière particularité (qui les distingue par exemple du Nématode, des Plantes et des Champignons) explique probablement l'absence du phénomène de « transitivity » du RNAi chez les Mammifères¹⁷, et peut-être également l'absence de diffusion systémique du RNAi¹⁸, puisque l'action de la RdRP aboutit à une amplification des siRNA, probablement nécessaire à la propagation du RNAi dans tout l'organisme.

l'activité de répression traductionnelle guidée par un petit ARN), il semble que le complexe répresseur de la traduction soit le complexe purifié en 2002 par Mourelatos *et al.* (Nelson *et al.*, 2004) : voir page 74.

¹⁷Chez le Nématode, une RdRP polymérise un ARN complémentaire des ARN-cibles hybridés à un brin de siRNA. L'ARN double brin issu de cette polymérisation (qui s'étend en direction du 5' de l'ARN-cible) devient lui-même un ARN déclencheur de RNAi, clivé en siRNA. La polymérisation peut donc s'amorcer sur des ARN matrices différents de l'ARN déclencheur, mais qui partagent avec lui des régions de similarité, suffisantes pour permettre l'hybridation du brin de siRNA (Sijen *et al.*, 2001) ; ce phénomène a été baptisé « transitivity du RNAi ». Chez les Plantes, un phénomène similaire permet l'extension du RNAi contre des séquences initialement absentes de l'ARN déclencheur (Vaistij *et al.*, 2002), mais le mécanisme en apparaît différent, puisque le RNAi se propage à la fois en direction du 5' et du 3' des ARN-cibles (Voinnet *et al.*, 1998 ; Vaistij *et al.*, 2002), et qu'il semble restreint à certains gènes (Vaistij *et al.*, 2002).

¹⁸Chez le Nématode (Fire *et al.*, 1998) et chez les Plantes (Palauqui *et al.*, 1996 ; Voinnet *et al.*, 1998), le RNAi initié dans un organe donné se propage ensuite dans le reste de l'organisme, alors qu'il semble rester confiné dans les cellules contenant l'ARN déclencheur chez les Mammifères.

1.4.2 Biogenèse des microARN et siRNA

La production des microARN chez les Mammifères

Les microARN sont transcrits sous forme de longs précurseurs, qui peuvent atteindre plusieurs centaines de nucléotides en longueur, et éventuellement contenir les séquences de plusieurs microARN différents (Lee *et al.*, 2002b). L'enzyme responsable de leur transcription est l'ARN-polymérase II (Lee *et al.*, 2004a). Ces transcrits primaires, appelés « pri-miRNA », sont ensuite maturés par la RNase III Drosha pour libérer les « pre-miRNA », tiges-boucles d'environ 70 nt de long (Lee *et al.*, 2002b ; Lee *et al.*, 2003b). Les pre-miRNA sont alors digérés par la RNase III Dicer, dont les produits stables sont les microARN matures d'environ 21 à 23 nt (Grishok *et al.*, 2001 ; Hutvagner *et al.*, 2001 ; Ketting *et al.*, 2001).

La localisation subcellulaire de ces réactions semblait bien établie : l'enzyme Drosha s'accumule dans le noyau (Lee *et al.*, 2003b), compartiment où les pri-miRNA sont convertis en pre-miRNA (Lee *et al.*, 2002b). Les pre-miRNA sont exportés dans le cytoplasme par l'exportine 5 (Yi *et al.*, 2003 ; Lund *et al.*, 2004), et y sont maturés en microARN matures (Lee *et al.*, 2002b) par l'enzyme Dicer, qui semble exclusivement cytoplasmique (Billy *et al.*, 2001). Notamment, la maturation *in vitro* de précurseurs de microARN s'interrompt au stade « pre-miRNA » si un extrait cytoplasmique n'est pas ajouté à la réaction (Lee *et al.*, 2002b).

Toutefois, quelques travaux semblent indiquer que la maturation intégrale des microARN pouvait avoir lieu dans le noyau, au moins dans les systèmes mammifères :

- des microARN matures peuvent être détectés dans le noyau (Zeng et Cullen, 2003 ; Meister *et al.*, 2004) ; ils pourraient cependant avoir été maturés dans le cytoplasme, puis importés dans le noyau ;
- un long ARN en tige-boucle incapable de sortir du noyau, car dépourvu de coiffe, de queue poly(A), et d'intron, est mûré malgré tout en petits ARN d'environ 21 à 22 nt de long (Shinagawa et Ishii, 2003).

Il est remarquable que les deux enzymes responsables de la maturation des microARN, Drosha et Dicer, appartiennent à la même famille, celle des RNases III. Les substrats de ces enzymes sont des ARN double brin, et leurs produits, des ARN double brin aux extrémités 3' sortantes (de 2 nt), non phosphorylées, et aux extrémités 5' phosphorylées. Cette caractéristique pourrait constituer un point de contrôle de la qualité des petits ARN : l'exportine 5 présente une affinité pour les ARN double brin aux extrémités 3' sortantes très supérieure à son affinité pour les autres ARN double brin (Gwizdek *et al.*, 2003 ; Lund *et al.*, 2004) ; puis, la structure particulière des produits de Drosha pourrait faciliter leur prise en charge par Dicer¹⁹ ; enfin, les protéines Argonaute du complexe effecteur (RISC, ou le complexe répresseur de la traduction) possèdent toutes le domaine PAZ, qui interagit préférentiellement avec les petits ARN double brin dont l'extrémité 3' est sortante (Lingel *et al.*, 2003 ; Song *et al.*, 2003 ; Yan *et al.*, 2003). Ainsi, l'extrémité caractéristique des produits de clivage des RNases III serait un indicateur de la qualité des différentes étapes de la maturation de l'ARN double brin, lui permettant d'accéder à l'étape suivante (Carmell et Hannon, 2004).

Finalement, le produit de Dicer est un petit ARN double brin, similaire à un siRNA ; or, à quelques exceptions près, un seul brin s'accumule sous la forme d'un microARN mature. Il a été remarqué que le brin dont l'extrémité 5' est la moins stablement appariée, dans ce duplex, semble être systématiquement celui qui s'accumulera (Khvorova *et al.*, 2003 ; Schwarz *et al.*, 2003). Il est donc vraisemblable qu'une hélicase à ARN prenne en charge le produit (bicaténaire) de Dicer, et

¹⁹En effet, les produits de Dicer sont plus abondants, et leurs tailles sont plus homogènes, si le substrat de Dicer était un pre-miRNA, produit de Drosha, plutôt qu'une tige-boucle aux extrémités quelconques (Lee *et al.*, 2003b). Dicer cliverait donc les pre-miRNA à environ 22 pb du site de coupure de Drosha sur les pre-miRNA (voir figure 1.13), alors qu'elle libérerait des produits d'environ 22 pb de n'importe quelle position d'un pri-miRNA (Carmell et Hannon, 2004) (la séquence ou la structure secondaire du substrat au voisinage des sites de clivage pourrait alors altérer le choix de la liaison phosphodiester hydrolysée, donc la taille des produits de Dicer).

qu'elle engage l'extrémité 5' d'un des deux brins dans un complexe ribonucléoprotéique qui, par la suite, protégera ce brin contre les nucléases cellulaires ; l'autre brin, non protégé, serait alors rapidement dégradé. Dans ce cas, la facilité avec laquelle cette hélicase ouvrirait l'extrémité 5' d'un brin donné conditionnerait la stabilité de ce brin. La même sélection semble s'opérer sur les brins des siRNA : le brin qui guide le RNAi le plus efficacement est le moins stablement apparié sur son extrémité 5' dans le siRNA (Aza-Blanc *et al.*, 2003 ; Khvorova *et al.*, 2003 ; Schwarz *et al.*, 2003).

La figure 1.13 résume le modèle actuel de biogenèse des microARN.

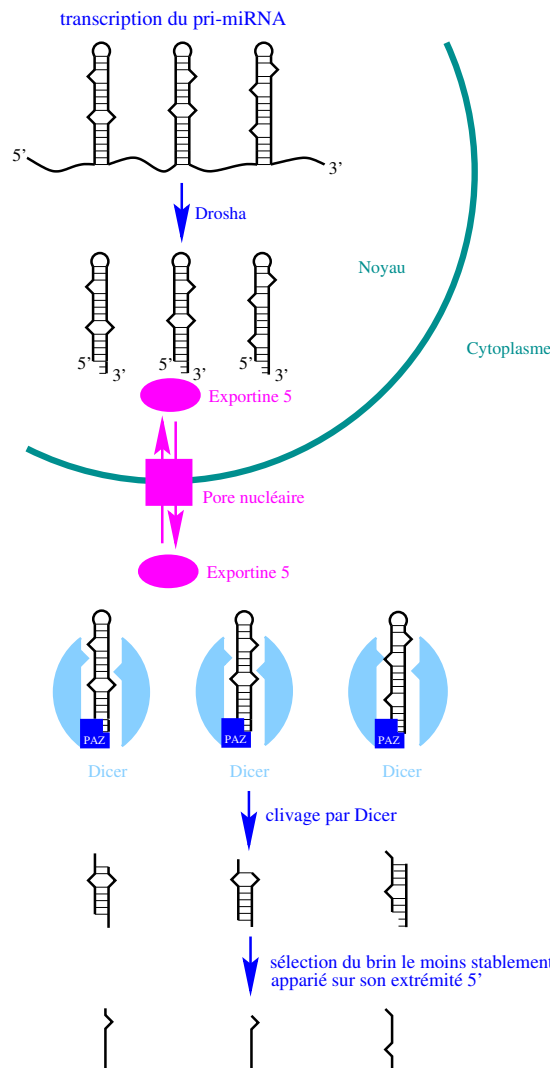


FIG. 1.13 – MODÈLE DE MATURATION DES MICROARN DE MAMMIFÈRES. Voir texte pour le détail.

Biogenèse des siRNA

Les siRNA proviennent de la digestion de l'ARN double brin déclencheur du RNAi (Parrish *et al.*, 2000 ; Yang *et al.*, 2000 ; Zamore *et al.*, 2000), par Dicer (Bernstein *et al.*, 2001 ; Knight et Bass, 2001). La protéine Dicer humaine, qui possède un domaine RBD (domaine de liaison à l'ARN double brin) et un domaine PAZ, qui interagit préférentiellement avec les ARN double brin aux extrémités 3' sortantes (Lingel *et al.*, 2003 ; Song *et al.*, 2003 ; Yan *et al.*, 2003), peut lier les ARN double brin *in vitro* et les cliver en petits ARN d'environ 22 nt de long (Provost *et al.*, 2002a ; Zhang *et al.*, 2002). L'enzyme clive préférentiellement le double brin à ≈ 22 nt de son extrémité ; le site de coupure, qui devient une nouvelle extrémité, au 3' sortant de 2 nt, peut alors servir de nouveau point d'ancrage pour Dicer, dont le prochain clivage se trouvera ≈ 22 nt en aval, et ainsi de suite : l'enzyme digère l'ARN double brin de manière processive, par pas d'environ 22 nt (Elbashir *et al.*, 2001a ; Zhang *et al.*, 2002).

La comparaison de Dicer humaine avec des RNases III bactériennes, et l'analyse des propriétés catalytiques *in vitro* de formes mutées de Dicer humaine, ont amené le groupe de W. Filipowicz à proposer un modèle d'interaction de Dicer avec son substrat (Zhang *et al.*, 2004). Dans ce modèle, chacun des deux domaines RNase III de Dicer contribue pour moitié à la catalyse : chacun hydrolyse une liaison phosphodiester sur un brin du substrat (voir figure 1.14).

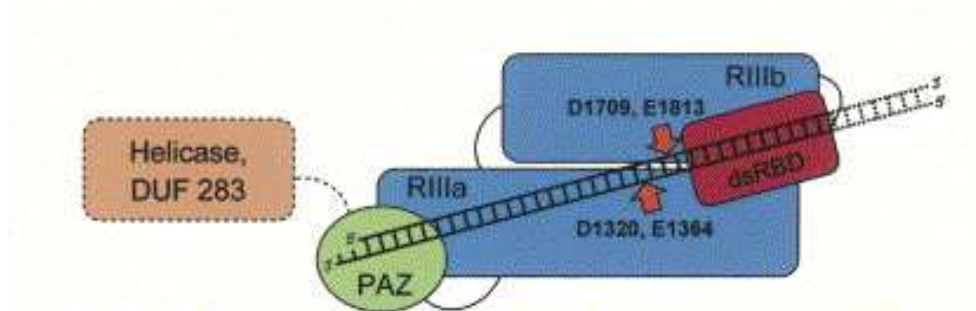


FIG. 1.14 – MODÈLE D'INTERACTION DE DICER AVEC SON SUBSTRAT. Figure tirée de Zhang *et al.*, 2004. La fonction des domaines hélicase et DUF 283 (en brun) reste inconnue : leurs contours sont dessinés en pointillés. Le domaine PAZ (en vert), qui interagit préférentiellement avec les extrémités au 3' sortant (Lingel *et al.*, 2003 ; Song *et al.*, 2003 ; Yan *et al.*, 2003), participe probablement à l'interaction avec le substrat ; avec le domaine RNase IIIa (noté : « RIIa »), il positionnerait le substrat dans l'enzyme, de sorte que les clivages aient lieu à ≈ 22 pb de l'extrémité du double brin. Le domaine RBD (en rouge sombre), d'interaction avec les ARN double brin, stabiliserait le complexe entre l'enzyme et son substrat. Chacun des deux domaines RNase III (en bleu) clive un brin du substrat (flèches rouges) ; l'hydrolyse est catalysée par un résidu acide (leur nature et position sont indiquées). Le domaine RNase IIIa (qui clive le brin dont l'extrémité 3' est libre) est plus long que celui de la plupart des enzymes de la famille des RNases III : sur le schéma, il est représenté par un rectangle plus grand que celui du domaine RNase IIIb.

La protéine mutante sur le domaine PAZ (dont l'activité de digestion des ARN double brin est réduite, probablement en raison de son affinité réduite pour le substrat) libère un siRNA de même taille que celui produit par la protéine sauvage, par un clivage à ≈ 22 nt de l'extrémité du substrat : le domaine PAZ est donc dispensable pour le positionnement du substrat dans l'enzyme, qui détermine la taille des siRNA produits. Le domaine RNase IIIa pourrait participer au positionnement du substrat, par sa partie N-terminale, nettement plus longue que celle de la plupart des domaines RNase III (Zhang *et al.*, 2004).

In vivo, la protéine Dicer de Mammifères est associée avec des partenaires de petite taille, pour former un complexe légèrement plus volumineux que Dicer seule (Zhang *et al.*, 2004). Ses partenaires ne sont pas connus chez les Mammifères, mais chez la Drosophile, plusieurs autres protéines sont requises pour l'assemblage du complexe RISC : les protéines « aubergine » et « armitage » (Tomari *et al.*, 2004), et la protéine « R2D2 », qui interagit avec Dicer, et la stabilise biochimiquement (Liu *et al.*, 2003). Dicer et R2D2 forment notamment un complexe avec les siRNA, qui semble être le précurseur immédiat de RISC (Tomari *et al.*, 2004).

Le gène *aubergine* appartient à la famille Argonaute, dont la plupart des membres sont impliqués dans les phénomènes de RNAi, et qui possèdent tous le domaine PAZ (voir Carmell *et al.*, 2002, pour revue). *Armitage* code une hélicase à ARN putative, et son orthologue chez *Arabidopsis thaliana*, *SDE3*, est nécessaire au RNAi (Cook *et al.*, 2004). *R2D2* est également apparenté à un gène requis pour le RNAi dans une autre espèce : son orthologue chez le Nématode *Caenorhabditis elegans*, *RDE-4*, avait lui aussi été identifié par un crible génétique visant à identifier des mutations du processus de RNAi ; les protéines R2D2 et RDE-4 possèdent chacune deux domaines putatifs de liaison à l'ARN double brin (Liu *et al.*, 2003).

1.4.3 Modes d'action

MicroARN et siRNA participent à deux modes de répression post-transcriptionnelle de gènes spécifiques : le RNAi (une dégradation des ARN qui présentent une complémentarité parfaite à un microARN, ou à l'un des deux brins d'un siRNA) et une répression traductionnelle, guidée par un microARN ou un brin de siRNA imparfaitement complémentaire à une séquence de l'ARN_m-cible. Chacun des deux types de petit ARN peut participer à chacun de ces deux types de répression post-transcriptionnelle, en fonction de la qualité de sa complémentarité à l'ARN-cible (voir « Similitudes entre microARN et siRNA », page 62).

L'interférence à ARN

Le rôle des siRNA dans la réaction de RNAi a été interprété de trois façons : premièrement, un brin du siRNA pourrait s'hybrider avec l'ARN-cible, et servir d'amorce pour la polymérisation d'un deuxième brin d'ARN sur l'ARN-cible, par une polymérase d'ARN dépendant de l'ARN (RdRP, pour *RNA-dépendant RNA polymerase*) ; ce long ARN double brin serait ensuite clivé par Dicer, ce qui régénérerait une grande quantité de petits ARN susceptibles d'amorcer ensuite de nouvelles polymérisations, et qui détruirait simultanément l'ARN-cible (Lipardi *et al.*, 2001). Deuxièmement, les brins antisens²⁰ de nombreux siRNA pourraient s'hybrider tout le long de l'ARN-cible, et seraient ligués entre eux par une ligase cellulaire ; l'ARN double brin résultant serait, comme dans le premier modèle, dégradé par l'enzyme Dicer (Lipardi *et al.*, 2001 ; Nishikura, 2001). Troisièmement, les siRNA seraient des guides de clivage, ils seraient incorporés (sous forme de simple ou de double brin) dans un complexe endonucléolytique qui cliverait les ARN qui s'hybrident à l'un des deux brins du siRNA (Hammond *et al.*, 2000 ; Zamore *et al.*, 2000 ; Nykänen *et al.*, 2001). Contrairement à ce qui fut publié dans un premier temps (Lipardi *et al.*, 2001), des siRNA dont les extrémités 3', modifiées chimiquement, qui semblent incapables de servir d'amorce à une polymérase (premier modèle) ou de substrat à une ligase (deuxième modèle), peuvent quand même participer à une réaction de RNAi dans des cellules humaines et dans un lysat d'embryon de Drosophile (Schwarz *et al.*, 2002 ; Martinez *et al.*, 2002). Il semble donc que les siRNA soient des guides de clivage chez ces deux organismes.

Le mécanisme de ce guidage posait également problème : dans l'hypothèse la plus simple, le complexe effecteur RISC contient un seul brin du siRNA (deux RISC différents pourraient donc

²⁰Convention : on appelle « brin sens » le brin d'ARN du siRNA de séquence identique à la séquence portée par l'ARN-cible, et « brin antisens », le brin du siRNA de séquence complémentaire à la séquence de l'ARN-cible.

coexister en présence d'un siRNA donné : celui qui contient le brin sens, et celui qui contient le brin antisens, du siRNA), et si un ARN s'hybride à ce petit ARN simple brin, il est clivé (Zamore *et al.*, 2000). Un tel modèle expliquerait facilement pourquoi certaines modifications chimiques du siRNA affectent l'efficacité de l'interférence à ARN si elles sont portées par le brin antisens, alors qu'elles n'ont pas d'effet si elles sont portées par le brin sens (Parrish *et al.*, 2000).

Mais un petit ARN simple brin, antisens d'un ARN-cible, est beaucoup moins efficace qu'un petit ARN double brin (voire totalement inefficace) pour guider l'interférence à ARN (Yang *et al.*, 2000 ; Zamore *et al.*, 2000 ; Parrish *et al.*, 2000 ; Schwarz *et al.*, 2002) ; de surcroît, un ARN simple brin partiellement complémentaire du brin sens (tout comme un ARN simple brin partiellement complémentaire du brin antisens) inhibe la réaction d'interférence à ARN, comme si le complexe RISC contenait les deux brins du siRNA, et que l'hybridation de l'un des deux, quel qu'il soit, avec un ARN empêchait le clivage de l'ARN-cible (Yang *et al.*, 2000). Enfin, un siRNA double brin peut guider un clivage beaucoup plus efficace sur une cible donnée, que sur l'antisens de cette cible (et généralement, la cible clivée le plus efficacement est complémentaire du brin du siRNA dont l'extrémité 3' est la plus proche de l'extrémité de l'ARN double brin déclencheur) ; pour expliquer cette observation, un modèle a été proposé, dans lequel les deux brins du siRNA gardent une trace de leur orientation, en s'associant chacun à des protéines spécifiques. Le complexe RISC contiendrait alors les deux brins du siRNA, mais un seul (celui associé aux protéines qui le permettent) serait capable de guider le clivage d'un ARN-cible (Elbashir *et al.*, 2001a). Ces différents modèles sont présentés en figure 1.15, page 70.

La purification du complexe RISC, et une expérience de reconstitution à partir d'un ARN étiqueté à la biotine, ont finalement démontré que le complexe ne contient qu'un seul brin du siRNA (Martinez *et al.*, 2002) ; de plus, un petit ARN simple brin peut guider un clivage d'ARN-cible très similaire au clivage guidé par un siRNA double brin, en terme de position du site de clivage, et de conditions influençant l'efficacité de la réaction (Holen *et al.*, 2003). Les observations précédentes, en contradiction apparente avec ces résultats (*cf* ci-dessus), peuvent alors s'interpréter de la manière suivante :

- un siRNA double brin serait plus efficace qu'un siRNA simple brin pour guider le RNAi, parce que des systèmes spécialisés pourraient coupler l'ouverture des deux brins du siRNA avec l'entrée d'un des deux dans le complexe RISC²¹ ; alternativement, ou simultanément, la plus grande sensibilité des petits ARN simple brin aux nucléases pourrait justifier de leur faible activité : ils seraient dégradés très rapidement (en particulier dans le lysat d'embryon de *Drosophile*), ce qui limiterait leur activité apparente (Martinez *et al.*, 2002) ;
- un ARN complémentaire du brin sens d'un siRNA est un compétiteur du brin antisens, et sa présence à haute concentration peut faire disparaître le siRNA double brin, au profit d'un duplex entre le brin sens et son ARN complémentaire ; le brin antisens, dont une proportion notable serait en simple brin, serait alors peu incorporé dans le complexe RISC (*cf* ci-dessus) ;
- les deux brins d'un siRNA ne sont pas égaux devant le RNAi : il semble que le brin dont l'extrémité 5' est la plus faiblement appariée entre plus facilement dans le complexe RISC (Aza-Blanc *et al.*, 2003 ; Khvorova *et al.*, 2003 ; Schwarz *et al.*, 2003) ; ainsi, un ARN-cible peut être clivé plus efficacement que son ARN antisens en présence d'un siRNA donné, si le brin antisens à cet ARN-cible est apparié moins stablement sur son extrémité 5' que le brin sens (voir « La production des microARN chez les Mammifères », page 65).

²¹Par exemple, chez la *Drosophile*, une isoforme de Dicer participe à l'assemblage de RISC, en interagissant notamment avec le siRNA (Lee *et al.*, 2004b ; Pham *et al.*, 2004). La situation pourrait être différente dans d'autres systèmes, puisque l'enzyme Dicer humaine recombinante ne semble pas interagir avec les siRNA (Provost *et al.*, 2002a).

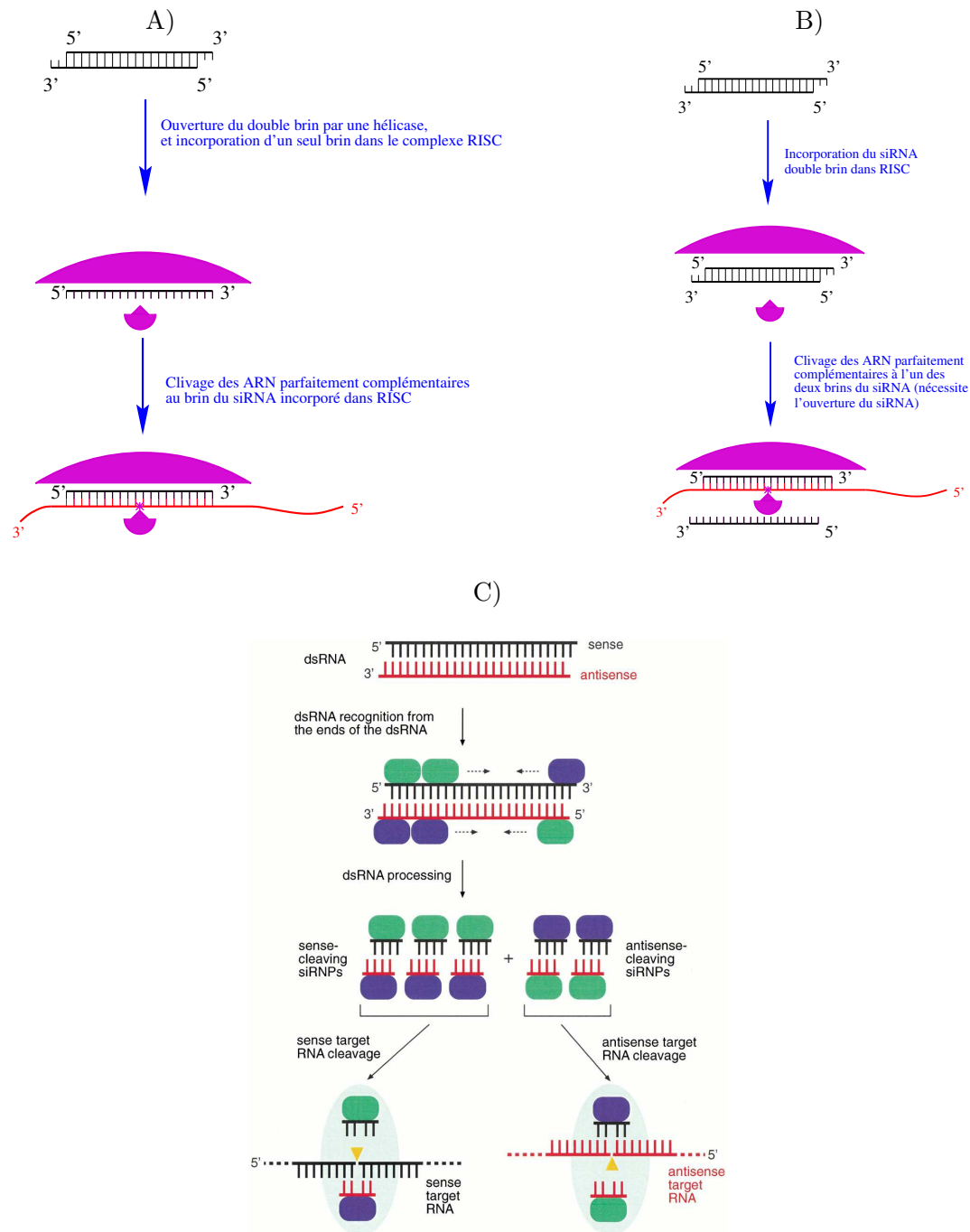


FIG. 1.15 – MODÈLES HISTORIQUES DE FONCTIONNEMENT DE L'INTERFÉRENCE À ARN. A) Le complexe RISC ne contient qu'un brin du siRNA, et clive les ARN qui lui sont complémentaires. B) Le complexe RISC contient le siRNA double brin, et clive les ARN complémentaires de l'un des deux brins ; la reconnaissance de l'ARN-cible nécessite donc un réarrangement du complexe, avec ouverture du siRNA. C) Figure tirée de Elbashir *et al.*, 2001a ; le complexe RISC contient le siRNA double brin, et seul le brin dont l'extrémité 3' était la plus proche de l'extrémité de l'ARN double brin déclencheur peut guider un clivage, car associé à une protéine (représentée en bleu sur le schéma) qui le permet. Cette particularité pourrait être due au mode de synthèse des siRNA : lorsque Dicer clive l'ARN déclencheur, elle le digère de façon processive, à partir des extrémités, et pourrait donc laisser des protéines spécifiques (représentées en bleu sur le schéma) sur le brin dont le 3' était le plus proche de l'extrémité de l'ARN déclencheur.

Plusieurs protéines du complexe RISC ont ensuite été identifiées, par des approches de biochimie :

1. Un complexe RISC d'environ 500 kDa a été purifié à partir de cellules S2 de *Drosophile* en culture (Hammond *et al.*, 2001); il contient les protéines suivantes :
 - Argonaute 2 (AGO2), qui contient le domaine PAZ (*cf* page 65); AGO2 est nécessaire au clivage par RNAi guidé par des siRNA, mais pas par des microARN (Okamura *et al.*, 2004); une autre protéine de la famille Argonaute, la protéine Ago2 de Mammifère, porte l'activité endonucléasique du complexe RISC (Liu *et al.*, 2004a; Song *et al.*, 2004);
 - VIG (pour : *Vasa intronic gene*, son gène se situe dans un intron du gène *Vasa*), de fonction inconnue, mais qui contient le motif RGG, un motif de liaison à l'ARN (Caudy *et al.*, 2002); cette protéine est également enrichie dans un complexe RISC de 250 kDa chez le Nématode (Caudy *et al.*, 2003);
 - dFXR, l'homologue de la protéine humaine FMRP (protéine du syndrome de l' X fragile), qui contient deux domaines KH et un domaine RGG, qui semblent tous les trois lier l'ARN (Caudy *et al.*, 2002; Ishizuka *et al.*, 2002);
 - Tudor-SN, une protéine de la famille des nucléases staphylococcales; elle présente une activité de nucléase, et semble associée à RISC également chez le Nématode, et dans des cellules humaines HeLa (Caudy *et al.*, 2003); puisque les enzymes de sa famille sont des exonucléases, Tudor-SN n'est probablement pas responsable de l'activité endonucléolytique de RISC, mais elle pourrait participer à la dégradation exonucléolytique des produits de clivage (Caudy *et al.*, 2003; Schwarz *et al.*, 2004).
2. Une équipe indépendante a démontré que, dans les cellules S2 également, l'hélicase à ARN Dmp68 était associée au complexe contenant dFXR, AGO2 et des petits ARN (Ishizuka *et al.*, 2002).
3. Un complexe RISC d'environ 160 kDa a été purifié à partir de cellules humaines HeLa (et le même protocole, appliqué à une préparation de cellules S2 de *Drosophile* permet la purification d'un complexe de taille voisine, donc nettement inférieure aux 500 kDa du complexe présenté ci-dessus, et aux 250 kDa d'un complexe purifié précédemment : Nykänen *et al.*, 2001; pourtant, le complexe humain de ≈ 160 kDa a conservé son activité de clivage endonucléolytique guidé par un petit ARN); ce complexe contient les protéines eIF2C1 et eIF2C2/GERp95, deux protéines similaires de la famille Argonaute (voir page 68).
4. À partir de cellules humaines HeLa, la protéine eIF2C2 et de petits ARN sont immunoprécipités avec les protéines Gemin3 et Gemin4, connues pour interagir avec de nombreuses particules ribonucléoprotéiques (Hutvagner et Zamore, 2002; Mourelatos *et al.*, 2002).

Des travaux très récents ont probablement permis d'identifier le constituant du complexe RISC qui porte l'activité endonucléasique, chez les Mammifères (ce constituant non identifié avait été baptisé « Slicer ») : la protéine Ago2, qui est systématiquement associée aux complexes compétents pour le clivage, à la différence des autres protéines Ago de Mammifères (Meister *et al.*, 2004), possède en outre un domaine très semblable au site catalytique de la RNase H, dont des mutations ponctuelles abolissent l'activité de RISC (Liu *et al.*, 2004a; Song *et al.*, 2004), et il est possible de reconstituer *in vitro* un complexe RISC minimal, actif, à partir d'un ARN simple brin de 21 nt et d'Ago2 purifiée²² (Liu *et al.*, 2004a).

Les ARN courts et le guidage de la répression traductionnelle

Les deux premiers microARN découverts, *lin-4* et *let-7*, inhibent la traduction d'ARN_m-cibles sans en affecter notablement la stabilité (Wightman *et al.*, 1993; Moss *et al.*, 1997; Reinhart

²²La protéine a été purifiée par immunoaffinité, et est contaminée par la protéine HSP90.

et al., 2000) (voir aussi Historique, page 59). Les 3'UTR de ces ARN-cibles présentent des complémentarités imparfaites aux microARN répresseurs, et des délétions qui suppriment une partie ou la totalité de ces sites de complémentarité abolissent la répression traductionnelle guidée par le microARN (Wightman *et al.*, 1991 ; Wightman *et al.*, 1993) ; réciproquement, la fusion de la 3'UTR à un gène rapporteur lui confère une répression, régulée au cours du développement, similaire à celle du gène-cible : le rapporteur est donc probablement devenu sensible au microARN (Wightman *et al.*, 1993 ; Reinhart *et al.*, 2000). Ainsi, les microARN répriment visiblement la traduction²³ d'ARN_m-cibles en s'appariant à leurs 3'UTR ; cet appariement ne serait pas parfait, des nucléotides non appariés (sur le microARN, ou sur l'ARN_m) interrompent les segments double brin (voir par exemple la figure 1.12, page 60). Effectivement, les microARN *let-7* et *lin-4* interagissent *in vitro* avec les ARN_m de *lin-14* et *lin-41*, respectivement (Ha *et al.*, 1996 ; Vella *et al.*, 2004).

Les deux modes de répression post-transcriptionnelle, RNAi et répression traductionnelle, sont alternatifs, et il semble que seule la qualité de l'appariement entre les deux ARN détermine l'action du microARN (ou : du brin de siRNA) sur l'ARN-cible : quand la complémentarité est parfaite, l'ARN-cible est clivé par RNAi, sinon (et s'il s'agit d'un ARN_m), sa traduction est réprimée (Hutvagner et Zamore, 2002 ; Doench *et al.*, 2003 ; Zeng *et al.*, 2003). Quelques mésappariements sont tout de même tolérés pour le RNAi, à condition qu'ils se situent aux extrémités du duplex (Tang *et al.*, 2003) : par exemple, si plus de cinq nucléotides en 5' du petit ARN sont mésappariés, la réaction de clivage *in vitro* est tellement lente qu'elle paraît difficilement compatible avec les conditions *in vivo* (Haley et Zamore, 2004).

La spécificité de la répression traductionnelle pose donc problème : si la complémentarité entre le petit ARN et sa cible est parfaite (sauf éventuellement aux extrémités), l'ARN-cible est clivé par RNAi ; si elle est trop partielle, l'ARN-cible n'est pas reconnu par le petit ARN (puisque les expériences de délétion des sites de complémentarité, qui abolissent la répression traductionnelle, montrent qu'une séquence quelconque n'est pas ciblée par la répression traductionnelle). Les limites de la définition d'une « complémentarité imparfaite », déterminant de la répression traductionnelle, compliquent donc énormément la recherche bio-informatique des sites de complémentarité fonctionnels dans les séquences d'ARN_m.

Deux types d'approche ont été employés, pour prédire *in silico* les complémentarités fonctionnelles :

- la recherche de complémentarités courtes au microARN, qui soient conservées dans des espèces différentes ; puisque les exemples connus de répression traductionnelle impliquent des séquences de la 3'UTR de l'ARN_m-cible, et que l'extrémité 5' du microARN semble appariée à la cible plus stablement que l'extrémité 3' du microARN (voir ci-dessous), ces analyses ont privilégié les complémentarités de ce type (Enright *et al.*, 2003 ; Lewis *et al.*, 2003 ; Stark *et al.*, 2003) ;
- la recherche des déterminants structuraux des appariements fonctionnels, par mutagenèse, et la recherche des complémentarités respectant ces règles (Kiriakidou *et al.*, 2004).

La première stratégie pose un problème conceptuel : la première de ces études a été réalisée alors que seuls trois cas de répression traductionnelle par des microARN étaient connus (*lin-14* et *lin-28* par *lin-4*, et *lin-41* par *let-7*)²⁴. Les sites de complémentarité identifiés se situaient tous dans la 3'UTR des ARN_m-cibles : une séquence de la 3'UTR de *lin-14* contenant cinq

²³En toute rigueur, ces expériences ne montrent pas que la répression est traductionnelle, elle pourrait également être post-traductionnelle : cf page 74.

²⁴Dans cette étude (Stark *et al.*, 2003), les auteurs affirment qu'un autre exemple est connu, celui du microARN *bantam* et de sa cible, l'ARN_m de *hid* de *Drosophila* ; cependant, la répression traductionnelle de *hid*, guidée par *bantam*, a été démontrée par la même équipe, et la complémentarité avait été identifiée grâce au programme décrit par Stark *et al.*, 2003 : au moment du développement de ce programme, le rôle de *bantam* n'était donc pas connu (Brennecke *et al.*, 2003).

des sites de complémentarité est nécessaire au contrôle de la traduction de *lin-14* par *lin-4* (Wightman *et al.*, 1991 ; Wightman *et al.*, 1993), et une séquence contenant quatre des sites de complémentarité confère la sensibilité à *lin-4* d'un gène rapporteur (Ha *et al.*, 1996). De la même manière, la fusion d'un gène rapporteur à la 3'UTR de *lin-28* ou de *lin-41* subit le même contrôle temporel que *lin-28* ou *lin-41* (respectivement), et cette régulation est attribuée à *lin-4*, ou *let-7* (respectivement) (Moss *et al.*, 1997 ; Reinhart *et al.*, 2000). Ces trois exemples ont encouragé plusieurs groupes à limiter les recherches de complémentarités aux 3'UTR (Enright *et al.*, 2003 ; Lewis *et al.*, 2003 ; Stark *et al.*, 2003) (les deux derniers chronologiquement, Enright *et al.*, 2003 et Lewis *et al.*, 2003, confortés par un quatrième exemple, celui de la régulation de *hid* par *bantam*, découvert par le groupe de Stark *et al.* : Brennecke *et al.*, 2003) ; bien entendu, les nouvelles complémentarités qu'ils ont identifiées se trouvaient donc toutes dans des 3'UTR, si bien que la liste de complémentarités identifiées entre ARN_m et microARN (dont la fonctionnalité a, parfois, été confirmée par des expériences de cotransfection du gène du microARN avec un gène rapporteur fusionné à la 3'UTR) ne contenait que des séquences-cibles localisées dans des 3'UTR. Entretemps, un crible génétique a toutefois mis en évidence un cinquième ARN_m-cible de répression traductionnelle, *cog-1* chez le Nématode (réprimé par le microARN *lsy-6*), et il présente un site de complémentarité à *lsy-6* dans sa 3'UTR (Johnston et Hobert, 2003).

Cependant, le déroulement des événements a visiblement biaisé les recherches de complémentarités en faveur des 3'UTR, puisque les critères de sélection d'ARN_m-cibles par ces programmes sont basés sur l'observation d'un petit nombre de cas. Il serait donc injustifié, sans davantage de données expérimentales, d'affirmer que toutes les complémentarités fonctionnelles se trouvent dans des 3'UTR. En particulier, il a été montré *ex vivo* que des complémentarités imparfaites dans la séquence codante d'un ARN_m en guidaient la répression traductionnelle (Saxena *et al.*, 2003) : il est donc tout à fait possible que des microARN endogènes inhibent la traduction d'ARN_m-cibles en s'appariant à leur séquence codante.

D'autre part, parmi les sites de complémentarité identifiés dans les quelques 3'UTR responsables d'une répression traductionnelle guidée par un microARN, plusieurs présentent une complémentarité relativement longue (de l'ordre de 8 à 11 nt), et phylogénétiquement conservée, à l'extrémité 5' du microARN : ce critère était donc également requis dans les recherches d'ARN-cibles (Enright *et al.*, 2003 ; Lewis *et al.*, 2003 ; Stark *et al.*, 2003). Enfin, puisque deux des trois (puis, trois des quatre, après la découverte de la répression de *hid* par *bantam*) ARN_m-cibles expérimentalement démontrés présentaient plusieurs sites de complémentarité imparfaite au microARN régulateur, il semblait probable que des appariements de microARN multiples, sur une 3'UTR, puissent réprimer la traduction de manière synergique ; chacun des trois programmes de recherche d'ARN_m-cibles favorise donc les cibles putatives qui portent plusieurs sites de complémentarité au microARN : en additionnant des scores, calculés pour chaque site de complémentarité, qui reflètent la stabilité prédite de l'appariement, corrigée de la stabilité prédite d'appariements aléatoires (Stark *et al.*, 2003) ; en additionnant des scores, calculés pour chaque site de complémentarité, qui dépendent du nombre d'appariements G-C, A-U et G-U, et des mésappariements dans le duplex prédit (Enright *et al.*, 2003) ; en additionnant les exponentielles de l'opposé de la stabilité prédite de chaque site de complémentarité, divisée par un paramètre appelé « *T* », et choisi pour maximiser le rapport entre le nombre d'ARN-cibles prédits pour les microARN de Mammifères, et le nombre d'ARN-cibles prédits pour des séquences aléatoires de microARN (Lewis *et al.*, 2003) ²⁵.

²⁵En dépit de ses apparences de probabilité d'occupation d'états énergétiques dans une distribution de Boltzmann ($p(E_i) \propto e^{-\frac{E_i}{RT}}$), la formulation de ce score n'a pas de justification théorique. Les auteurs ne justifient pas le choix de l'exponentielle, et le paramètre *T* n'a rien d'une température absolue : il a été « optimisé » à la valeur *T* = 20 après le constat que, parmi cinq valeurs (dont *T* = 20) données au paramètre *T*, la valeur *T* = 20 était celle qui maximisait le rapport : (nombre d'ARN-cibles prédits pour les microARN de Mammifères)/(nombre d'ARN-cibles prédits pour des microARN aléatoires).

La deuxième stratégie (recherche des déterminants structuraux des appariements fonctionnels par mutagenèse), plus rationnelle puisque basée sur l'expérience, présente toutefois l'inconvénient d'être lourde à mettre en œuvre : la mesure de l'effet de mutations affectant la stabilité du duplex entre le petit ARN (microARN ou brin de siRNA) et sa cible n'a été réalisée que sur trois exemples de petits ARN (Doench et Sharp, 2004 ; Kiriakidou *et al.*, 2004). Elle montre que, comme l'avaient suggéré les statistiques, la fonctionnalité de l'appariement est très sensible aux mutations qui déstabilisent l'appariement du 5' du petit ARN. Les conclusions des deux laboratoires divergent, en revanche, sur le rôle de l'appariement de l'extrémité 3' du petit ARN, facultatif dans une étude (Doench et Sharp, 2004), et relativement important dans l'autre (Kiriakidou *et al.*, 2004) ; cette différence pourrait signifier que, selon la séquence du petit ARN, les caractéristiques requises pour la fonctionnalité du duplex ne sont pas les mêmes. La différence des protocoles utilisés par les deux équipes pourrait aussi expliquer cette divergence : dans la première étude, la séquence-cible est clonée en double, entre deux copies non mutées de la complémentarité, alors que dans la deuxième étude, le gène rapporteur ne contient qu'une copie de la séquence-cible, et il est transcrit sous le contrôle d'un promoteur fort ; ces conditions, moins favorables aux effets du petit ARN, pourraient mettre plus facilement en évidence l'influence de mutations délétères. Enfin, les deux groupes ne proposent pas la même interprétation moléculaire des effets mesurés : l'appariement de l'extrémité 5' du petit ARN pourrait simplement stabiliser l'interaction entre les deux ARN, et les effets observés des mutations seraient principalement dus à l'altération de l'enthalpie libre de l'appariement (Doench et Sharp, 2004) ; ou alors, des interactions avec d'autres partenaires (protéiques, par exemple) seraient sensibles à la présence de nucléotides non appariés, sur l'un ou l'autre brin, qui modifie la structure spatiale du duplex (Kiriakidou *et al.*, 2004). En effet, les appariements les plus stables entre un petit ARN et sa cible ne sont pas nécessairement les plus efficaces pour la répression traductionnelle, probablement parce qu'ils interdisent l'association de protéines sur le duplex (Ha *et al.*, 1996).

À la différence de la réaction de RNAi, le processus de répression traductionnelle guidée par les microARN n'a pas encore été reconstitué *in vitro* ; la biochimie de ce phénomène est encore très mal connue.

Plusieurs protéines interagissent avec les microARN, chez les Mammifères : Gemin3 et Gemin4 (qui participent probablement à l'assemblage de nombreuses particules ribonucléoprotéiques), et eIF2C2 (protéine de la famille Argonaute) (Hutvagner et Zamore, 2002 ; Mourelatos *et al.*, 2002). Ce complexe contient aussi les trois protéines de la famille FMRP (la protéine FMRP, et ses deux paralogues, FXR1P et FXR2P ; ces protéines doivent leurs noms au syndrome de retard mental de l'X fragile, une maladie génétique humaine causée par la déficience de FMRP), connues pour réprimer la traduction d'ARN_m spécifiques ; il est donc possible que le mode d'action de ces protéines se résume à leur rôle dans la répression traductionnelle guidée par des microARN (Jin *et al.*, 2004). Ce complexe, baptisé « miRNP », cosédimente avec les polyribosomes, et il est physiquement associé à un ARN_m-cible de répression traductionnelle : il est donc probablement l'effecteur de la répression traductionnelle (Nelson *et al.*, 2004).

Le mode d'action de ce complexe n'est pas encore connu. Néanmoins, il a été établi que le profil d'association d'un ARN_m-cible avec les polyribosomes n'était pas modifié en présence de son microARN inhibiteur (Olsen et Ambros, 1999 ; He et Sontheimer, 2004). L'association de l'ARN_m avec plusieurs ribosomes suggère que l'initiation de la traduction n'est pas affectée par l'association de la miRNP : la traduction pourrait donc débuter, puis l'élongation de la traduction ou la libération de la chaîne peptidique pourrait être inhibée par la miRNP.

1.4.4 Fonctions biologiques

Les rôles des microARN et du RNAi chez les Vertébrés sont encore mal connus, en raison de la difficulté d'étude de ces systèmes. En revanche, de nombreux résultats ont été obtenus chez

les Plantes, la levure à fission, les Ciliés, le Nématode et la Drosophile, et de façon surprenante, ils indiquent que les microARN d'une part, et le RNAi d'autre part, seraient impliqués dans des processus communs chez ces espèces éloignées : le contrôle du développement pour les microARN, et la répression des éléments génétiques parasites (transposons et virus) et la structuration des centromères, pour le RNAi. Les raisons de cette similarité ne sont pas encore claires, mais il est possible qu'elles s'appliquent aussi aux Vertébrés, et que ces tendances observées chez les Plantes et les Invertébrés se généralisent aux Vertébrés, qui font l'objet de ma thèse. Cette partie présentera donc les fonctions biologiques des microARN et du RNAi chez les Plantes et les Invertébrés, et les données disponibles chez les Vertébrés.

MicroARN et contrôle du développement

Les deux premiers microARN découverts, *lin-4* et *let-7*, régulent des transitions entre stades larvaires chez le Nématode, et *let-7* semble participer au contrôle d'événements développementaux dans une grande variété d'espèces animales (Lee *et al.*, 1993 ; Moss *et al.*, 1997 ; Reinhart *et al.*, 2000 ; Pasquinelli *et al.*, 2000) (voir aussi p. 59). Leur patron temporel d'expression, associé à cette fonction, leur a valu le nom de « petits ARN temporels » (stRNA, pour : *small temporal RNAs*) (Pasquinelli *et al.*, 2000).

Le rôle biologique d'un troisième microARN de Nématode est connu : le microARN *lsy-6* est nécessaire à l'asymétrisation d'une paire de neurones chez le Nématode (ces deux neurones, l'un à droite du plan de symétrie de l'animal, l'autre à gauche, expriment des chémorécepteurs différents, et cette différence fonctionnelle entre les deux neurones est indispensable pour que le ver puisse distinguer certains stimuli olfactifs) (Johnston et Hobert, 2003). Ce microARN inhibe la traduction d'un ARN_m-cible, qui code l'une des protéines impliquées dans le choix de la destinée cellulaire de cette paire de neurones.

Chez la Drosophile, la génétique a montré que le locus *bantam* était impliqué dans le contrôle de la croissance : les larves des homozygotes mutants sont plus petites que les larves sauvages, et meurent au début du stade pupal, tandis que les adultes hétérozygotes sont de taille réduite. Réciproquement, la surexpression du locus augmente la croissance (Hipfner *et al.*, 2002). Chez les mutants *bantam*, la taille des cellules est normale, leur nombre est simplement réduit. L'analyse de la mutation a montré que le gène responsable du phénotype était un gène de microARN (Brennecke *et al.*, 2003). Le microARN *bantam* est détecté dans les embryons, puis son expression décroît au fil des stades larvaires, et il est à nouveau exprimé à partir du milieu du stade pupal ; l'observation de larves transgéniques, exprimant une protéine fluorescente dont l'ARN_m contient des sites de complémentarité parfaite à *bantam* (le RNAi guidé par *bantam*, dans les cellules où il est exprimé, réprime l'expression de ce transgène ; le patron de fluorescence de ces larves donne donc une image, en négatif, de l'expression de *bantam*) a montré que l'expression du microARN corrèle avec la prolifération cellulaire ; *bantam* est aussi un anti-apoptotique, en réprimant la traduction de l'ARN_m du gène pro-apoptotique *hid* (Brennecke *et al.*, 2003).

Un autre crible génétique, visant à identifier des répresseurs du pro-apoptotique *reaper*, en a localisé un dans un locus connu pour héberger un gène de microARN déjà annoté, *miR-14*. Des expériences supplémentaires ont montré que ce gène était bien un répresseur de *reaper*, et que sa délétion provoquait une forte létalité au stade pupal, et une grande sensibilité des animaux à plusieurs types de stress (Xu *et al.*, 2003). Les mutants pour *mir-14* manifestent aussi une nette dérégulation du métabolisme des lipides : ils accumulent de grandes quantités de di- et triglycérides dans les adipocytes. Ce deuxième phénotype pourrait être lié au premier, puisque plusieurs partenaires sont communs aux voies de signalisation de l'apoptose, et au métabolisme des lipides (Xu *et al.*, 2003).

Chez les Plantes, de nombreux exemples de complémentarité parfaite, ou presque parfaite, ont été signalés, suggérant que la plupart des microARN de Plantes agissent en guidant le RNAi,

plus qu'une répression traductionnelle, sur leurs ARN-cibles (voir Bartel, 2004, pour une revue ; voir aussi p. 68). Et effectivement, des produits de clivage par RNAi des ARN-cibles sont détectés dans plusieurs cas ; et de manière remarquable, chacun de ces ARN-cibles est impliqué dans des processus développementaux.

Les microARN miR-165 et miR-166 d'*Arabidopsis thaliana* (dont les séquences ne diffèrent que d'un nucléotide ; ils pourraient remplir des fonctions redondantes, d'autant que le génome du Riz ne contient pas de gène *mir-165*, mais six copies de *mir-166*) guident le clivage de l'ARN_m du gène *REVOLUTA in vivo* (Emery *et al.*, 2003), et des ARN_m de *PHAVOLUTA* et *PHABULOSA in vitro* (Tang *et al.*, 2003). Ces trois gènes, de la même famille multigénique, participent au contrôle de la polarité des organes latéraux, et une mutation du site d'interaction de l'ARN_m *REVOLUTA* avec miR-165 et miR-166 (qui ne change pas la séquence protéique de *REVOLUTA*) lui confère un phénotype gain-de-fonction (Tang *et al.*, 2003). Le microARN miR-171 (aussi appelé : miRNA 39) dirige le clivage par RNAi des ARN_m de plusieurs membres de la famille de facteurs de transcription putatifs *Scarecrow-like*, impliqués dans l'asymétrie de certaines mitoses dans les racines (nécessaires à la polarisation radiale des racines), et dans des voies de transduction du signal (Llave *et al.*, 2002b). Le microARN JAW (ou : miR-JAW) guide le clivage des ARN_m de plusieurs gènes connus pour contrôler le développement de la feuille, les gènes *TCP* (Palatnik *et al.*, 2003) ; les feuilles des mutants du gène *JAW* ont une morphologie altérée.

Chez les Vertébrés, on ne connaît pour le moment qu'une cible de microARN, et il s'agit, là encore, de l'ARN_m d'un gène de contrôle du développement, le gène *HOXB8* (Yekta *et al.*, 2004). Le microARN miR-196a en guide le clivage par RNAi. *HOXB8* appartient à la famille des gènes à homéoboîte, des facteurs de transcription impliqués dans la mise en place de structures régionalisées, pendant l'embryogenèse ; les gènes *HOX* sont regroupés en « complexes », dont tous les gènes sont transcrits dans le même sens. Au cours de l'embryogenèse, ces différents gènes sont activés séquentiellement, et l'ordre dans lequel ils sont activés suit l'ordre spatial de leur disposition dans le complexe (voir Kmita et Duboule, 2003, pour revue). Une carte des complexes *HOX* de Mammifères et d'Insectes est donnée en figure 1.16.

Il est remarquable que le gène de miR-196a (et celui de miR-196b, très similaire) se trouve précisément dans les complexes *HOX* ; sa localisation génomique pourrait signifier que son expression est régulée temporellement comme celle des gènes *HOX*. Hormis *HOXB8*, plusieurs gènes *HOX* présentent des complémentarités imparfaites à miR-196a et miR-196b, qui pourraient être responsables d'une répression traductionnelle par ces microARN ; lorsque ces séquences sont fusionnées à un gène rapporteur, elles en provoquent la répression, dépendante des microARN miR-196, dans des cellules transfectées (Yekta *et al.*, 2004).

La fonction d'un autre microARN de Vertébré, miR-181, est probablement connue, même si son rôle au niveau moléculaire n'a pas été découvert : ce microARN, chez la Souris, est faiblement exprimé dans les progéniteurs des cellules sanguines, puis surexprimé quand elles se différencient en lymphocytes B. La surexpression de ce microARN dans des progéniteurs augmente, *in vivo*, la proportion de progéniteurs qui se différencient en lymphocytes B (Chen *et al.*, 2004). Cet effet est également observé *ex vivo*, sur des progéniteurs en culture, et deux autres microARN, miR-142s et miR-223 ont un effet inverse dans ce test *ex vivo* : leur surexpression augmente la proportion de cellules qui s'engagent dans la différenciation en lymphocytes T.

Ces microARN de Vertébrés gouvernent donc, eux aussi, des processus développementaux. Cependant, dans le cas des trois microARN impliqués dans la différenciation des lymphocytes, les auteurs ont délibérément recherché des microARN régulés au cours de l'hématopoïèse, partant du principe que les microARN d'Invertébrés jouent fréquemment des rôles importants dans le développement (Chen *et al.*, 2004) ; en somme, ils ne pouvaient trouver que des microARN régulés au cours du développement, et il n'est pas surprenant que ceux qu'ils ont trouvés participent à ce type de contrôle.

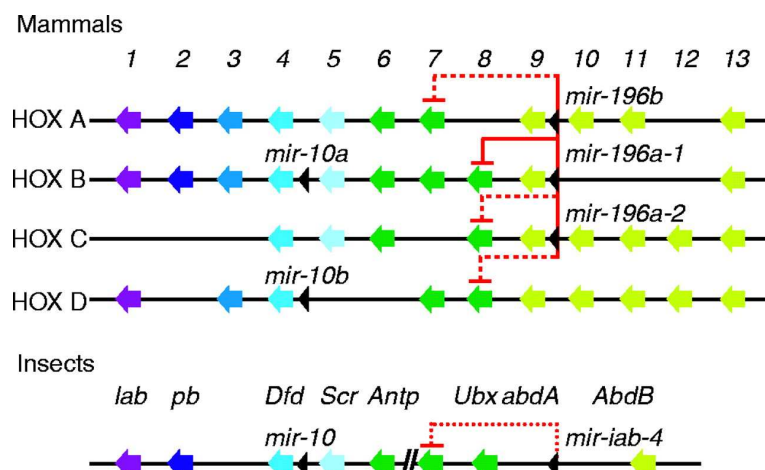


FIG. 1.16 – ORGANISATION GÉNOMIQUE DES GÈNES HOX. Figure tirée de Yekta *et al.*, 2004. Chez les Mammifères, les gènes HOX se trouvent dans quatre complexes (HOX A, HOX B, HOX C et HOX D), dont l'organisation est très similaire : les 13 paralogues (numérotés de 1 à 13 sur la figure, et représentés par des flèches de couleur) sont toujours dans le même ordre, et plusieurs contiennent des gènes de microARN apparentés (représentés par des flèches noires). Les relations de régulation par les microARN sont indiquées en rouge (trait plein : guidage d'un clivage par RNAi ; pointillés : répression traductionnelle, prédite et éventuellement confirmée *ex vivo*).

Outre les microARN dont l'effet est avéré dans des processus développementaux, plusieurs recherches bio-informatiques d'ARN_m-cibles ont révélé des complémentarités entre des microARN et des ARN_m. Ces prédictions *in silico* ont parfois été vérifiées expérimentalement, par des co-transfections dans des cellules en culture, du gène du microARN (ou d'un siRNA synthétique censé mimer le microARN) et de sa séquence-cible présomptive, fusionnée à un gène rapporteur. Cette méthode n'est pas aussi convaincante que les approches *in vivo* présentées ci-dessus, puisque, dans de nombreux cas, rien n'indique que, *in vivo*, il existe des types cellulaires où le microARN et sa cible supposée sont coexprimés, et où les concentrations intracellulaires des deux partenaires seraient comparables à celles obtenues après transfection. De plus, dans certains travaux, le gène rapporteur est fusionné à plusieurs copies de la séquence-cible, ce qui pourrait faciliter les artefacts (Bartel, 2004). Toutes ces difficultés expliquent peut-être l'effet qui a été rapporté, d'un microARN humain sur la traduction d'un ARN_m : les auteurs ont cru observer une répression traductionnelle de cet ARN_m guidée par le microARN, sensible aux mutations du microARN, mais restaurée par des mutations compensatoires sur l'ARN_m, alors que la protéine dont ils suivaient l'accumulation n'était pas celle codée par l'ARN_m en question (Kawasaki et Taira, 2003a ; Kawasaki et Taira, 2003b).

Avec cet avertissement à l'esprit, si l'on analyse la liste des gènes-cibles prédits pour les microARN de Plantes (Llave *et al.*, 2002a ; Reinhart *et al.*, 2002 ; Rhoades *et al.*, 2002 ; Llave *et al.*, 2002b ; Kasschau *et al.*, 2003 ; Park *et al.*, 2002a ; Xie *et al.*, 2003) et d'Animaux (Lai, 2002 ; Stark *et al.*, 2003 ; Lewis *et al.*, 2003), on s'aperçoit que la grande majorité est impliquée dans des contrôles d'événements développementaux (Bartel et Bartel, 2003 ; Bartel, 2004).

Les fonctions de contrôle du développement de la plupart des microARN expliquent sans doute les effets pléiotropiques de l'inactivation de Dicer chez *Arabidopsis thaliana* (Jacobsen *et al.*, 1999)²⁶, chez le Poisson-zèbre (Wienholds *et al.*, 2003) et la Souris (Bernstein *et al.*,

²⁶ *Arabidopsis thaliana* possède plusieurs isoformes de Dicer ; celle qui a été invalidée ici est l'enzyme responsable de la maturation des microARN (Finnegan *et al.*, 2003).

2003). Les phénotypes associés à la mutation de Dicer chez le Poisson-zèbre et chez la Souris sont toutefois différents : alors que l'embryogenèse du poisson-zèbre homozygote mutant s'accomplit jusqu'à un stade avancé, celle de la souris homozygote mutante s'interrompt au cours de la première semaine. Cette différence semble due, au moins en partie, à la forte contribution maternelle chez le poisson, puisqu'un traitement des embryons homozygotes mutants du stade « une cellule » avec des morpholino-oligonucléotides (qui titrent les ARN_m Dicer éventuellement hérités maternellement) arrête le développement du mutant à un stade plus précoce que la simple mutation homozygote ; Dicer pourrait même être indispensable à une étape antérieure, et une contribution maternelle en protéine Dicer permettrait à l'embryon mutant d'y survivre.

Les microARN, molécules de la pluricellularité ?

Tous les microARN décrits jusqu'à présent ne sont exprimés que dans des organismes pluricellulaires. De surcroît, le clonage de petits ARN phosphorylés en 5' et hydroxylés en 3' chez *Schizosaccharomyces pombe* n'a identifié (en plus des produits de dégradation d'ARN abondants) que des siRNA issus des séquences centromériques, dont les précurseurs semblent être des ARN double brin bimoléculaires, plutôt que des ARN monomoléculaires, repliés en tige-boucle (Reinhart et Bartel, 2002 ; Volpe *et al.*, 2002).

Les microARN pourraient donc présenter un avantage spécifique aux organismes pluricellulaires. Ils pourraient, par exemple, aider chaque type cellulaire à acquérir son identité, et donc, optimiser le partage des tâches entre cellules, une caractéristique des espèces pluricellulaires. Plusieurs auteurs ont suggéré que le contenu de chaque type cellulaire en microARN spécifiques (dont on sait à présent qu'il existe une grande variété) constituerait un environnement particulier, qui influencerait l'expression des gènes (Bartel et Chen, 2004 ; Hobert, 2004). Le répertoire de microARN exprimés par un type cellulaire donné pourrait réguler finement la traduction de tous les ARN_m de la cellule, principalement chez les Animaux, où les microARN semblent guider des répressions traductionnelles, plus que des clivages par RNAi : dans ces conditions, chaque ARN_m pourrait s'hybrider à une classe donnée de microARN (dont les contributions respectives sur la répression traductionnelle s'ajouteraient), et réciproquement, chaque microARN pourrait réguler une certaine population d' ARN_m . La faible spécificité apparente des microARN (ils reconnaissent leurs cibles par une quinzaine d'appariements, sans contrainte structurale évidente : cf « Les ARN courts et le guidage de la répression traductionnelle », page 71) leur permettrait alors de réguler de nombreux gènes ; comme le font remarquer Bartel et Chen, certains microARN sont plus abondants dans la cellule que la plupart des ARN_m , de deux ou trois ordres de grandeur : ils sont donc susceptibles de réguler des centaines d' ARN_m différents. La régulation de la combinaison de microARN exprimés offre donc l'opportunité de réguler finement l'expression de tous les ARN_m de la cellule en modulant leur traduction, donc de contrôler les caractéristiques d'un type cellulaire donné.

La nature post-transcriptionnelle des régulations guidées par les microARN présente un autre avantage pour les organismes pluricellulaires : après une division asymétrique, dont chaque cellule-fille est appelée à remplir une fonction particulière, les ARN_m hérités du cytoplasme de la cellule-mère, communs aux deux cellules-filles, pourraient polluer les ARN_m exprimés spécifiquement par chaque cellule-fille. La spécialisation des cellules-filles serait donc retardée, jusqu'à la dégradation complète des ARN_m de la cellule-mère. Des microARN, exprimés par une cellule-fille immédiatement après la division, pourraient réprimer ces ARN_m (par RNAi, ou par répression traductionnelle) beaucoup plus rapidement (Rhoades *et al.*, 2002 ; Bartel et Chen, 2004).

RNAi et extinction des parasites génétiques

L'invalidation des gènes impliqués dans le RNAi se traduit souvent par une dérégulation des séquences répétées (le plus souvent, des transposons), et une sensibilité accrue aux virus. Cette observation a conduit plusieurs auteurs à proposer que la fonction principale du RNAi serait la défense contre les séquences envahissantes (Voinnet, 2001 ; Matzke *et al.*, 2002 ; Plasterk, 2002 ; Wassenegger, 2002 ; Zamore, 2002).

Dans la plupart des isolats du Nématode *Caenorhabditis elegans*, les transposons sont réprimés dans la lignée germinale. Un crible génétique a identifié plusieurs gènes qui participent à la répression de différentes familles de transposons, et la plupart des mutations révélées par ce crible affectent aussi le processus de RNAi (Ketting *et al.*, 1999). Simultanément, un crible visant à isoler des gènes impliqués dans le RNAi chez le Nématode a mis en évidence plusieurs gènes nécessaires à la répression des transposons (Tabara *et al.*, 1999). Ainsi, chez le Nématode, l'une des fonctions du RNAi semblait être l'extinction des transposons dans la lignée germinale. Les extrémités de l'ARN du transposon, issues de la transcription des répétitions inversées qui bordent la séquence du transposon, se replient en tiges-boucles, et constitueraient l'ARN double brin déclencheur du RNAi (Sijen et Plasterk, 2003) ; des siRNA, provenant essentiellement de ces séquences particulières du transposon, sont effectivement détectés, et semblent guider la dégradation, post-transcriptionnelle, d'ARN rapporteurs portant une séquence homologue aux répétitions inversées du transposon (Ambros *et al.*, 2003 ; Sijen et Plasterk, 2003). Des gènes requis pour la répression des transposons, mais qui ne le sont pas nécessairement pour le RNAi de manière générale, sont nécessaires à la production des siRNA issus des répétitions inversées des transposons (Sijen et Plasterk, 2003) ; si de tels gènes étaient spécifiquement exprimés dans la lignée germinale, la spécificité de tissu de l'extinction des transposons chez le Nématode pourrait alors être due à une spécificité de tissu de la production des siRNA.

Chez la Drosophile, la mutation de *spindle E/homeless* (une hélicase à ARN, homologue du gène *rde-1* du Nématode ; tous deux sont requis pour le RNAi : Tabara *et al.*, 1999 ; Kennerdell *et al.*, 2002) lève la répression de rétrotransposons, et des séquences répétées *mst40* (Aravin *et al.*, 2001). Le rétrovirus *gypsy* semble lui aussi réprimé par RNAi, puisque la mutation de *piwi* (ce gène est essentiel pour le RNAi : Pal-Bhadra *et al.*, 2002) dérégule *gypsy*, et des petits ARN, d'environ 25 à 27 nt, et homologues à la 5'UTR de *gypsy*, sont détectés dans les ovaires de Drosophile (Sarot *et al.*, 2004).

Le RNAi est également impliqué dans la régulation du locus *Stellate* chez la Drosophile. Ce gène (abrégé : *Ste*), porté par le chromosome X, est responsable de la formation de cristaux dans les spermatocytes primaires des mâles de génotype XO (donc : dépourvus de chromosome Y), et la forme de ces cristaux dépend de l'allèle de *Ste* porté par le chromosome X (l'un des allèles leur donne une forme étoilée, d'où le nom du locus). L'effet de *Stellate* est réprimé par le « suppresseur de *Stellate* », *Su(Ste)*, porté par le chromosome Y (Hardy *et al.*, 1984 ; Livak, 1984), et la délétion de *Su(Ste)* s'accompagne de la formation des cristaux dans les spermatocytes, mais aussi de défauts de disjonction des chromosomes homologues, à la première division de méiose (Hardy *et al.*, 1984). Les gènes *Ste* et *Su(Ste)* sont constitués de séquences répétées homologues, apparentées à la séquence codant une sous-unité de la caséine kinase II (les répétitions de *Su(Ste)* contiennent, en plus, une séquence spécifique) (Livak, 1984 ; Balakireva *et al.*, 1992). La répression de *Ste* par *Su(Ste)* est sensible aux mutations des gènes *aubergine/sting* et *spindle-E/homeless*, requis pour le RNAi, et de petits ARN d'environ 25 nt, homologues de chacun des deux brins de *Ste*, sont détectés dans les testicules (Aravin *et al.*, 2001 ; Aravin *et al.*, 2004).

Enfin, chez la Drosophile au moins, le RNAi dégrade les ARN du virus FHV après infection ; en particulier, des ARN d'environ 22 nt sont détectés dans les cellules infectées, et l'absence de la protéine AGO2 (nécessaire au clivage guidé par des siRNA, voir page 64) se traduit par une accumulation des ARN viraux (Li *et al.*, 2002). Le virus, quant à lui, produit une protéine

qui réprime le RNAi : il aurait donc évolué de façon à neutraliser ce système de défense des cellules-hôtes. D'autres exemples de ce type de « course à l'armement » entre le virus et l'hôte ont également été décrits pour plusieurs virus végétaux, qui inhibent aussi le RNAi (Vance et Vaucheret, 2001 ; voir aussi ci-dessous).

Chez les Plantes, les transposons et les virus sont des cibles endogènes du RNAi. Ainsi chez l'Algue unicellulaire *Chlamydomonas reinhardtii*, le gène *Mut6*, identifié par un crible génétique en tant qu'acteur du RNAi, est également requis pour la répression d'un transposon et d'un rétrotransposon (Wu-Scharf *et al.*, 2000). Chez le Maïs, de petits ARN de 21 à 26 nt homologues d'un transposon ont été observés, mais ce ne sont probablement pas des siRNA, puisque l'un des deux brins (le brin antisens) s'accumule nettement plus que l'autre ; les deux brins ne sont donc pas présents en quantités stœchiométriques pour former des duplex (Rudenko *et al.*, 2003).

L'effet du RNAi sur les infections virales chez les Plantes est connu de longue date ; ainsi, la résistance naturelle aux virus, qui est partiellement post-transcriptionnelle (Al-Kaff *et al.*, 1998), peut être mimée, et renforcée, par l'expression de transgènes portant des séquences du virus (Lindbo et Dougherty, 1992a ; Lindbo et Dougherty, 1992b ; Lindbo *et al.*, 1993 ; Baulcombe, 1996a ; Goodwin *et al.*, 1996 ; Sijen *et al.*, 1996). L'analyse mécanistique de ce qui fut appelé le « système immunitaire des Plantes » a progressivement démontré qu'il fonctionnait par interférence à ARN (Waterhouse *et al.*, 1998 ; Mourrain *et al.*, 2000 ; Papaefthimiou *et al.*, 2001). Au cours de l'évolution, certains virus ont acquis des gènes de suppression du RNAi, qui leur permettent de s'affranchir de ce système de défense de la plante. Ces supprimeurs du RNAi s'attaquent à diverses étapes du processus de RNAi (initiation, maintien, propagation systémique) (Anandalakshmi *et al.*, 1998 ; Brigneti *et al.*, 1998 ; Kasschau et Carrington, 1998 ; Voinnet *et al.*, 1999 ; Voinnet *et al.*, 2000 ; Guo et Ding, 2002 ; Pfeffer *et al.*, 2002 ; Silhavy *et al.*, 2002 ; Takeda *et al.*, 2002 ; Yelina *et al.*, 2002 ; Bucher *et al.*, 2003 ; Reed *et al.*, 2003 ; Liu *et al.*, 2004b).

1.4.5 RNAi et guidage de l'hétérochromatinisation

Ces deux dernières années, le champ d'action du RNAi s'est étendu dans une direction surprenante : chez des espèces très distinctes (la levure à fission *Schizosaccharomyces pombe* et deux Vertébrés, l'Homme et le Poulet), des gènes effecteurs du RNAi se sont avérés nécessaires à la cohésion des chromatides sœurs en mitose (Provost *et al.*, 2002b ; Volpe *et al.*, 2003 ; Schramke et Allshire, 2003 ; Fukagawa *et al.*, 2004). Le RNAi semble en effet nécessaire à l'hétérochromatinisation des centromères, qui est elle-même requise pour le recrutement de la cohésine Rad21, qui assure l'arrimage des deux chromatides.

Un autre rôle inattendu s'est révélé récemment, avec l'implication du RNAi dans les réarrangements chromosomiques qui accompagnent la différenciation du macronoyau des Ciliés (Mochizuki *et al.*, 2002 ; Garnier *et al.*, 2004). Là encore, l'action du RNAi semble passer par l'hétérochromatinisation de séquences spécifiques.

RNAi et hétérochromatinisation chez *Schizosaccharomyces pombe*

En 2002, T. Volpe et ses collaborateurs ont rapporté une observation surprenante : la mutation, chez la levure à fission *Schizosaccharomyces pombe*, de l'un des trois gènes *ago1* (l'unique gène argonaute chez *S. pombe*), *dcr1* (l'unique gène dicer) et *rdp1* (la polymérase d'ARN dépendant de l'ARN, abrégée RdRP en anglais) aboutit à l'expression d'ARN issus de la transcription des répétitions centromériques externes, à partir des deux brins, et à la dérégulation de transgènes insérés dans ces répétitions (Volpe *et al.*, 2002). La figure 1.17 présente l'organisation des centromères de *S. pombe*.

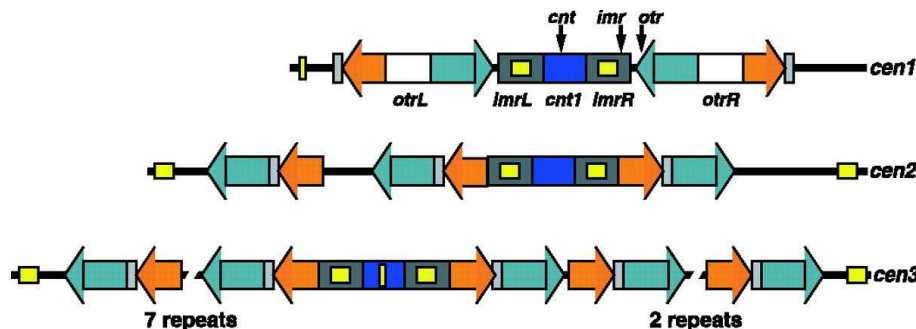


FIG. 1.17 – ORGANISATION DES CENTROMÈRES DE *Schizosaccharomyces pombe*. Figure tirée de Volpe *et al.*, 2002. Les répétitions externes sont représentées par des flèches orange et vertes, et appelées « otr » ; les régions internes sont appelées « imr », et la région centrale, « cnt ». Le schéma représente les centromères des chromosomes n°1 (« cen1 »), 2 (« cen2 ») et 3 (« cen3 »). Les rectangles jaunes indiquent les régions contenant un ou plusieurs gène(s) d'ARN_t.

Chez la levure sauvage, l'un des deux brins (appelé *forward* par convention) n'est pas transcrit, et l'autre (appelé *reverse*) est transcrit, mais ses ARN sont dégradés post-transcriptionnellement, puisqu'ils sont indétectables dans l'ARN total. Chez chacun des trois mutants pour la machinerie du RNAi, en revanche, les deux brins sont transcrits à des taux comparables, et leurs ARN s'accumulent. Enfin, chez le mutant de Swi6 (Swi6 est l'orthologue de la protéine de l'hétérochromatine HP1 des Vertébrés et de la Drosophile), les deux brins sont transcrits, mais seul l'ARN *forward* est détectable (Volpe *et al.*, 2002). Ainsi, ces mutations du RNAi dérèglent post-transcriptionnellement le brin *reverse*, et transcriptionnellement le brin *forward*. La dérégulation transcriptionnelle semble due à la perte de la structure hétérochromatique chez les trois mutants du RNAi : l'histone H3 de la chromatine des répétitions centromériques externes est moins méthylée sur la lysine 9, et plus méthylée sur la lysine 4, chez ces mutants que dans les cellules sauvages. Cet effet serait responsable de la perte de la protéine Swi6 sur le transgène inséré dans les répétitions centromériques, d'où sa dérégulation transcriptionnelle. Enfin, une expérience d'immuno-précipitation montre que la RdRP Rdp1 est associée à la chromatine des répétitions centromériques externes, et que Dcr1 n'y est pas associée.

Ainsi, chez la levure sauvage, l'ARN *reverse*, faiblement transcrit, serait dégradé en permanence par la machinerie du RNAi. Cette dégradation pourrait être le fait de Dcr1 (et il faudrait alors que l'ARN *reverse* soit sous forme double brin ; il faudrait aussi expliquer la dépendance à Ago1 : pourquoi la dégradation par Dcr1 nécessiterait-elle un gène *ago1* fonctionnel ?), ou bien, l'ARN *reverse* pourrait être clivé par RNAi guidé par des siRNA, produits par Dcr1. Dans la première hypothèse, le brin apparié à l'ARN *reverse* pourrait provenir de la transcription du deuxième brin d'ADN (mais, puisqu'il semble nettement moins transcrit, il est difficile de concevoir que cette transcription suffise à produire autant d'ARN *forward* qu'il y a d'ARN *reverse*, pour que la totalité de l'ARN *reverse* soit sous forme bicaténaire), ou, plus probablement, de la polymérisation par la RdRP, sur l'ARN *reverse*, d'un brin complémentaire. Dans la seconde hypothèse, la réaction de clivage guidée par les siRNA nécessiterait Ago1 et l'activité de Rdp1, qui fournirait un substrat à Dcr1 (Rdp1 pourrait donc contribuer à la production des siRNA, d'où l'effet de la mutation de *rdp1*). Simultanément à ces travaux, une autre équipe a cloné 61 petits ARN phosphorylés en 5', et hydroxylés en 3', à partir de *S. pombe* : parmi ces 61 ARN, 49 étaient des produits de dégradation d'ARN_r ou d'ARN_t, mais les 12 autres étaient tous issus des répétitions centromériques, ou de séquences immédiatement voisines (Reinhart et Bartel, 2002). Puisque les ARN *forward* et *reverse* ne semblent pas pouvoir se replier en structures en

tiges-boucles, ces petits ARN ne sont probablement pas des microARN, mais plutôt les produits de la digestion par Dcr1 de l'ARN double brin formé par les ARN *forward* et *reverse* (Volpe *et al.*, 2002). Ces petits ARN pourraient ensuite guider la modification des histones de leur locus ; par exemple, Rdp1, qui interagit avec la chromatine, et probablement aussi avec les siRNA, pourrait jouer un rôle dans la réaction (Volpe *et al.*, 2002).

Peu de temps après, il était démontré que le RNAi est aussi responsable de la nucléation de l'hétérochromatinisation de la région de type sexuel (appelée : région *mat*), qui contrôle le type sexuel de *S. pombe*²⁷ : ce locus contient une séquence, appelée *cenH*, très similaire aux répétitions centromériques externes, et dont la machinerie du RNAi est nécessaire à l'hétérochromatinisation (Hall *et al.*, 2002). Plus précisément, le RNAi est dispensable pour le maintien de la structure hétérochromatique de la région *mat*, mais si un traitement chimique, ou un contexte génétique particulier, a éliminé les marques hétérochromatiniennes de la région, les mutants du RNAi sont incapables de la ré-hétérochromatiser. L'hétérochromatinisation semble se nucléer sur la séquence *cenH*, puis se propager (par un mécanisme faisant intervenir Swi6 et la méthyl-transférase de la lysine 9 de l'histone H3, Clr4) aux séquences voisines, jusqu'aux premiers éléments frontières (Hall *et al.*, 2002).

Cependant, un mécanisme supplémentaire, indépendant de la machinerie du RNAi, finit par nucléer, avec une faible efficacité, l'hétérochromatinisation de la région *mat* chez les mutants *ago1*⁻, *dcr1*⁻ et *rdp1*⁻ (Hall *et al.*, 2002). Ce deuxième mécanisme semble reposer sur la reconnaissance de deux séquences du locus par les protéines Atf1 et Pcr1, qui seraient capables, tout comme les composants du RNAi, de provoquer une hétérochromatinisation qui se répand sur tout le locus (Jia *et al.*, 2004).

Le guidage de l'hétérochromatinisation par le RNAi semble pouvoir fonctionner aussi en *trans*, puisque l'expression d'un ARN capable de se replier en une longue tige-boucle provoque l'hétérochromatinisation d'un locus qui porte la séquence de l'un des brins de la tige (Schramke et Allshire, 2003). Curieusement, ce phénomène (et, plus précisément, la production de petits ARN d'une vingtaine de nucléotides, à partir de la tige-boucle) est dépendant de Clr4. Cette méthyl-transférase de la lysine 9 de l'histone H3 est donc nécessaire à la digestion, par Dcr1, d'un long ARN double brin.

La même étude a montré que les longues répétitions terminales (LTR, pour : *long terminal repeats*, en anglais) de rétrotransposons pouvaient contrôler l'expression de gènes voisins par une hétérochromatinisation dépendante du RNAi : plusieurs gènes spécifiquement exprimés après la privation de nutriments azotés (connue pour induire la différenciation sexuelle chez *S. pombe*) sont localisés à proximité de LTR, et sont dérprimés dans un milieu riche en nutriments azotés si l'un des gènes du RNAi (*ago1*, *dcr1* ou *rdp1*) ou de protéines de l'hétérochromatine (*swi6* ou *clr4*) est muté. Or, la chromatine des LTR du voisinage de deux de ces gènes est condensée en hétérochromatine dans les cellules sauvages, et décondensée si l'un des trois gènes du RNAi, ou *clr4*, est muté ; ces LTR sont nécessaires à la répression, en milieu riche en nutriments azotés, des deux gènes voisins. Enfin, alors qu'aucun transcrit issu de ces LTR n'est détecté dans les levures sauvages, des ARN issus de la transcription des deux brins sont détectés chez les mutants *ago1*Δ, *dcr1*Δ, *rdp1*Δ et chez un mutant *clr4*, et l'ARN issu de la transcription d'un seul brin est détecté chez le mutant *swi6*Δ (Schramke et Allshire, 2003) : la répression de ces LTR est donc très similaire à la répression des répétitions centromériques externes (voir plus haut), et semble, là aussi, impliquer une contribution transcriptionnelle (sur l'un des brins), et une contribution post-transcriptionnelle, par RNAi, sur l'autre brin.

²⁷Ce locus de 20 kb contient deux séquences, *mat2* et *mat3*. Une recombinaison peut copier l'une de ces séquences au site d'expression *mat1*, et la nature de la séquence portée par *mat1* détermine le type sexuel de la levure. La région de type sexuel est donc une région transcriptionnellement inactive, qui stocke les séquences *mat2* et *mat3* qui servent à chaque changement de type sexuel.

Enfin, la purification d'un complexe associé à Chp1 (une protéine impliquée dans une étape précoce de l'hétérochromatinisation, et dont la mutation conduit à un phénotype semblable à celui des mutants de la machinerie du RNAi) a montré que cette protéine interagissait avec Ago1, une protéine de fonction alors inconnue, baptisée Tas3 (pour : *targeting complex subunit 3*), et des petits ARN d'environ 22 à 25 nt, dont la production dépend de Dcr1 (Verdel *et al.*, 2004). Ce nouveau complexe a été baptisé RITS, pour : *RNA-induced initiation of transcriptional gene silencing*. Les petits ARN du complexe RITS contiennent, au moins en partie, les siRNA hétérochromatiques issus des répétitions centromériques, clonés par Reinhart et Bartel, 2002 (voir plus haut). Toutefois, le complexe RITS ne semble pas contenir de siRNA issus de LTR. Les auteurs suggèrent que les siRNA issus de LTR sont beaucoup moins abondants que les siRNA issus des répétitions centromériques, ce qui compliquerait leur détection.

Le complexe RITS est associé à la chromatine des répétitions centromériques, et l'association des composants du complexe à la chromatine dépend à la fois de Tas3 et Ago1, mais également de Dcr1 et Rdp1, qui ne semblent pourtant pas participer au complexe. La localisation du complexe sur ses cibles chromatiniques serait donc due à ses composants ribonucléiques, les siRNA, qui guideraient le complexe sur les régions génomiques homologues.

La figure 1.18, page 84, présente l'hypothèse, présentée par les auteurs de ce travail, pour expliquer l'intervention de la machinerie du RNAi dans le guidage de l'hétérochromatinisation.

L'hétérochromatinisation des séquences centromériques, nucléée par l'action du RNAi, est à la fois nécessaire (Volpe *et al.*, 2003) et suffisante (Schramke et Allshire, 2003) au recrutement de la cohésine Rad21 sur la chromatine centromérique. Cette protéine, dont la localisation aux centromères dépend de Swi6, est responsable de la cohésion des centromères des deux chromatides sœurs avant l'anaphase (Bernard *et al.*, 2001). Ce rôle de l'hétérochromatinisation des centromères dans la cohésion des chromatides explique vraisemblablement les anomalies de la ségrégation des chromosomes, à la mitose, chez les mutants *ago1*, *dcr1*, et *rdp1* (Volpe *et al.*, 2003).

RNAi et méthylation de l'ADN chez les Plantes

Chez les Plantes, de nombreux exemples de répression post-transcriptionnelle dépendant de l'homologie (appelée : PTGS) ont été décrits depuis la fin des années 1980 ; il apparaît à présent que le PTGS est une interférence à ARN (Cogoni et Macino, 2000 ; Hannon, 2002 ; Matzke et Matzke, 2004).

Plusieurs cas de PTGS sont associés à une méthylation du gène-cible de cette répression (Ingelbrecht *et al.*, 1994 ; Baulcombe, 1996b ; Sijen *et al.*, 1996 ; Smith *et al.*, 1996 ; Wassenegger et Péliissier, 1998 ; Mette *et al.*, 1999). La relation entre la méthylation des gènes impliqués et leur répression post-transcriptionnelle, a fait débat : la méthylation de l'ADN pouvait perturber sa transcription, et aboutir à la production d'ARN dits aberrants (par exemple, tronqués), qui seraient reconnus en tant que tels par la machinerie du PTGS, et dégradés (la méthylation de l'ADN serait alors une cause du PTGS) (Baulcombe, 1996b). Alternativement, la méthylation de l'ADN pourrait n'être qu'une conséquence du PTGS (Wassenegger et Péliissier, 1998). En effet, chez *Neurospora crassa*, un mutant incapable de méthyliser les cytosines est quand même réprimé, et, chez les Plantes, on connaît de nombreux exemples de PTGS sans méthylation d'ADN, et une souche de Tabac qui manifeste un PTGS pour le gène *npt II* transcrit le gène au même niveau que les souches réprimées. Cette observation rappelle celle de Van Blokland et coll., qui avaient montré que le gène *chs A*, réprimé post-transcriptionnellement, était transcrit en entier (Van Blokland *et al.*, 1994).

Depuis, il a été établi que les gènes impliqués dans le RNAi sont nécessaires à la méthylation *de novo* d'au moins deux loci (Chan *et al.*, 2004) chez *Arabidopsis thaliana*, et les siRNA sont bien capables de guider la méthylation de séquences d'ADN homologues chez *A. thaliana* et

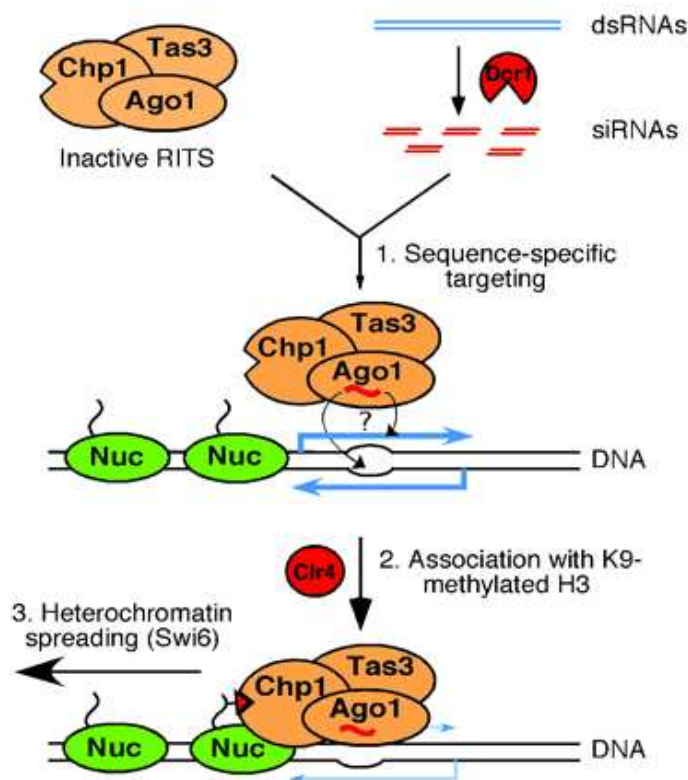


FIG. 1.18 – RNAi ET GUIDAGE DE L'HÉTÉROCHROMATINISATION DES RÉPÉTITIONS CENTROMÉRIQUES. Figure tirée de Verdel *et al.*, 2004. Le clivage, par Dicer (Dcr1), de longs ARN double brin issus de la transcription des répétitions centromériques, libère des siRNA qui peuvent guider le complexe RITS (représenté en orange; le complexe dit « inactif » est constitué des protéines de RITS, sans son composant ARN) sur l'ADN des répétitions centromériques, ou les transcrits naissants de ces régions (flèches bleues). L'association du complexe RITS à la chromatine recruterait la protéine Clr4, laquelle méthyle les lysines 9 des histones H3 localement; cette hétérochromatinisation se propagerait ensuite aux séquences avoisinantes, par un mécanisme impliquant la protéine Swi6.

Nicotiana benthamiana (Jones *et al.*, 1999 ; Aufsatz *et al.*, 2002). En particulier, un transgène transcrit en ARN susceptible de se replier en une longue tige-boucle (et qui produit de petits ARN d'environ 23 nt) provoque la méthylation d'un transgène homologue sur les cytosines dites « symétriques » (c'est à dire : les cytosines des dinucléotides CpG et des trinucléotides CpNpG) et sur les cytosines dites « asymétriques » (les cytosines des autres contextes de séquence). Lorsque, après reproduction sexuée, le transgène répresseur (transcrit en un ARN qui se replie en tige-boucle) et le transgène-cible ségrègent, la méthylation de ce dernier est perdue sur les sites asymétriques, mais maintenue sur les sites symétriques, reflétant probablement l'action d'une méthyl-transférase de maintien. La production des petits ARN semble donc nécessaire pour la méthylation *de novo* de la séquence-cible. La méthyl-transférase *de novo* responsable de l'activité catalytique serait DRM1 ou DRM2 (Cao *et al.*, 2003).

RNAi et réarrangements chromosomiques chez les Ciliés

Les Ciliés sont des Protozoaires qui possèdent deux types de noyaux : le micronoyau est diploïde, et transcriptionnellement inactif, alors que le macronoyau, polypléide (il contient plu-

sieurs centaines de copies du génome), est transcriptionnellement actif. Après chaque événement de reproduction sexuée (par conjugaison avec un autre individu, ou par autoconjugaison ; les partenaires échangent des noyaux haploïdes, issus de la méiose de leur micronoyau), un noyau diploïde dit « zygotique », issu de la fusion de deux noyaux haploïdes, se divise par mitose ; l'un des noyaux issus de cette mitose devient le nouveau micronoyau de la cellule, et l'autre se différencie en macronoyau (selon les Ciliés, le nombre de mitoses du noyau zygotique peut varier, et avec lui, le nombre de micronoyaux et de macronoyaux par cellule). L'ancien macronoyau est dégradé pendant les cycles cellulaires suivants. Au cours de la différenciation du macronoyau, chaque chromosome subit de nombreuses réplifications, les chromosomes sont fragmentés, et des télomères ajoutés aux nouvelles extrémités (les chromosomes macronucléaires sont donc plus courts que les chromosomes micronucléaires), et de nombreuses séquences sont excisées, et leurs régions flanquantes, raboutées. Ces séquences, spécifiquement micronucléaires, sont appelées « IES », pour : *internal eliminated sequences*. Alors que la fragmentation des chromosomes est relativement imprécise (les zones de fragmentation sont définies à quelques centaines de paires de bases près), l'excision des IES est précise au nucléotide près. Le déroulement de la conjugaison des Ciliés est schématisé sur la figure 1.19.

Lorsqu'un plasmide portant la séquence d'une IES est introduit dans le macronoyau parental, cette IES n'est pas excisée du macronoyau en développement (Duharcourt *et al.*, 1995 ; Chalker et Yao, 1996), et réciproquement, la délétion dans le macronoyau parental d'une séquence habituellement retenue dans le macronoyau provoque son élimination dans le macronoyau en développement (Forney *et al.*, 1996). Il s'établit donc une communication entre l'ancien macronoyau et le macronoyau en développement, qui influence le choix des séquences éliminées de manière spécifique. Une telle spécificité de séquence fait probablement intervenir des appariements entre acides nucléiques : il avait donc été proposé que des ARN pourraient diffuser entre l'ancien macronoyau et le macronoyau en développement, et participer à la sélection des séquences éliminées en comparant les séquences des deux macronoyaux (Duharcourt *et al.*, 1995 ; Meyer *et al.*, 1997 ; Meyer et Garnier, 2002). Effectivement, de petits ARN (d'environ 28 nt) homologues de séquences spécifiquement micronucléaires sont détectés au cours du développement macronucléaire chez le Cilié *Tetrahymena* ; leur accumulation dépend d'une protéine de la famille PPD (cette famille doit son nom aux domaines *PAZ* et *Piwi* ; ses membres sont impliqués dans le RNAi dans une grande variété d'espèces), appelée *Twilp* (Mochizuki *et al.*, 2002). Lorsque le gène *Twil1* est délété, *Tetrahymena* n'élimine pas les IES du macronoyau en développement, et fragmente peu les chromosomes du futur macronoyau ; les petits ARN ne sont pas détectés dans le mutant *Twil1*. Ces observations, associées à celle du comportement de la protéine *Twilp* au cours du développement macronucléaire (cytoplasmique au début du processus, elle commence par se localiser dans l'ancien macronoyau, puis dans le macronoyau en développement), suggèrent que *Twilp* et les petits ARN participent, ensemble, à une comparaison des deux macronoyaux, et que leur action se traduit par la délétion, dans le macronoyau en développement, des séquences absentes de l'ancien macronoyau (par exemple, ces petits ARN pourraient provenir de transcrits du génome micronucléaire, et dans l'ancien macronoyau, les ARN complémentaires de séquences macronucléaires seraient dégradés, et les ARN restants, donc complémentaires des séquences spécifiquement micronucléaires, pourraient guider l'élimination, dans le macronoyau en développement, des séquences homologues) (Mochizuki *et al.*, 2002). Ces petits ARN sont sensiblement plus longs que les siRNA observés chez les autres espèces, mais ils ont des extrémités (5' phosphorylée, 3' hydroxylée) typiques des produits de clivage de Dicer. Simultanément, un autre laboratoire montrait que la chromatine des séquences spécifiquement micronucléaires est modifiée (la lysine 9 de l'histone H3 est méthylée) au début du développement macronucléaire, et que le ciblage de cette modification épigénétique sur une séquence donnée semble suffisant pour en promouvoir l'excision dans le futur macronoyau (Taverna *et al.*, 2002).

Il a ensuite été démontré que l'injection, dans le cytoplasme de cellules de *Tetrahymena* après

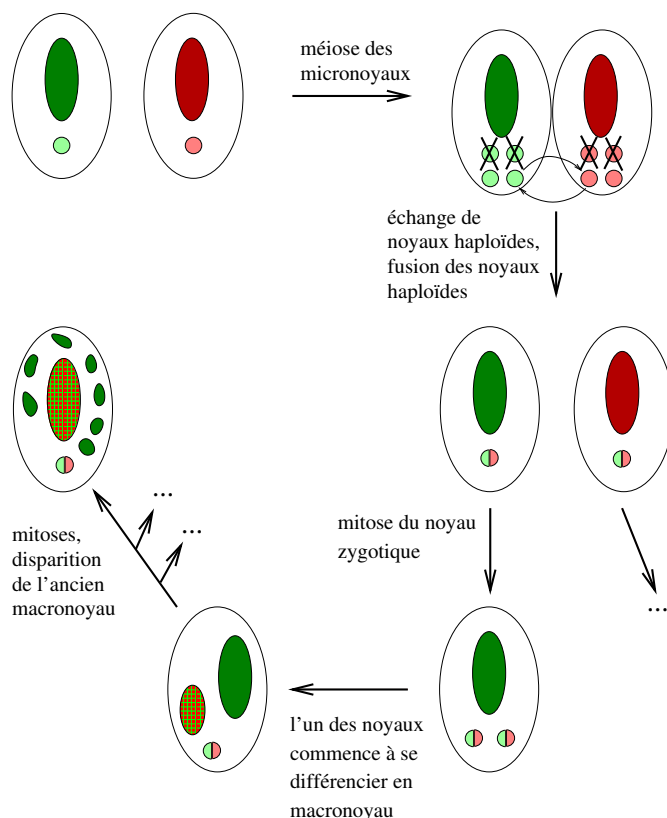


FIG. 1.19 – LA CONJUGAISON DES CILIÉS. Les micronoyaux sont représentés en teinte claire (vert clair, et rouge clair), et les macronoyaux en teinte sombre (vert sombre et rouge sombre). Le micronoyau de chaque partenaire subit une méiose, et deux des noyaux haploïdes résultants sont dégradés ; l'un des deux autres est transmis à l'autre partenaire, et le dernier fusionne avec le noyau reçu de l'autre partenaire. Le noyau diploïde ainsi constitué (« noyau zygotique ») subit un nombre variable de mitoses, selon les espèces (sur ce schéma, il n'en subit qu'une), et la moitié des noyaux formés deviendra des micronoyaux, l'autre moitié se différenciera en macronoyaux (par polyploïdisation, fragmentation des chromosomes et ajout de télomères, et élimination des IES ; voir texte). *N.B.* : selon les Ciliés, le nombre de micro- et de macronoyaux par cellule peut varier, en raison de différences entre les nombres de duplications entre chaque étape.

conjugaison, de longs ARN double brin provoquait l'excision des séquences homologues dans le macronoyau en développement Yao *et al.*, 2003, et que l'abolition de la méthylation de la lysine 9 de l'histone H3 était épistatique sur la production des petits ARN (Liu *et al.*, 2004c). Il semble donc que, dans la suite d'événements qui conduit à la sélection des séquences éliminées, les petits ARN jouent un rôle plus précoce que celui de la méthylation de la lysine 9 de l'histone H3 ; il est donc probable que, comme chez la levure à fission (voir p. 80), de petits ARN produits par Dicer guident cette modification de la chromatine ; ensuite, les séquences portant des histones H3 méthylées sur la lysine 9 seraient excisées (Liu *et al.*, 2004c).

Le modèle de comparaison des deux macronoyaux par l'intermédiaire de petits ARN est présenté en figure 1.20, page 87.

Il faut toutefois noter que la comparaison entre les deux macronoyaux (ancien et nouveau) n'est pas le seul déterminant de la sélection des séquences excisées, puisque les expériences de guidage artificiel d'excisions par le ciblage de la méthylation de la lysine 9 de l'histone H3 (Taverna *et al.*, 2002), ou par des ARN double brin (Yao *et al.*, 2003), mettent bien en évidence

l'élimination spécifique des séquences visées, mais les bornes de ces excisions sont moins précises que dans le cas de l'élimination des IES naturelles. Le contexte de séquence des IES pourrait donc influencer leur excision, ce qui expliquerait quelques difficultés du modèle de guidage par les petits ARN (Mochizuki *et al.*, 2002).

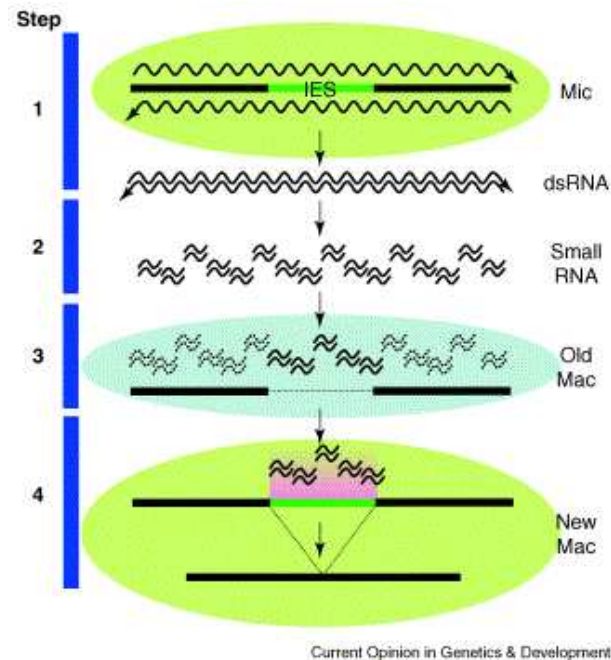


FIG. 1.20 – MODÈLE DE SÉLECTION DES SÉQUENCES ÉLIMINÉES LORS DU DÉVELOPPEMENT MACRONUCLÉAIRE DE *Tetrahymena*. Figure tirée de Mochizuki et Gorovsky, 2004. Le génome du micronoyau (représenté par l'ellipse vert pomme, en haut) est transcrit à partir des deux brins ; les ARN double brins résultants (flèches noires en zig-zag) seraient clivés en siRNA, et rejoindraient l'ancien macronoyau (ellipse vert pastel). Tous les siRNA homologues de séquences macronucléaires (ces séquences d'ADN sont représentées par un trait noir épais) seraient dégradés (ils sont représentés par des zig-zags en pointillés), et seuls les siRNA homologues des séquences éliminées (ces séquences d'ADN sont représentées par un trait en pointillés) se maintiendraient (ils sont représentés par des zig-zags en trait plein). Ils migreraient ensuite dans le macronoyau en développement (représenté par l'ellipse vert pomme, en bas), où ils s'apparieraient aux séquences spécifiquement micronucléaires (représentées en vert vif) ; cette reconnaissance aboutirait à une hétérochromatinisation des séquences spécifiquement micronucléaires, avant qu'un système spécialisé n'élimine les séquences hétérochromatinisées.

Un phénomène très similaire semble prendre place chez un autre Cilié, la Paramécie, puisqu'un traitement aux ARN double brin (exprimés par les bactéries dont se nourrissent les paramécies) provoque la délétion imprécise des séquences homologues (Garnier *et al.*, 2004). De petits ARN sont également détectés chez la Paramécie, mais leur taille est différente (ils font principalement 22 à 23 nt) et ils sont aussi observés pendant la vie végétative, où ils semblent responsables de répressions par RNAi, alors que les petits ARN de *Tetrahymena* ne sont détectés que pendant la conjugaison. Les ARN de 22 à 23 nt de la Paramécie ne sont donc pas forcément les effecteurs de la comparaison entre les deux macronoyaux : ce pourraient être des siRNA simplement capables de guider des dégradations d'ARN complémentaires par RNAi. Une autre classe de petits ARN réaliserait la comparaison des génomes de l'ancien et du futur macronoyau, comme chez *Tetrahymena*. Dans le cadre de cette hypothèse, il est tentant d'imaginer que l'acide

nucléique avec lequel s'hybrident ces petits ARN dans l'ancien macronoyau n'est pas l'ADN, mais plutôt les ARN transcrits dans l'ancien macronoyau, qui pourraient représenter la totalité de sa séquence. Dans ce cas, les siRNA de 22 à 23 nt, capables de guider la dégradation des ARN de l'ancien macronoyau, pourraient dépléter l'ancien macronoyau des ARN homologues, qui ne seraient donc pas perçus comme séquences macronucléaires, et dont les séquences seraient excisées du macronoyau en développement.

Ce modèle expliquerait une observation surprenante : alors que l'introduction d'un nombre limité de copies d'une séquence donnée dans l'ancien macronoyau provoque sa rétention dans le macronoyau en développement (*cf* plus haut), l'introduction d'un grand nombre de copies a l'effet inverse, elle induit l'élimination des séquences homologues dans le macronoyau en développement. Or, seules les séquences introduites en grand nombre de copies provoquent la production de siRNA, et répriment les gènes homologues, en croissance végétative. Ainsi, dans ce modèle, les séquences réprimées post-transcriptionnellement en croissance végétative seraient, en plus, excisées du génome macronucléaire après la conjugaison suivante (Garnier *et al.*, 2004).

Implications génétiques du RNAi dans l'hétérochromatinisation chez les Métazoaires

Chez la *Drosophile*, trois mutations qui affectent la maturation des siRNA (des mutations sur les gènes *aubergine/sting*, *spindle-E/homeless* et *piwi*) provoquent également la délocalisation de la protéine de l'hétérochromatine HP1, et une réduction de la méthylation de la lysine 9 de l'histone H3 (Pal-Bhadra *et al.*, 2004). Cette délocalisation se traduit notamment par la perte de répression d'un gène rapporteur *white* inséré dans la chromatine péricentromérique. La quantité globale de protéine HP1 n'est pas affectée, donc il semble que l'adressage de la protéine vers ses séquences-cibles, plus que sa stabilité, dépende de la machinerie du RNAi. De plus, la mutation du gène *piwi* dérèprime un transgène normalement réprimé transcriptionnellement, sans que la base moléculaire de cette dérèpression soit connue (en particulier, les modifications des histones n'ont pas été analysées) (Pal-Bhadra *et al.*, 2002).

Chez les Vertébrés, des siRNA pourraient également diriger l'hétérochromatinisation de séquences homologues : des cellules hybrides Poulet-Homme (ce sont des cellules DT40 de Poulet, dans lesquelles un chromosome 21 humain a été introduit par micromanipulation), dans lesquelles le gène Dicer est sous le contrôle d'un promoteur répressible par la tétracycline, présentent de graves défauts de ségrégation des chromosomes (qui concernent aussi bien le chromosome 21 humain que les chromosomes de Poulet) après traitement à la tétracycline (Fukagawa *et al.*, 2004). Ces défauts semblent dus tout à la fois à une séparation précoce des chromatides sœurs, et à une anomalie du « point de contrôle de la mitose » (système de contrôle qui inhibe la transition en anaphase tant que tous les chromosomes ne sont pas alignés sur la plaque métaphasique, chaque chromatide reliée au fuseau mitotique par les deux faces du kinétochore). Comme chez les mutants du RNAi de *S. pombe* (voir p. 83), la séparation précoce des chromatides est la conséquence d'une délocalisation de la cohésine Rad21, qui est diffuse dans le noyau des cellules privées de Dicer, et concentrée aux centromères dans les cellules non traitées à la tétracycline. Quant au défaut du point de contrôle mitotique, il pourrait être dû à une mauvaise reconnaissance des centromères par le complexe du point de contrôle, dont l'un des constituants, BubR1, n'est pas retrouvé sur les centromères des chromatides séparées prématurément (sa localisation habituelle, dans des cellules sauvages).

Les auteurs proposent que les séquences répétées des centromères (les satellites α 21-I et α 21-II, sur le chromosome humain 21) sont faiblement transcrites, à partir des deux brins, et que l'ARN double brin résultant soit digéré par Dicer en siRNA qui guideraient ensuite l'hétérochromatinisation de leur locus ; que le modèle qui avait été proposé pour la levure à fission (voir page 81) s'appliquerait donc chez les Vertébrés. Et effectivement, des transcrits sens et antisens des satellites α (à peine détectables en conditions normales) s'accumulent en

l'absence de Dicer (Fukagawa *et al.*, 2004). Mais les petits ARN prédits ne sont pas détectés de façon convaincante en conditions normales, et le signal que les auteurs leur attribue est incomparablement plus intense en l'absence de Dicer (les auteurs suggèrent que ces ARN sont issus d'une activité résiduelle de Dicer).

En revanche, un transgène qui permet l'expression de longs ARN double brin contre un gène-cible, dans les ovocytes de Souris, n'en guide pas la méthylation (Svoboda *et al.*, 2004b). Ce résultat n'est peut-être pas généralisable à tous les types cellulaires de Vertébrés, mais les auteurs font remarquer que l'ovocyte de Souris manifeste une activité de méthylation de l'ADN *de novo* (responsable notamment de la mise en place des empreintes maternelles sur l'ADN du futur gamète : voir page 26), et que les conditions étaient donc favorables à ce que le gène-cible soit méthylé. Il ne leur a pas été possible d'étudier l'effet de la transcription de l'ARN double brin sur la méthylation en *cis* de son propre gène pour des raisons techniques. Il reste donc possible que, dans ce système, les ARN bicaténaires guident une méthylation de l'ADN, mais que cet effet ne puisse avoir lieu qu'en *cis*, sur leur propre gène.

Dans un système cellulaire en culture, un groupe indépendant a confirmé que la transcription d'ARN en tige-boucle ne guide pas la méthylation d'un promoteur homologue chez les Vertébrés ; mais un traitement qui facilite l'accès des acides nucléiques au noyau lui permet de réprimer la transcription contrôlée par ce promoteur avec une grande efficacité, et cette répression s'accompagne d'une méthylation du promoteur (Morris *et al.*, 2004). Il semble donc que les petits ARN ont la capacité intrinsèque de guider la méthylation de séquences d'ADN dans ce système, mais que leur localisation cytoplasmique les en empêche.

Cependant, un autre laboratoire rapporte un cas de méthylation d'ADN guidé par des siRNA homologues, dans des cellules humaines, sans qu'un traitement particulier leur facilite l'accès au noyau (Kawasaki et Taira, 2004). Dans ce système (des cellules MCF-7 — ce sont des cellules épithéliales humaines tumorales — et des cellules épithéliales humaines saines en culture), la méthylation guidée par les siRNA nécessite à la fois la présence de la méthyl-transférase de maintien Dnmt1, et la méthyl-transférase *de novo* Dnmt3b.

Régulation du RNAi

Alors que de nombreuses mutations abolissent le RNAi (c'est ainsi qu'ont été identifiés la plupart des acteurs du RNAi), peu de mutations l'activent. C'est toutefois le cas des deux mutations décrites par Dehio et Schell en 1994, qui permettent au RNAi de réprimer un transgène dans les graines d'*Arabidopsis thaliana*, alors que les graines sauvages ne répriment pas le transgène (Dehio et Schell, 1994). Les deux loci identifiés dans ce crible, baptisés *egs1* et *egs2* (pour : *enhancer of gene silencing*) n'ont cependant pas été caractérisés plus en détail.

Chez le Nématode, les cellules nerveuses sont réfractaires au RNAi : la plupart des ARN_m ne sont pas dégradés dans les neurones, après un traitement aux ARN double brin. Un mutant qui réalise un RNAi efficace dans les neurones a été isolé : le gène sur lequel il est muté, appelé *eri-1* (pour : *enhanced RNAi*), code une nucléase qui clive spécifiquement les siRNA *in vitro* (Kennedy *et al.*, 2004). Plusieurs Vertébrés, et la levure à fission, possèdent un orthologue de *eri-1*, et la protéine humaine clive elle aussi les siRNA *in vitro*. L'expression de *eri-1* est régulée, chez le Nématode : il est principalement exprimé dans quelques neurones et dans la spermathèque (un autre organe où le RNAi est peu efficace), mais aussi, faiblement, dans le reste de l'organisme, ce qui explique probablement ses effets ubiquitaires sur certains gènes (Kennedy *et al.*, 2004).

Une autre protéine, RRF-3, inhibe le RNAi dans les neurones du Nématode, et curieusement, elle appartient à une famille de polymérases à ARN dépendant de l'ARN (RdRP) putatives, alors que des RdRP sont nécessaires au processus de RNAi chez plusieurs espèces (Simmer *et al.*, 2002). Les auteurs proposent que cette protéine entre en compétition avec des RdRP impliquées dans le RNAi, pour l'association à des partenaires communs.

Enfin, des observations étonnantes suggèrent que les déaminases à adénine de l'ARN (qui convertissent les adénines en inosines dans les ARN double brin) pourraient réguler également le RNAi chez le Nématode (et, à nouveau, ces protéines sont fortement exprimées dans les neurones) : des transgènes (transcrits en ARN qui se replie en longues tiges-boucles) deviennent sensibles au RNAi lorsque les deux déaminases à adénine de l'ARN du Nématode sont mutées (Knight et Bass, 2002) ; et la mutation de gènes nécessaires au RNAi sauve le phénotype de mutants des déaminases à adénine de l'ARN (Tonkin et Bass, 2003). Il est possible que les déaminases, en convertissant les adénines en inosines, détruisent la structure bicaténaire de leurs substrats, et donc, en empêchent le clivage par Dicer et ses partenaires (Tonkin et Bass, 2003).

1.5 ARN non-codants et contrôle transcriptionnel

1.5.1 ARN non-codants et compensation de dose

Le nombre de copies des hétérosomes (ou : chromosomes sexuels) présente la particularité de dépendre du sexe de l'individu : ainsi, chez les Oiseaux, les mâles portent deux chromosomes sexuels homologues, appelés « Z » (les mâles sont dits « homogamétiques »), et les femelles, un chromosome Z et un chromosome appelé « W » (elles sont dites « hétérogamétiques »), alors que, par exemple, chez les Insectes, les mâles, hétérogamétiques, possèdent un chromosome appelé « X » et un chromosome appelé « Y » et les femelles, homogamétiques, possèdent deux chromosomes X. Chez les Mammifères également, les femelles sont homogamétiques (leurs chromosomes sexuels sont aussi notés « X »), alors que les mâles sont hétérogamétiques (et leurs chromosomes sexuels sont aussi notés « X » et « Y »). Des mécanismes de régulation, appelés « compensation de dose », normalisent l'expression des gènes liés aux chromosomes sexuels dans les deux sexes chez des espèces animales diverses (voir Meller, 2000 pour revue) : dans un système particulièrement étudié (l'inactivation du chromosome X chez les femelles de Mammifères), l'un des deux chromosomes X est inactivé dans sa globalité (à l'exception de quelques gènes — voir plus bas — tous les gènes de ce chromosome sont transcriptionnellement réprimés), si bien que les femelles, au même titre que les mâles, ne possèdent qu'un chromosome X actif. Un autre cas a été relativement bien décrit : l'hyperactivation du chromosome X des mâles de *Drosophile*. Dans cet exemple, les gènes de l'unique chromosome X des mâles sont hypertranscrits, et de nombreux gènes portés par les chromosomes X des femelles seraient de plus réprimés post-transcriptionnellement (Kelley *et al.*, 1995). Finalement, mâles et femelles de *Drosophile* expriment les gènes liés au chromosome X à des taux comparables.

Ces deux phénomènes, l'inactivation d'un chromosome X chez les femelles de Mammifères, et l'hyperactivation du chromosome X chez les mâles de *Drosophile*, font intervenir des ARN non-codants, qui modulent l'activité transcriptionnelle de chromosomes entiers, en contrôlant la mise en place de marques épigénétiques spécifiques²⁸.

Les ARN non-codants Xist et Tsix et l'inactivation du chromosome X chez les Mammifères

Chez les femelles de Mammifères, l'un des deux chromosomes X adopte une structure chromatinienne très particulière, qui lui donne un aspect reconnaissable en microscopie (cette particule dense, même pendant l'interphase, avait été décrite dès 1959 : Barr, 1959, et est connue sous le nom de « corps de Barr »), et qui inhibe la transcription de tous les gènes du chromosome. Tous ? Non. Plusieurs gènes portés par le chromosome X sont insensibles à la compensation de dose (notamment dans les « régions pseudoautosomales », régions communes aux chromosomes X et Y) (voir Brown et Greally, 2003, pour revue).

Alors que les cellules du placenta²⁹ des embryons femelles de Souris inactivent systématiquement le chromosome X d'origine paternelle (Takagi et Sasaki, 1975), les cellules des embryons femelles de Souris, et les cellules des embryons femelles humains, et de leurs tissus extra-embryonnaires (Migeon et Do, 1979), inactivent aléatoirement le chromosome X d'origine paternelle ou maternelle (Lyon, 1961). Chez les Marsupiaux en revanche, les cellules des embryons femelles inactivent toujours le chromosome X paternel (Cooper *et al.*, 1971 ; Sharman,

²⁸La compensation de dose n'implique cependant pas systématiquement des ARN non-codants : ainsi, chez le Nématode *Caenorhabditis elegans*, la transcription sur chacun des deux chromosomes X des hermaphrodites XX est divisée par deux, en comparaison à la transcription de l'unique chromosome X des mâles XY ou XO. Le complexe responsable de la répression transcriptionnelle chez les hermaphrodites est constitué uniquement de protéines (voir la revue de Meller, 2000).

²⁹N.B. : le placenta dérive du trophoctoderme qui, tout comme l'embryon, est issu de l'œuf fécondé ; il est donc isogénique avec l'embryon.

1971). Ainsi, chez les Marsupiaux et dans le placenta de Souris, l'inactivation du chromosome X (qui dépend de l'origine parentale du chromosome) est un phénomène d'empreinte génomique parentale (voir section 1.2), alors que dans les embryons (puis tout au long de la vie de l'individu) des Mammifères placentaires, un mécanisme différent est à l'origine du choix, arbitraire, du chromosome X qui sera inactivé.

Plus précisément, il apparaît que le chromosome X paternel est systématiquement inactivé, même dans l'embryon, au stade pré-implantatoire, mais que cette empreinte est perdue spécifiquement dans les cellules de l'embryon pendant la maturation du blastocyste (Mak *et al.*, 2004 ; Okamoto *et al.*, 2004) ; le chromosome X inactif est ensuite choisi aléatoirement dans chaque cellule de l'embryon.

L'étude de translocations, qui fusionnaient des séquences du chromosome X avec d'autres chromosomes, a mis en évidence une région régulatrice, appelée Xic (pour : *X inactivation center*, centre d'inactivation du X, en anglais), nécessaire (Russell, 1963) et suffisante (Eicher, 1970) pour nucléer l'inactivation, qui peut ensuite se propager aux séquences adjacentes. La définition de ce centre d'inactivation a été progressivement précisée, grâce à l'analyse de nombreux réarrangements (voir par exemple Brown *et al.*, 1991b), en particulier dans un système cellulaire qui récapitule le phénomène d'inactivation (Rastan et Robertson, 1985). Le Xic minimal est une région de 80 kb, décrite en 1999 (Lee *et al.*, 1999b). Le « comptage » (évaluation du nombre de chromosomes X par jeu d'autosomes, responsable de la restriction de l'inactivation du chromosome X aux cellules femelles) implique également des séquences de Xic : des cellules qui ne possèdent qu'un chromosome X peuvent inactiver cet unique chromosome s'il est délété pour une partie du locus Xic, alors qu'il reste actif si Xic est intact (Clerc et Avner, 1998 ; Morey *et al.*, 2004).

Des séquences localisées dans Xic affectent la probabilité, pour un chromosome X donné, d'être choisi pour l'inactivation. Ces séquences constituent l'« élément de contrôle du X » (*Xce*, pour : *X controlling element*), dont il existe plusieurs allèles, dont les influences sur le choix du chromosome qui les porte sont différentes : alors que les cellules homozygotes pour *Xce* inactivent l'un ou l'autre de leurs chromosomes X avec des probabilités très voisines (environ 0,5 pour chacun, donc), les cellules hétérozygotes vont préférentiellement inactiver l'un des deux (son allèle *Xce* est dit « faible »), et maintenir l'autre chromosome X actif (son allèle *Xce* est dit « fort ») (Rastan, 1982).

Le locus Xic héberge de plus le gène d'un ARN non-codant, qui est exprimé à partir du chromosome X inactif uniquement : cet ARN a été baptisé « Xist », pour : *X inactive-specific transcript*, transcrit spécifique de l'X inactif, en anglais (Brown *et al.*, 1991a ; Borsani *et al.*, 1991 ; Brockdorff *et al.*, 1991). Cet ARN non-codant est épissé, long de 15 (chez la Souris) à 17 kb (chez l'Homme), et s'accumule dans le noyau, sur le chromosome X inactif (Brockdorff *et al.*, 1992 ; Brown *et al.*, 1992 ; Clemson *et al.*, 1996). Les cellules souches embryonnaires (cellules ES) femelles de Souris reproduisent l'inactivation aléatoire du chromosome X au cours de leur différenciation (Rastan et Robertson, 1985). Lorsque l'un des deux allèles de *Xist* de ces cellules est délété, si un chromosome est inactivé, après différenciation des cellules ES, c'est systématiquement le chromosome portant l'allèle sauvage de *Xist* ; certaines cellules n'inactivent aucun chromosome X (Penny *et al.*, 1996). Il semble donc que le gène *Xist* agisse en aval de *Xce*, puisque la délétion de *Xist* n'affecte pas la probabilité de son chromosome d'être choisi pour l'inactivation, elle ne fait qu'empêcher l'inactivation si le chromosome qui la porte a été choisi (Penny *et al.*, 1996). *In vivo*, un phénomène analogue est observé (Marahrens *et al.*, 1997) : les mâles dont l'unique allèle *Xist* est une délétion sont phénotypiquement sauvages, et fertiles. Des femelles ayant hérité la délétion par leur mère sont phénotypiquement sauvages, mais une analyse plus fine montre que leur chromosome X inactif est toujours le chromosome X hérité de

leur père (le chromosome X maternel, délété de *Xist*, semble donc incapable de s'inactiver)³⁰. Enfin, des embryons femelles ayant hérité de la délétion de *Xist* par leur père meurent en début de gestation ; dans l'embryon, toutes les cellules observées inactivent le chromosome X maternel, et la létalité de ces embryons semble due à la dérégulation des gènes liés au chromosome X dans le placenta : le chromosome X paternel (habituellement inactivé dans toutes les cellules du placenta, *cf* plus haut) y est actif, au même titre que le chromosome X maternel, et ce défaut d'empreinte génomique parentale des gènes du chromosome X dans le placenta (qui aboutit à une surexpression, d'un facteur 2, de tous les gènes liés au chromosome normalement soumis à l'empreinte) perturbe vraisemblablement la fonction du placenta dans le développement de l'embryon (Marahrens *et al.*, 1997). La démonstration vient de l'observation que des souris femelles $X_{pat}.O_{mat.}$ (c'est à dire : ne possédant qu'un chromosome sexuel, un X, d'origine paternelle) dont l'unique chromosome sexuel est délété de *Xist*, se développent normalement : leur placenta n'exprime qu'un allèle des gènes liés au chromosome X. Ainsi, l'expression de l'ARN *Xist* à partir d'un chromosome X donné, ou le gène *Xist* lui-même, est nécessaire à l'inactivation de ce chromosome X, que ce soit par inactivation aléatoire (dans l'embryon), ou par empreinte génomique parentale (dans le placenta).

Chez la Souris, un ARN antisens à *Xist*, appelé *Tsix*, semble réguler l'activité de *Xist* : *Tsix* est un autre ARN non-codant, d'environ 40 kb, transcrit à partir d'un promoteur situé 15 kb en aval de *Xist* (Lee *et al.*, 1999a). Avant la différenciation des cellules ES femelles, *Xist* et *Tsix* sont tous deux exprimés à partir de chacun des deux chromosomes X, puis l'expression de *Tsix* devient monoallélique, et l'ARN *Xist* devient métaboliquement plus stable (Panning *et al.*, 1997 ; Sheardown *et al.*, 1997), et se propage le long du chromosome qui n'exprime pas *Tsix* : ce chromosome, couvert d'ARN *Xist* en moins d'un cycle cellulaire, devient le chromosome X inactif (Clemson *et al.*, 1996 ; Lee *et al.*, 1999a). Lorsque ce processus de sélection du chromosome X inactif est terminé, l'expression de *Tsix* s'interrompt également sur le chromosome X actif. Le comportement de ces deux ARN suggère que *Tsix* est un répresseur de l'activité de *Xist* chez la Souris : tant que le futur X inactif n'a pas été choisi, *Tsix*, exprimé à partir de chacun des deux chromosomes X, réprimerait la propagation de *Xist*, puis lorsqu'un événement (encore inconnu) sélectionne le futur chromosome X inactif, l'expression de *Tsix* à partir de ce chromosome s'arrêterait, laissant *Xist* le recouvrir, et en provoquer l'inactivation (Lee *et al.*, 1999a). Puisque la délétion du promoteur de *Tsix* (qui abolit l'expression de l'ARN *Tsix*) sur un chromosome X conduit, dans des cellules ES femelles, à l'inactivation systématique de ce chromosome (Debrand *et al.*, 1999 ; Lee et Lu, 1999), *Tsix* semble en effet responsable de la répression en *cis* de *Xist*. Et plus précisément, *Tsix* serait l'un des acteurs du mécanisme aboutissant au choix du chromosome X à inactiver, puisque cette délétion du promoteur de *Tsix* se comporte comme un allèle *Xce* infiniment faible : alors que *Xist* semble être un effecteur de l'inactivation du chromosome, qui agit en aval du processus de choix (*cf* plus haut), *Tsix* serait impliqué dans le choix. La force des allèles *Xce* semble corrélérer inversement avec la méthylation d'un îlot CpG dans lequel se trouve le promoteur de *Tsix* (Simmler *et al.*, 1993 ; Courtier *et al.*, 1995 ; Avner *et al.*, 1998 ; Lee et Lu, 1999). Des séquences non-codantes transcrites, baptisées *Xite* (pour : *X-inactivation intergenic transcription elements*), contrôlent l'expression de *Tsix*, et des polymorphismes de séquence de *Xite* semblent responsables des différences de force des allèles de *Xce* (Ogawa et Lee, 2003) : finalement, l'élément de contrôle *Xce* pourrait être le locus *Xite*, et les différences de force entre allèles de *Xce* pourraient s'expliquer par des différences d'expression de *Tsix* (Lee et Lu, 1999 ; Ogawa et Lee, 2003).

³⁰En revanche, ce travail *in vivo* suggère que *Xist* n'agit pas seulement en aval du mécanisme de choix, mais qu'il joue également un rôle dans le choix du chromosome X qui sera inactivé : le chromosome X qui porte la délétion n'est jamais choisi pour être inactivé (Marahrens *et al.*, 1998). La délétion étudiée dans ce travail *in vivo* n'est pas la même que celle étudiée par Penny *et al.*, 1996, dans les cellules ES : elle pourrait donc affecter un élément important dans le processus de choix, qui n'était pas délété chez Penny *et al.*, 1996.

Tsix est également impliqué dans la répression de *Xist* sur le chromosome X maternel des cellules du placenta de Souris, puisque la délétion du promoteur de *Tsix* sur le chromosome X maternel provoque l'inactivation du chromosome maternel (en plus du chromosome paternel), donc la répression biallélique des gènes liés au chromosome X dans le placenta, probablement responsable de la létalité de la presque totalité des embryons (Lee, 2000).

Le mode de répression de la propagation de *Xist* par *Tsix* reste énigmatique : le comportement des deux ARN, décrit par Lee *et al.*, 1999a, pourrait s'interpréter de plusieurs façons (l'ARN *Tsix* pourrait réprimer la propagation de l'ARN *Xist* en inhibant sa transcription, en accélérant sa dégradation, ou, en s'y appariant, en masquant des domaines fonctionnels de *Xist*). Cependant, l'extinction de *Tsix* ne précède pas toujours l'accumulation de l'ARN *Xist* (Debrand *et al.*, 1999). Une observation intéressante pourrait apporter quelques indices quant au mode de répression de *Xist* par *Tsix* : dans des cellules ES murines femelles non différenciées, une délétion qui inclut notamment le promoteur de *Tsix* provoque la délocalisation de l'ARN *Xist*, qui n'est plus alors concentré sur son site de transcription, mais plus diffus dans le noyau, et cette délocalisation s'accompagne d'une augmentation de la quantité globale d'ARN *Xist* (Morey *et al.*, 2001). L'ARN *Tsix* pourrait donc, d'une manière ou d'une autre, séquestrer l'ARN *Xist* à proximité de leur site de transcription. Ce confinement de *Xist* pourrait être nécessaire à sa répression, ou alors, ce pourrait être une conséquence de son association à l'ARN *Tsix*, elle-même requise pour la répression de *Xist*.

Enfin, toutes ces données sur le gène *Tsix* ont été collectées chez la Souris, et, alors qu'un orthologue humain de l'ARN *Tsix* est détecté, il semble incapable de réprimer *in cis* l'action de l'ARN *Xist*, puisque dans les embryons femelles humains, ces deux ARN sont transcrits à partir du même chromosome, le chromosome X inactif (Migeon *et al.*, 2002).

Le mécanisme de l'inactivation du chromosome X par l'ARN *Xist* reste également mystérieux. Il a été montré très tôt que le locus *Xic*, requis pour l'initiation de l'inactivation d'un des chromosomes X dans les cellules femelles, était dispensable pour le maintien de cet état inactif (Brown et Willard, 1994) ; donc, l'ARN *Xist*, même exprimé constitutivement sur le chromosome X inactif, ne semble plus jouer de rôle particulier après l'initiation de son inactivation. Des modifications épigénétiques, héréditaires par division cellulaire, semblent donc prendre le relais, et assurer une répression transcriptionnelle indépendante de l'ARN *Xist* (*cf* plus bas).

L'étude de délétions de transgènes *Xist* a montré que des séquences de l'ARN distinctes étaient nécessaires à son association à la chromatine du chromosome X inactif, et à la répression transcriptionnelle des gènes du chromosome (Wutz *et al.*, 2002). Notamment, des séquences répétées situées en 5' de l'ARN *Xist*, appelées « répétitions A », sont indispensables à la fonction de répression de *Xist* ; or (en raison de l'activité d'un promoteur alternatif en aval des répétitions A) la plupart des molécules *Xist* en sont dépourvues, si bien que seules 50 à 80 molécules actives (*i.e.* : contenant les répétitions A) par cellule suffisent à inactiver le chromosome entier, d'environ 150 Mb ; il se pourrait donc que chaque molécule d'ARN *Xist* active soit responsable de la répression d'environ 2 Mb du chromosome (Brockdorff, 2002). Et effectivement, l'ARN *Xist* s'associe à des loci particuliers sur le chromosome X inactif (visibles sur le chromosome métaphasique) (Duthie *et al.*, 1999). À partir de la détermination du chromosome X inactif, sa chromatine subit une série de modifications épigénétiques successives (Brockdorff, 2002). Pendant l'inactivation d'un chromosome X choisi aléatoirement (dans les cellules ES femelles de Souris), les événements les plus précoces semblent être une déméthylation de la lysine 4 de l'histone H3, et une déacétylation et une méthylation de la lysine 9 de l'histone H3 (Heard *et al.*, 2001 ; Boggs *et al.*, 2002 ; Mermoud *et al.*, 2002 ; Peters *et al.*, 2002) (qui s'accompagnent d'un retard à la réplication du chromosome X inactif), puis l'hypoacétylation de l'histone H4 (Keohane *et al.*, 1996), suivie du recrutement du variant d'histone macroH2A1.2 (Mermoud *et al.*, 1999). Le dernier événement semble être la méthylation des dinucléotides CpG (Keohane

et al., 1996). La plupart de ces modifications sont certainement interdépendantes, et en toute logique, la première d'entre elles au moins doit être guidée par l'ARN Xist (Brockdorff, 2002). À l'heure actuelle, le lien entre l'ARN Xist et ces modifications épigénétiques (qui fera probablement intervenir la fixation de complexes de remodelage de la chromatine sur les répétitions A de l'ARN Xist) n'a pas encore été établi.

Il semble que le recrutement par l'ARN Xist des protéines Eed et Ezh2³¹, au cours de l'initiation de l'inactivation du chromosome, provoque la méthylation de la lysine 27 de l'histone H3 ; cependant, dans l'embryon, cette modification épigénétique n'est pas nécessaire à l'inactivation du chromosome (Plath *et al.*, 2003).

Dans les tissus extra-embryonnaires de Souris, où l'inactivation du chromosome X est un phénomène d'empreinte génomique parentale qui dépend aussi de l'ARN Xist (voir page 93), les modifications épigénétiques qui affectent le chromosome X inactif ne sont pas rigoureusement les mêmes que dans l'embryon : il semble que les dinucléotides CpG n'y soient pas méthylés (Kratzer *et al.*, 1983 ; Sado *et al.*, 2000), que la réplication du chromosome X inactif soit précoce, et non tardive (Takagi *et al.*, 1982), enfin, le recrutement du variant d'histone macroH2A1.2 semble plus précoce, dans la série de modifications épigénétiques (Costanzi *et al.*, 2000). Les protéines du groupe polycomb Eed et Enx1 s'associent au chromosome X inactif, et cette association est dépendante de l'ARN Xist, et résiste à la mitose ; sur le chromosome métaphasique, les protéines se concentrent sur des loci bien particuliers, rappelant la distribution de l'ARN Xist (Mak *et al.*, 2002). Or, le gène *Eed* est requis, chez la Souris, pour la méthylation des histones du chromosome X inactif (Silva *et al.*, 2003) (tant dans l'embryon que dans le placenta) et le maintien de l'inactivation de ce chromosome dans les tissus extra-embryonnaires (Wang *et al.*, 2001a). Cependant, Eed n'est pas requise pour le maintien de l'inactivation du chromosome X dans l'embryon, ni pour l'initiation de l'inactivation du chromosome X dans les tissus extra-embryonnaires (Wang *et al.*, 2001a), ce qui indique que, si Eed devait être ce lien entre la propagation de l'ARN Xist et l'initiation des modifications épigénétiques du chromosome X inactif, alors d'autres systèmes, redondants, peuvent le remplacer dans un mutant *Eed*.

L'ARN roX et l'hypertranscription du chromosome X chez la Drosophile

Chez les mâles *Drosophila melanogaster*, il semble que deux mécanismes indépendants conduisent à la surexpression des gènes portés par l'unique chromosome X par rapport aux deux allèles des femelles : une répression post-transcriptionnelle des gènes portés par le chromosome X, chez les femelles (qui dépendrait d'un motif porté par la 3'UTR de leurs ARN_m), et une hypertranscription des gènes portés par le chromosome X des mâles (Kelley *et al.*, 1995). C'est ce dernier phénomène, qui fait intervenir un ARN non-codant et des modifications épigénétiques du chromosome X, qui sera exposé ici.

Le complexe ribonucléoprotéique responsable de l'hypertranscription est appelé « complexe MSL », pour *male-specific lethal*. Il est assemblé autour d'un ARN non-codant, appelé roX (pour : *RNA on the X chromosome*) (Meller *et al.*, 2000), qui peut exister sous deux isoformes, chacune transcrite à partir d'un gène propre : roX1, dont la forme majeure est longue de 3,7 kb, et roX2, dont la forme majeure est longue de 0,6 kb (Amrein et Axel, 1997 ; Kelley, 2004). Le complexe MSL contient de plus six protéines : l'acétyl-transférase de l'histone H4 MOF (pour : *males-absent on the first*) (Akhtar et Becker, 2000 ; Smith *et al.*, 2000), la kinase de l'histone H3 JIL1 (Wang *et al.*, 2001b), la protéine de fonction inconnue MSL1 (pour : *male-specific lethal*) (Palmer *et al.*, 1993), la protéine à domaine RING finger (domaine d'interaction avec

³¹Les orthologues de Eed chez la Drosophile et chez l'Homme ont une activité de méthyl-transférase sur les lysines n°3 et 27 de l'histone H3 (Cao *et al.*, 2002 ; Czermin *et al.*, 2002 ; Kuzmichev *et al.*, 2002 ; Müller *et al.*, 2002).

l'ADN) MSL2 (Bashaw et Baker, 1995 ; Kelley *et al.*, 1995 ; Zhou *et al.*, 1995), la protéine à chromodomaine MSL3 (Koonin *et al.*, 1995), et l'hélicase MLE (pour : *maleless*) (Kuroda *et al.*, 1991 ; Lee *et al.*, 1997).

Plusieurs de ces protéines ont des activités de modification d'histone qui peuvent rendre compte de l'activation transcriptionnelle des gènes-cibles. En revanche, le rôle de l'ARN non-codant roX est mal compris. Les deux isoformes, roX1 et roX2, sont interchangeables, et l'absence d'un seul des deux ne confère pas de phénotype mutant (Meller et Rattner, 2002). Le complexe MSL, à l'image de l'ARN Xist chez les Mammifères (voir ci-dessus), couvre le chromosome X des cellules mâles (Meller *et al.*, 2000), ainsi que quelques dizaines de loci sur les autosomes (Demakova *et al.*, 2003). L'absence d'ARN roX (lorsque les gènes *roX1* et *roX2* sont invalidés simultanément) est létale pour les mâles (Meller et Rattner, 2002), et au niveau cellulaire, elle dissocie les protéines du complexe MSL du chromosome X (Franke et Baker, 1999). Pourtant un sous-complexe MSL peut se former en l'absence d'ARN roX, et, si ses constituants protéiques sont surexprimés, de nombreux mâles viables sont obtenus (Oh *et al.*, 2003). Ainsi, il semblerait que l'ARN roX participe à la localisation du complexe MSL sur le chromosome X, et que son défaut puisse être sauvé en augmentant la quantité du complexe purement protéique.

Selon une hypothèse séduisante, l'ARN roX pourrait s'hybrider avec l'ADN des gènes *roX*, qui se trouvent tous les deux sur le chromosome X, et donner un point d'ancrage au complexe MSL sur le chromosome (Kelley, 2004). Ensuite, le remodelage de la chromatine pourrait se propager en *cis*, pour finalement concerner tout le chromosome. Il a en effet été démontré que, en l'absence des protéines MOF ou MSL3, les autres composants du complexe gardent la capacité de s'associer au chromosome X, mais seulement sur une trentaine de sites discrets (Lyman *et al.*, 1997 ; Kelley *et al.*, 1999 ; Gu *et al.*, 2000), comme si le complexe avait été correctement ciblé sur des sites d'entrée sur le chromosome, mais qu'il avait été incapable de se propager aux séquences flanquantes. Or, deux de ces sites sont précisément les gènes *roX* (Kelley *et al.*, 1999).

Cependant, l'adressage du complexe MSL sur le chromosome X apparaît bien moins prévisible que ce que ce modèle laisserait prévoir (Kelley, 2004). En particulier, insérer un gène *roX* du chromosome X dans un autosome ne permet qu'une association très limitée du complexe MSL vers le locus *roX* autosomal, et les délétions successives des gènes *roX* du chromosome X, qui augmentent progressivement la propagation du complexe aux séquences flanquantes du transgène autosomal, ne parviennent pas à déplacer massivement le complexe du chromosome X vers l'autosome (Park *et al.*, 2002b). Ainsi, même si le chromosome X ne porte plus de gène *roX*, il reste la cible préférentielle du complexe MSL. D'autre part, l'insertion d'un deuxième transgène *roX* autosomal abolit complètement la propagation du complexe MSL à partir du premier transgène autosomal (Park *et al.*, 2002b).

Il semble donc qu'une propriété particulière du chromosome X lui permette d'attirer le complexe MSL soluble, assemblé en *trans* sur un transgène autosomal. De plus, le promoteur des gènes *roX* semble également jouer un rôle particulier, puisque la propagation en *cis* du complexe MSL, observée autour d'un transgène *roX* sous le contrôle du promoteur *roX*, n'est pas reproduite si le transgène est sous le contrôle d'un autre promoteur (Meller et Rattner, 2002 ; Park *et al.*, 2002b ; Stuckenholtz *et al.*, 2003).

Finalement, l'adressage du complexe MSL semble reposer sur deux mécanismes distincts (Kelley, 2004) : l'assemblage du complexe MSL autour du transcrit roX naissant, qui concentrerait le complexe sur ces deux sites du chromosome X ; et la reconnaissance, par un complexe MSL soluble, de sites d'entrée sur le chromosome X, absents des autosomes. Ce deuxième mécanisme semble dépendre lui aussi de l'ARN roX, puisque les protéines du complexe MSL sont incapables de se localiser sur le chromosome X chez les mutants *roX1⁻ roX2⁻* (Franke et Baker, 1999).

1.5.2 ARN et hétérochromatine péricentromérique chez les Mammifères

Le traitement de cellules murines en culture, perméabilisées, avec une RNase fait disparaître les foyers d'hétérochromatine péricentromérique (identifiées par la concentration de protéine HP1 et d'histone H3 méthylée sur la lysine 9, en microscopie) (Maison *et al.*, 2002). L'addition d'ARN nucléaires des mêmes cellules (des fibroblastes de Souris immortalisés) aux cellules traitées reconstitue les foyers d'hétérochromatine. Les centromères de Mammifères sont transcrits à partir des deux brins (Rudert *et al.*, 1995 ; Lehnertz *et al.*, 2003) (voir aussi page 88), et des ARN transcrits à partir des centromères pourraient vraisemblablement être ce composant ARN, capable de concentrer les protéines de l'hétérochromatine sur les centromères. Cependant, puisqu'une incubation des cellules traitées à la RNase dans une préparation d'ARN suffit à restaurer les foyers d'hétérochromatine, la transcription des centromères en tant que telle n'est probablement pas responsable de l'hétérochromatinisation des centromères.

Des ARN transcrits à partir des centromères pourraient être impliqués dans la spécificité de la formation de ces foyers, en s'appariant à l'ADN centromérique, ou à des transcrits naissants. En particulier, des ARN transcrits à partir de séquences centromériques sont mis en évidence dans des cellules de Mammifères, après un choc thermique, et ils restent associés à leur site de transcription (Jolly *et al.*, 2004). Alternativement, un ARN transcrit en *trans*, à partir d'un gène différent, pourrait participer à l'assemblage d'un complexe de structuration de la chromatine. En particulier, ces foyers sont détectés avec un anticorps dirigé contre un épitope constitué de quatre extrémités N-terminales d'histone H3 méthylées sur la lysine 9, liées de façon covalente (cet anticorps reconnaît donc probablement l'histone H3 méthylée sur la lysine 9 dans une conformation de la chromatine où les histones H3, méthylées, sont à proximité les unes des autres), et ils ne sont pas détectés avec un anticorps dirigé contre une unique extrémité N-terminale d'histone H3 méthylée sur la lysine 9 (Maison *et al.*, 2002). Le composant ARN mis en évidence par ces expériences n'a donc probablement pas de rôle dans la réaction de méthylation de la lysine 9 de l'histone H3 (la cinétique de disparition des foyers — après seulement 10 min. de traitement à la RNase — semble trop rapide), mais plus certainement, dans le repliement, à plus grande échelle, de la chromatine dont l'histone H3 est méthylée sur la lysine 9.

1.5.3 ARN non-codants et contrôle transcriptionnel spécifique chez les Mammifères

Alors que l'ARN non-codant Xist contrôle la transcription d'un chromosome entier (voir page 91), et que le RNAi semble pouvoir réprimer la transcription de séquences quelconques (voir page 88), homologues aux siRNA, quelques exemples d'ARN non-codants pourraient réguler la transcription de gènes spécifiques chez les Mammifères. Ils sont présentés dans cette partie.

L'ARN double brin NRSE et l'activation transcriptionnelle des gènes neuronaux

Le petit ARN double brin NRSE, long d'une vingtaine de nucléotides, a été cloné à partir de neurones de Rat ; il est peu exprimé dans les progéniteurs neuronaux, surexprimé les quatre premiers jours de la différenciation neurale, puis son expression décroît pendant la suite de la différenciation (Kuwabara *et al.*, 2004). La surexpression simultanée des deux brins de ce petit duplex, dans des progéniteurs neuronaux en culture, provoque leur différenciation en neurones, alors que leur hydrolyse par un ribozyme inhibe la différenciation. Les deux brins s'accumulent dans le noyau, et le duplex interagit avec la protéine NRSF/REST, un facteur de transcription qui réprime l'expression des gènes neuronaux dans les progéniteurs.

Il n'est pas démontré que le petit ARN NRSE n'agit pas par répression traductionnelle (les effets observés pourraient alors être dus à un effet indirect de NRSE, par l'intermédiaire d'un gène qu'il inhiberait post-transcriptionnellement) ; mais la localisation nucléaire de l'ARN double brin NRSE, et son interaction avec NRSF/REST, suggèrent aux auteurs un rôle de l'ARN

dans un changement de spécificité du facteur de transcription, qui deviendrait activateur après interaction avec l'ARN.

Interférence transcriptionnelle

Chez un sujet humain, une délétion a rapproché le gène HBA2 (qui code l' α 2-globine) d'un gène transcrit dans l'autre orientation, LUC7L (qui code une protéine de la particule ribonucléoprotéique associée au petit ARN U1) ; la carte du locus sauvage, et du locus délété (appelé : α - Z^F), est présentée en figure 1.21.

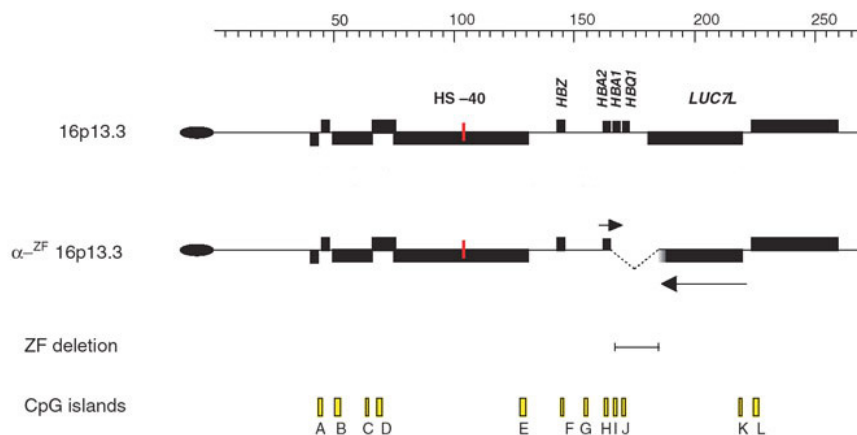


FIG. 1.21 – LA DÉLÉTION α - Z^F . Figure adaptée de Tufarelli *et al.*, 2003. Les gènes des α -globines (HBZ, HBA2, HBA1 et HBQ1), chez l'Homme, sont hébergés par le locus 16p13.3, à proximité d'un télomère (les répétitions télomériques sont représentées par une ellipse noire). Les gènes sont représentés par des rectangles noirs : ceux qui sont dessinés au-dessus de la ligne horizontale sont transcrits en direction du centromère, ceux qui sont dessinés en-dessous, en direction du télomère. La barre rouge symbolise l'élément régulateur de l' α -globine, HS-40. L'allèle α - Z^F est délété de 18 351 pb (la délétion est représentée par la barre horizontale). Les douze rectangles jaunes représentent les îlots CpG annotés du locus, tous hypométhylés chez les sujets sauvages. L'échelle (en haut de la figure) est exprimée en kb.

Chez cet individu, le promoteur de HBA2 est hyperméthylé, et le gène est réprimé ; la transcription de LUC7L s'étend, en orientation antisens, dans le gène HBA2, au moins jusqu'à l'îlot CpG du promoteur de HBA2 (Tufarelli *et al.*, 2003). Un transgène qui imite la disposition des gènes HBA2 et LUC7L dans cette délétion reproduit le phénomène *in vivo* : des souris transgéniques dans lesquelles LUC7L est en orientation antisens vis-à-vis de HBA2 répriment HBA2, et le promoteur de HBA2 est hyperméthylé. En revanche, un transgène homologue (qui porte la séquence du locus des gènes d' α -globine sauvage) n'est pas méthylé *in trans*. Des cellules souches embryonnaires de Souris, transformées avec le même type de transgène, reproduisent également la méthylation et la répression du promoteur de HBA2. Ce système a permis de démontrer que l'inactivation de HBA2 (par mutation du promoteur et d'un élément activateur) ne suffit pas à provoquer la méthylation du promoteur, et que la substitution de LUC7L par un autre gène, également transcrit en orientation antisens de HBA2, conduit elle aussi à la méthylation du promoteur (Tufarelli *et al.*, 2003).

On appelle « interférence transcriptionnelle », ou : « occlusion de promoteur », la répression d'un promoteur par l'activité d'un promoteur fort situé en amont : le passage des polymérases à

ARN provenant du promoteur amont, d'autant plus fréquent que ce promoteur est fort, empêche probablement la fixation de facteurs de transcription sur le promoteur aval, par gêne stérique ou par modification de la structure de l'ADN. Cet effet a été observé chez les Procaryotes (Hausler et Somerville, 1979 ; Ward et Murray, 1979 ; Horowitz et Platt, 1982 ; Adhya et Gottesman, 1982) et les Eucaryotes (Cullen *et al.*, 1984 ; Proudfoot, 1986 ; Corbin et Maniatis, 1989 ; Greger *et al.*, 2000). En particulier chez la Levure de boulanger *Saccharomyces cerevisiae*, il a été montré que la répression du promoteur aval par interférence transcriptionnelle était en compétition avec la fixation des facteurs de transcription : la surexpression d'un facteur de transcription capable de se fixer sur le promoteur aval réduit l'interférence transcriptionnelle (Greger *et al.*, 2000).

Le même phénomène a été rapporté récemment, chez la Levure de boulanger à nouveau, et l'ARN issu de la transcription par le promoteur amont a été détecté ; c'est un ARN non-codant, et sa longueur indique que sa transcription traverse le promoteur aval, et empiète sur la région transcrite du gène aval (Martens *et al.*, 2004).

Si cet effet est dû à une gêne stérique ou une modification de la conformation de l'ADN par le passage des polymérases à ARN, on peut prédire qu'il est indépendant de l'orientation relative des deux promoteurs : alors que la plupart des exemples cités ci-dessus mettent en scène deux promoteurs de même orientation, deux articles présentent, dans un système procaryotique, des effets similaires quand les deux promoteurs se font face (Ward et Murray, 1979 ; Horowitz et Platt, 1982). Telle est peut-être l'origine de la répression du gène HBA2 dans la délétion humaine α^{-ZF} ; dans ce cas, la méthylation du promoteur de HBA2 pourrait n'être qu'une conséquence de l'inactivité de HBA2, plutôt que sa cause.

Chapitre 2

Résultats

Tout ça c'est de la théorie, et la pratique n'est guère différente de la théorie — en théorie.
(David Madore, ens.forum.sciences.physique message n°2484)

2.1 Un deuxième locus de gènes d'ARN C/D répétés

Suite à la découverte des ARN C/D MBII-13, MBII-52 et MBII-85, à expression restreinte au cerveau de l'adulte (voir page 53), le laboratoire avait réalisé une banque d'ARN C/D à partir d'ARN de cerveau de Rat, et identifié un nouveau snoRNA spécifique du cerveau chez le Rat, RBII-36, et dont le gène, intronique, est répété (Cavaillé *et al.*, 2001). À la différence de MBII-13, MBII-52 et MBII-85, le snoRNA RBII-36 n'a pas d'homologue flagrant chez les autres Mammifères.

En revanche, son gène est situé dans le locus 6q31-32 du Rat, locus orthologue des domaines 14q32 humain et distal 12 murin, connus pour héberger des gènes soumis à l'empreinte génomique parentale (voir « Régulation de l'empreinte génomique parentale sur quelques loci », page 36) : ce nouveau gène de snoRNA C/D pouvait donc être soumis à l'empreinte. Dans ce cas, le gène de RBII-36 partagerait trois caractéristiques étonnantes avec les gènes de snoRNA du locus 15q11-13 (nomenclature humaine ; voir page 53) : une expression spécifique du cerveau, un gène répété, et soumis à l'empreinte génomique parentale. Cette perspective nous a encouragés à analyser en détail le locus 14q32 humain et distal 12 murin : il était possible que ce locus soit très similaire au locus 15q11-13, et que des gènes d'ARN C/D répétés et soumis à l'empreinte génomique parentale soient également hébergés par le locus 14q32 / distal 12. Les séquences nucléotidiques des gènes d'ARN C/D pouvaient alors être différentes entre le Rat, la Souris et l'Homme, mais leur organisation répétée pouvait être conservée.

Nous avons donc recherché les séquences répétées dans le locus humain : il contient trois types de répétitions (baptisées « cluster I », « cluster II » et « cluster R »). Les séquences répétées dans les clusters I et II portent les caractéristiques des séquences des ARN C/D (une taille de l'ordre de 75 pb, les boîtes C, D, C' et D', et la complémentarité des extrémités 5' et 3' ; voir page 46 pour la description des ARN C/D), et elles sont exprimées sous la forme de petits ARN de la taille prédite, et principalement dans le cerveau et la membrane utérine, chez l'Homme. Tout comme les gènes d'ARN C/D du locus 15q11-13, ils sont localisés dans des introns d'unités transcriptionnelles non-codantes ; en revanche, les ARN C/D de ce nouveau locus sont exprimés au cours du développement embryonnaire, à la différence des ARN C/D du locus 15q11-13, et leurs séquences sont plus divergentes (entre les différentes copies dans une même espèce, et entre espèces). Enfin, ils sont soumis à l'empreinte génomique parentale, au moins chez la Souris, et sont exprimés à partir de l'allèle maternel uniquement. L'empreinte de ces nouveaux gènes a été démontrée grâce à l'analyse des ARN d'embryons de Souris présentant des disomies uniparentales pour le chromosome 12. Ces échantillons d'ARN nous ont été fournis par le laboratoire d'Anne Ferguson-Smith, en Angleterre, et leur procédé d'obtention est présenté en figure 1.7, en Introduction (page 41).

L'organisation génomique du locus 14q32 (*Homo sapiens*) / distal 12 (*Mus musculus*) présente donc des similitudes troublantes avec celle du locus 15q11-13 / 7C, qui pourraient révéler un lien inattendu entre l'empreinte génomique parentale et les répétitions de gènes d'ARN C/D.

Cette étude fait l'objet de l'article présenté ci-après, « Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region » (Cavaillé *et al.*, 2002).

2.1.1 Article n°1

© 2002 Oxford University Press

Human Molecular Genetics, 2002, Vol. 11, No. 13 1527–1538

Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader–Willi/Angelman syndrome region

Jérôme Cavaillé^{1,*}, Hervé Seitz¹, Martina Paulsen², Anne C. Ferguson-Smith³ and Jean-Pierre Bachelier¹

¹LBME-CNRS (UMR5099), Université P. Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France,

²Universität des Saarlandes, FR Genetik, Postfach 151150, D-66041 Saarbrücken, Germany and ³Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

Received February 18, 2002; Revised and Accepted April 17, 2002

A human imprinted domain at 14q32 contains two co-expressed and reciprocally imprinted genes, *DLK1* and *GTL2*, which are expressed from the paternally and maternally inherited alleles, respectively. We have previously shown that another imprinted locus, on human 15q11–q13, contains a large number of tandemly repeated C/D small nucleolar RNA genes (or C/D snoRNAs) only expressed from the paternal allele. Here we show that the region downstream from the *GTL2* gene is also characterized by a transcription unit spanning many repeated intron-encoded C/D snoRNA genes, most of them arranged into two tandem arrays of 31 and 9 copies. Intriguingly, these snoRNAs depart from previously reported rRNA or snRNA methylation guides by their tissue-specific expression and by their lack of complementarity to rRNA or snRNA within their sequences. Analysis of the orthologous region in the mouse shows that the previously reported maternally expressed *Rian* gene, located downstream of *Gtl2* on the distal 12 chromosome, encodes at least nine C/D snoRNAs. Through a systematic search in rodents, we could identify other C/D snoRNA genes in this domain. All snoRNAs identified on mouse distal 12 are brain-specific and only expressed from the maternally inherited allele. The human imprinted 14q32 domain therefore shares common genomic features with the imprinted 15q11–q13 loci. This link between tandemly repeated C/D snoRNA genes and genomic imprinting suggests a role for these snoRNAs and/or their host non-coding RNA genes in the evolution and/or mechanism of the epigenetic imprinting process.

INTRODUCTION

A subset of mammalian genes undergo genomic imprinting, an epigenetic phenomenon that determines gene expression (or repression) according to parental origin (only the allele inherited from the father or that from the mother is expressed). Imprinted genes are often clustered, and multiple regulatory gene expression networks are predicted to operate within the same imprinted locus, involving allele-specific methylation, chromatin modification, antisense RNAs, chromatin boundaries and silencers (1–4). However, if common genomic features conferring local epigenetic control exist within imprinted domains, they have not yet been identified.

Among the most extensively studied imprinted domains are the insulin-like growth factor-2 (IGF2)/H19 domain and the

PWS/AS region involved in the Beckwith–Wiedemann and Prader–Willi/Angelman syndromes, respectively. The IGF2/H19 domain is located at the human 11p15.5 locus (mouse chromosome 7F) and possesses two co-expressed and reciprocally imprinted genes located 70 kb apart at one end of a larger cluster of imprinted genes. In IGF2/H19, the parental-specific gene expression is mediated by a germline-specific differentially methylated region (DMR) located between *IGF2* and *H19* that functions as an insulator element upon the binding of CTCF, a methylation-sensitive DNA-binding protein (5–8). The Prader–Willi/Angelman syndrome chromosomal region at human 15q11q13 (mouse chromosome 7C) is characterized by another gene cluster containing several paternally expressed genes (the bistrionic *SNURF–SNRPN* locus and the *MKRN3*, *NDN* and *MAGEL2* genes), while two genes, *UBE3A* and

*To whom correspondence should be addressed. Tel: + 33 5 61335934; Fax: + 33 5 61335886; Email: cavaille@ibcg.biotoul.fr
Correspondence may also be addressed to J.-P. Bachelier.

1528 Human Molecular Genetics, 2002, Vol. 11, No. 13

ATP10C, are expressed from the maternally inherited alleles (9). Imprinted expression within this large (approximately 2–3 Mb) domain is coordinated by a bipartite *cis*-acting element encompassing a DMR and located upstream from the paternally expressed *SNURF-SNRPN* gene (10–13).

We and others have recently shown that the Prader–Willi/Angelman syndrome chromosomal region and its counterpart in mouse also encode multiple, tandemly repeated C/D boxes containing small nucleolar RNA genes (or C/D snoRNAs) expressed only from the paternal chromosome (14–17). C/D snoRNAs guide the formation of 2'-O methylations of both

rRNA and U snRNAs through a specific RNA duplex at each modification site (18–20). However, the C/D snoRNAs at PWS lack a canonical antisense segment against rRNA and/or U snRNAs, suggesting that they can have other functions or modify other cellular RNAs such as mRNA (14). Remarkably, Runte and colleagues (21) have recently shown that all the human PWS-encoded snoRNAs are processed from a single paternally expressed transcription unit. It starts at the IC and overlaps the maternally expressed gene *UBE3A* in an antisense direction. Deletion of the PWS-IC on the paternal allele in mouse results in the loss of *Ube3a* antisense expression and

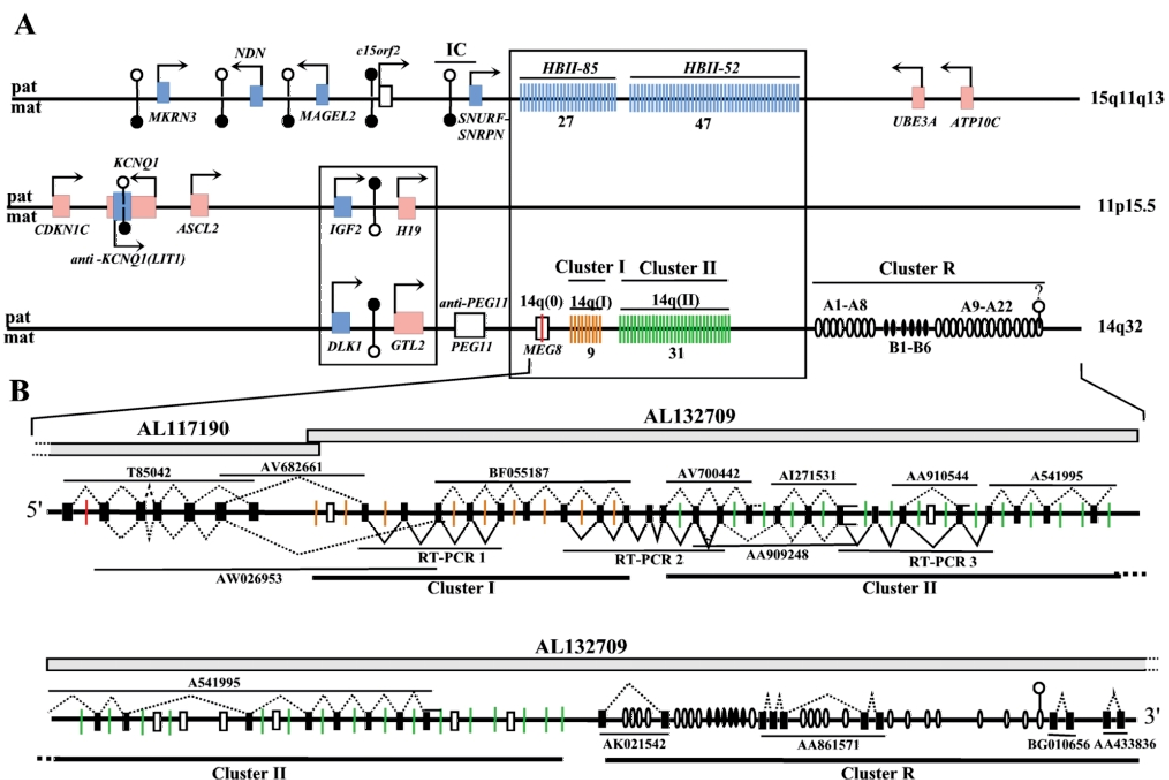


Figure 1. Identification of tandemly repeated C/D snoRNA genes at the human 14q32 imprinted locus. (A) Common imprinting features identified at the three human imprinted loci at 11p15.5, 14q32 and 15q11q13. The position and direction of transcription of the several imprinted genes within each human imprinted locus are shown by squares and arrows, respectively. Blue genes are paternally expressed while pink ones are expressed from the maternal allele only. The imprinting status of white squares remains to be checked in human. The C/D snoRNA genes at 14q32 and at 15q11q13 are depicted by vertical bars [note that HBII-13 (14) and HBII-436, HBII-437 and HBII-438A/B snoRNAs (21) are not indicated], with the number of snoRNA genes being indicated below. Several differentially methylated regions characterized within each domain loci are represented by circles (filled, methylated regions; open, unmethylated regions). IC, imprinting center; mat and pat, maternally and paternally inherited chromosomes. Repeated sequences of A-type (22 copies) or B-type (6 copies) within the repeat cluster (cluster R) are represented by open and filled ovals, respectively. Common structural gene organization features between these imprinted loci are boxed. (B) C/D snoRNAs at human 14q32 are intron-encoded and processed from a single transcript. Schematic representation of the genomic region spanning the three clusters of repeated sequences (derived from AL117190 and AL132709 clones). C/D snoRNA genes are indicated by vertical bars [red, 14q(0) snoRNA; brown, 14q(I) snoRNAs; green, 14q(II) snoRNAs], while exons detected by genomic/ESTs comparison or predicted by *in silico* analysis are represented by black or white boxes, respectively. Dotted and filled lines denote splicing events identified by ESTs analysis or RT-PCR experiments, respectively. Nucleotide positions of exons within AL117190: (?–134517); (138432–138562); (145574–145640); (146416–146610); (149709–149772); (152908–152998); (154149–?); within AL132709 (reverse complementary strand): (5510–5582); (7041–7109); (12940–12993); (14714–14827); (16345–16430); (17822–17898); (19307–19392); (21704–21780); (23751–23814); (25976–26062); (27920–27979); (28466–28538); (29344–29414); (30374–30436); (31363–?); (32906–32986); (33864–33948); (34827–34891); (37204–37285); (43295–43381); (44257–44334); (45587–45677); (46581–46641); (47590–47666); (48401–48487); (51206–51288); (53068–53145); (54353–54438); (55008–55094); (55897–55983); (56547–56636); (59519–59605); (60516–60602); (61446–61527); (62403–62484); (63290–63376); (64597–64682); (65561–65648); (66486–66572); (68636–68714); (?–89220); (106288–106357); (?–123635); (123746–123858); (123948–124126); (130009–130211); (130544–?); (?–148997); (149384–?); (?–150933); (151301–?). Nucleotide positions of predicted exons are indicated in italics. Other symbols are as in (A). The cartoons are not drawn to scale.

activation of the paternal *Ube3a* allele (22). Although an RNA-independent chromatin mechanism might be involved in this regulation, this observation raises the possibility of a role for this snoRNA host gene in the silencing of paternal *UBE3A* gene expression (21–23).

Indeed, there are several examples of imprinted, protein-coding genes being physically linked to non-coding RNAs and exhibiting a reciprocally imprinted expression pattern. Transcribed from the other parental chromosome, these mono-allelically expressed, non-coding RNAs are often transcribed as antisense RNAs relative to the sense, protein-coding genes: *Igf2r/Air* (24), *KVLQT/LIT1* (25), *UBE3A/UBE3A-AS* (21,23) and *Nesp/Nesp-AS* (26). Reciprocally imprinted non-coding RNAs can also be transcribed in the same orientation as the protein-encoding gene, as described for both the *Igf2/H19* and *Dlk1/Gtl2* imprinted domains (3). In most cases, the function of these RNAs remains to be established, but the available evidence suggests that this particular gene organization relative to regional controlling elements is crucial in the regulation of the two pairs of reciprocally imprinted genes (3,27–29).

Like the two better-characterized clusters of imprinted genes at chromosomes 11p15.5 and 15q11q13, recent data indicate that the *DLK1–GTL2* region on chromosome 14q32/mouse distal 12 may also be part of a larger cluster of imprinted genes. Although still under investigation, the gene organization at this locus has been studied in the sheep and human, where two genes downstream of *GTL2*, namely *PEG11* and *MEG8*, have been described and shown in the sheep to exhibit paternal-specific and maternal-specific allelic expression, respectively (30,31). In the mouse, *Rian*, a brain-specific non-coding RNA expressed from the maternal allele, has also been mapped to this region, although its precise location relative to *Gtl2* has not been determined (32). Comparative genomic analysis of imprinted domains between or within species can identify conserved genomic features that may be functionally important in imprinting control (31,33–36). Using this approach, we describe here the identification of tandemly repeated C/D snoRNA genes located at the human 14q32 locus, adjacent to the imprinted *DLK1/GTL2* domain. Most of these are intron-encoded and processed from a complex transcription unit mapping to the *MEG8* gene. Furthermore, we have identified other putative C/D snoRNA genes at orthologous loci in rodents. All of these C/D snoRNAs are brain-specific and expressed from the maternal allele only. Thus, the human imprinted 14q32 domain shares common genomic features with the Prader–Willi/Angelman syndrome chromosomal region (Fig. 1A).

RESULTS

Searching for tandemly repeated C/D snoRNA genes at human 14q32

We have recently identified a rat C/D snoRNA, RBII-36, encoded in the *Bsr* gene (37). The gene organization and tissue-specific expression pattern of RBII-36 are highly reminiscent of those of imprinted C/D snoRNAs at human 15q11–13/mouse 7C loci (14,38). The RBII-36 snoRNA gene maps to 6q31q32, a locus that is orthologous to mouse distal chromosome 12 and

human 14q32, which harbour imprinted genes. Among the most well-characterized imprinted genes at this locus are *Dlk1*, a paternally expressed gene encoding a cell-surface membrane protein of the Notch–Delta family, and *Gtl2/Meg3*, a maternally expressed gene located 80 kb downstream and lacking conserved open reading frames (39,40). To identify putative human homologues of RBII-36 snoRNA genes, we searched the human 14q32 imprinted locus for the presence of direct tandemly repeated C/D snoRNAs. Three direct-repeat clusters (two snoRNA clusters I and II and one cluster of direct repeats, cluster R) have been detected in BAC sequences AL132709 (Fig. 1B). Cluster I (nucleotide position ~3700–24700) and cluster II (nucleotide position ~28750–72260), are located approximately 40 kb downstream from *PEG11* and contain 9 and 31 copies, respectively, of 70–75 nt-long related snoRNA-like sequences (Fig. 2A). Copies of these snoRNA-like sequences contain the characteristic C/D/C/D motifs (in that order) and also a terminal 5'–3' stem structure exhibiting the conserved sequence 5'-GGACC...GGTCC-3'. Despite a relatively high overall similarity of the snoRNA gene copies within each cluster, their sequence tracts upstream from either box D or box D' exhibit a substantial number of nucleotide differences among copies, as exemplified for snoRNA copies within cluster I, and do not represent potential antisense elements against rRNA or U snRNAs (Fig. 2A). Moreover, in agreement with our failure to detect specific RBII-36 homologues in human (37), none of these sequences shows significant similarities with rat RBII-36 snoRNA. These sequences therefore represent novel human C/D snoRNA genes, and we propose to name them 14q(I) and 14q(II) for snoRNAs encoded in clusters I and II, respectively. By focusing on sequences conserved between human and ovine *MEG8*, we were able to identify a third intron-encoded C/D snoRNA gene present as a single copy and termed 14q(0) since it does not belong to a cluster (Fig. 1). Again, this C/D snoRNA, conserved between human and sheep (Fig. 2A), does not contain any obvious antisense element against rRNAs or U snRNAs. Further analysis downstream of the snoRNAs (nucleotide position ~102260–144600) resulted in identification of two classes of ~80–100 nt-long direct repeats (22 and 6 copies of A- and B-type, respectively) that do not exhibit snoRNA hallmarks (Figs 1 and 2B). All are in head-to-tail orientation, with a relatively irregular spacing between A repeats, while five of six B repeats are clustered and separated from each other by a 350–380 nt-long spacer sequence. Interestingly, one of these repeats, A22, is embedded within a conserved CpG island at the end of cluster R (Fig. 1B).

The C/D snoRNAs are intron-encoded and processed from the tissue-specific non-coding human *MEG8* RNA

To learn more about the snoRNA gene organization and mode of biosynthesis, GenBank was searched for expressed sequence tags (ESTs) covering the snoRNA gene clusters. We identified several human spliced ESTs and putative exons overlapping this region, all of which appear to be devoid of significant protein coding potential (Fig. 1B). Comparisons of genomic and cDNA sequences and RT–PCR analysis revealed that most of the tandemly repeated C/D snoRNA genes and also several A and B repeats are positioned within introns (Fig. 1B).

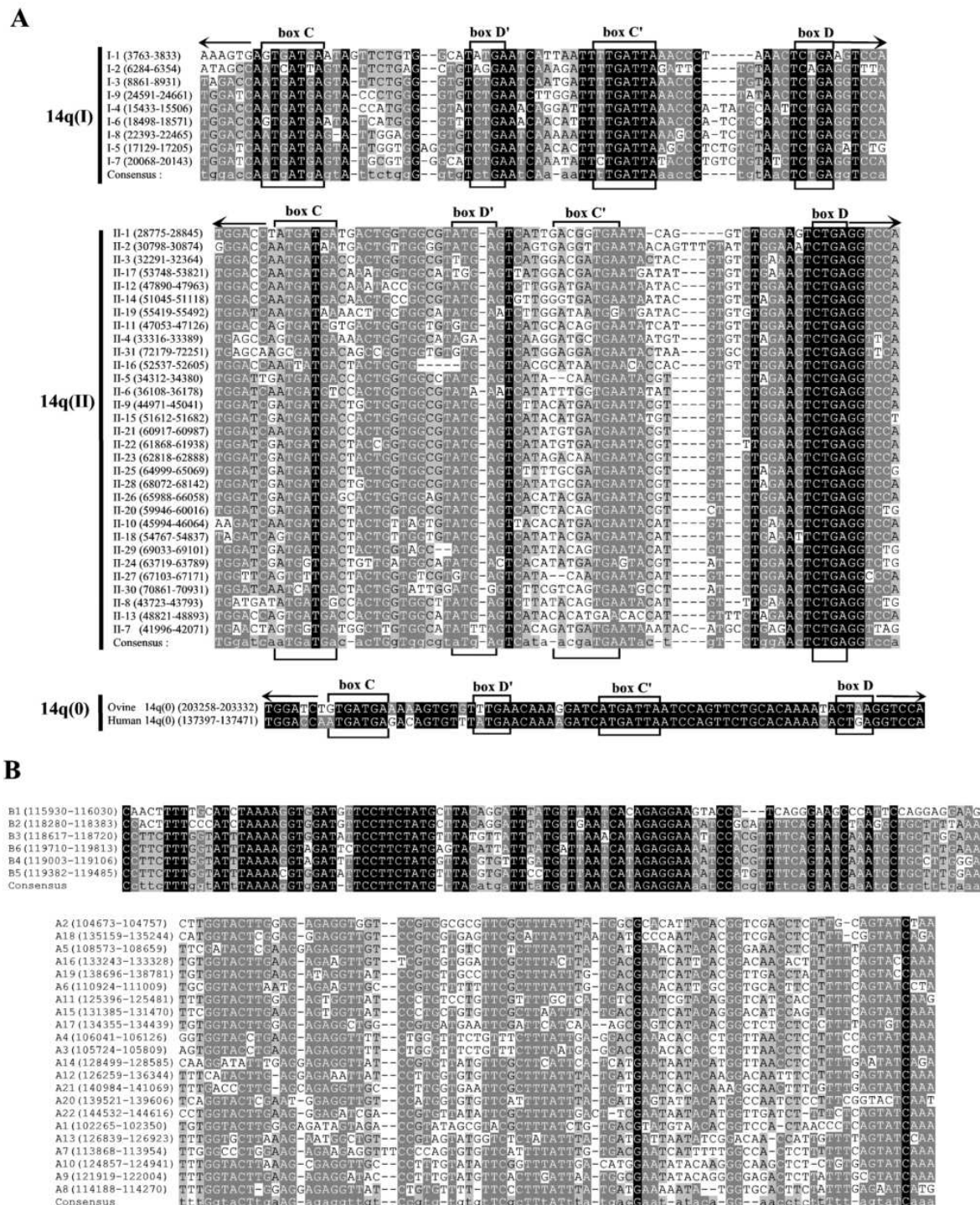
1530 *Human Molecular Genetics, 2002, Vol. 11, No. 13*

Figure 2. Sequence alignment of the repeated human C/D snoRNA genes and A and B repeats. (A) Sequence alignment of human 14q(I) and 14q(II) snoRNA gene copies within clusters I and II, respectively, and of 14q(0) human/ovine snoRNA genes. Nucleotide positions of snoRNA genes within BAC AL117190 [14q(0)]; AL132709 [14q(I)] and AF354168 [ovine 14q(0)] are given in brackets. Box C/C' and box D/D' motifs are boxed and the terminal 5'-3' stem is indicated by arrows in the opposite orientation. (B) Sequence alignment of A and B repeats. Nucleotide positions of these repeats within AL132709 are shown in brackets. Conservation of each nucleotide position is denoted by shading according to GeneDoc software: black (100%), dark grey (80-99%) and clear grey (60-79%).

Interestingly, the AW026953 EST clearly connects the 3' part of *MEG8* RNA with the 5' end of the snoRNA gene cluster I. Clusters I and II also appear to be transcribed from the same unit, as suggested by sequencing of a RT-PCR product 2 (Fig. 1B) overlapping the two types of snoRNA genes. Taken together, this analysis suggests that the three novel types of human snoRNAs are processed from a single transcription unit that includes the *MEG8* non-coding RNA gene. Northern blot analysis performed on total RNA extracted from a panel of human tissues showed that the three C/D snoRNAs, 14q(0), 14q(I) and 14q(II), exhibit the same tissue-specific expression pattern. The strongest expression was observed for the brain and the uterine mucous membrane and to a lesser extent within the heart, while only a very faint signal could be detected for all other human tissues analysed (Fig. 3). This expression pattern was compared with that of the snoRNAs encoded at 15q11-q13. While HBII-52 snoRNA was only detected in the brain, HBII-85 was detected at relatively high levels in the brain, uterine mucous membrane and kidney and at low levels in several other tissues (Fig. 3) (14).

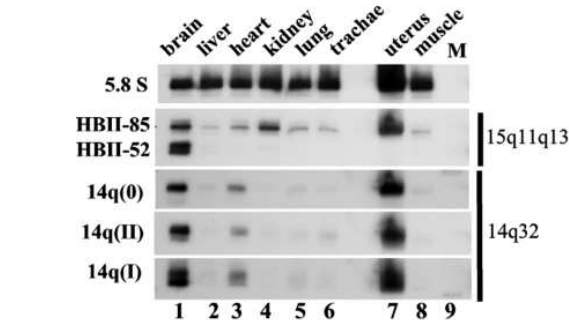


Figure 3. Tissue-specific expression of the novel snoRNAs analysed by northern hybridization. Total RNAs (5 µg/lane) purchased from Clontech (lanes 1–6) or extracted from normal adult human tissues (lanes 7 and 8) were fractionated on a 6% acrylamide gel/7 M urea and analysed by northern blot with ³²P labelled oligonucleotide probes against 14q32- or 15q11q13-encoded snoRNAs (see Materials and Methods). A 5.8S rRNA probe has also been used for gel loading control. M, DNA marker.

The maternally expressed *Rian* transcript at mouse distal 12 chromosome encodes at least 9 C/D snoRNAs

Hatada *et al.* (32) recently described a novel 5.4-kb maternally expressed brain-specific non-coding RNA gene, named *Rian*, that is closely linked to the *Gtl2* gene on distal mouse chromosome 12. BLAST N database searches identified several mouse ESTs positive for *Rian* (data not shown), among them

AK017440 cDNA harbouring four segments exhibiting 100% identity to the *Rian* transcript. Sequence analysis revealed that these segments of homology correspond to exons (exons a–d), since they are interrupted by sequences displaying splice donor and acceptor consensus intronic hallmarks at both ends (data not shown; Fig. 4A). Moreover, AK017440 also displays homology with a 43 kb-long mouse contig 474407. By further sequence analysis, we could identify within AK017440

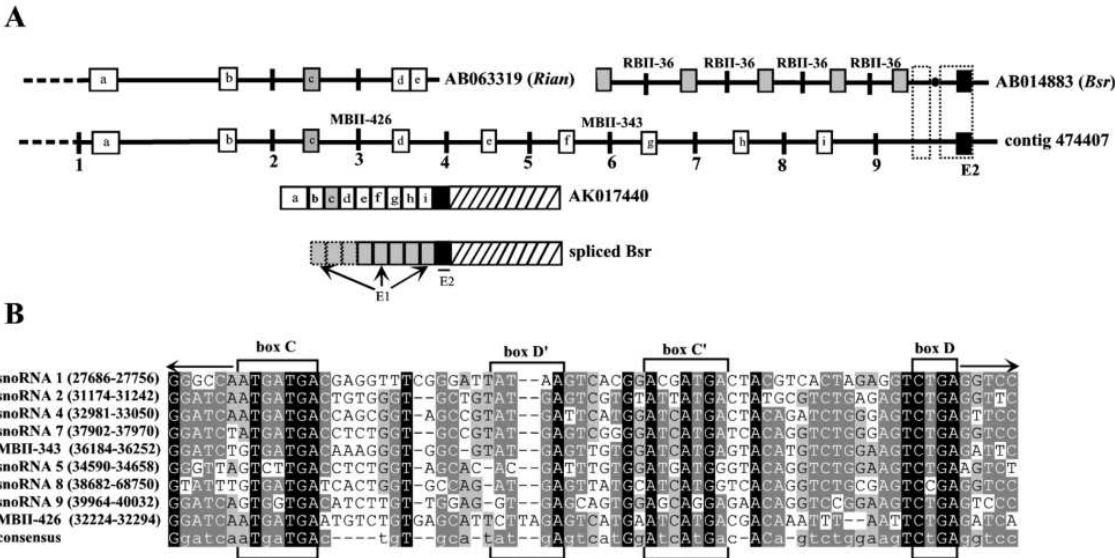


Figure 4. The maternally expressed mouse *Rian* gene encodes nine related C/D snoRNAs. (A) Schematic representation of the *Rian* gene (contig 474407), hemi-processed *Rian* RNA (AB063319) as described in (32), fully spliced *Rian* RNA (AK017440), the *Bsr* gene (AB014883) and spliced *Bsr* RNAs as described in (37). Exons are depicted by boxes, introns by a thick horizontal line and snoRNAs by a vertical bar. *Rian* exons closely similar to repeated E1 and E2 *Bsr* exons (37) are filled in grey and black, respectively. The repetitive element Lx7LINE/L1 at the 3' end of both spliced *Bsr* and *Rian* RNAs is denoted by a hatched box. Intronic sequences conserved between mouse *Rian* and rat *Bsr* genes are denoted by dotted boxes (the black oval in between depicts a repetitive ID element in the *Bsr* intron not found within the *Rian* gene). (B) Sequence alignment of C/D snoRNA genes encoded in *Rian* introns. Nucleotide positions of C/D snoRNAs within contig 474407 are shown in brackets. Symbols are as in Fig. 2. Average percentages of identity to the consensus sequence of the *Rian*-encoded snoRNAs and RBII-36 snoRNA are 70.6% and 88.7%, respectively (according to GeneDoc software).

1532 Human Molecular Genetics, 2002, Vol. 11, No. 13

additional spliced exons (e-i), suggesting strongly that the *Rian* gene, as originally described, might represent an RNA-processing intermediate. Remarkably, AK017440 cDNA also shares some sequence and structural features in common with the rat *Bsr* cDNA. First, *Rian* exon c displays 80% identity to the repeated E1 exons from *Bsr* RNA (37) (Fig. 4A). Second, *Rian* and *Bsr* cDNAs both exhibit a conserved exon E2 sequence (82% identity) immediately followed by an Lx7/LINE1 repetitive element. Third, sequences of the last

Bsr and *Rian* introns are very similar to each other (with two ~264 and 155 nt-long segments exhibiting 72% and 76% identity; Fig. 4A). These common features prompted us to look for the presence of RBII-36-related snoRNAs within the *Rian* gene. We could identify nine C/D snoRNAs genes, all of which are intron-encoded (Fig. 4A). Surprisingly, they do not display significant homology with RBII-36 and are relatively divergent from each other (Fig. 4B). C/D snoRNAs located within introns 3 and 6 correspond to mouse MBII-426 and MBII-343

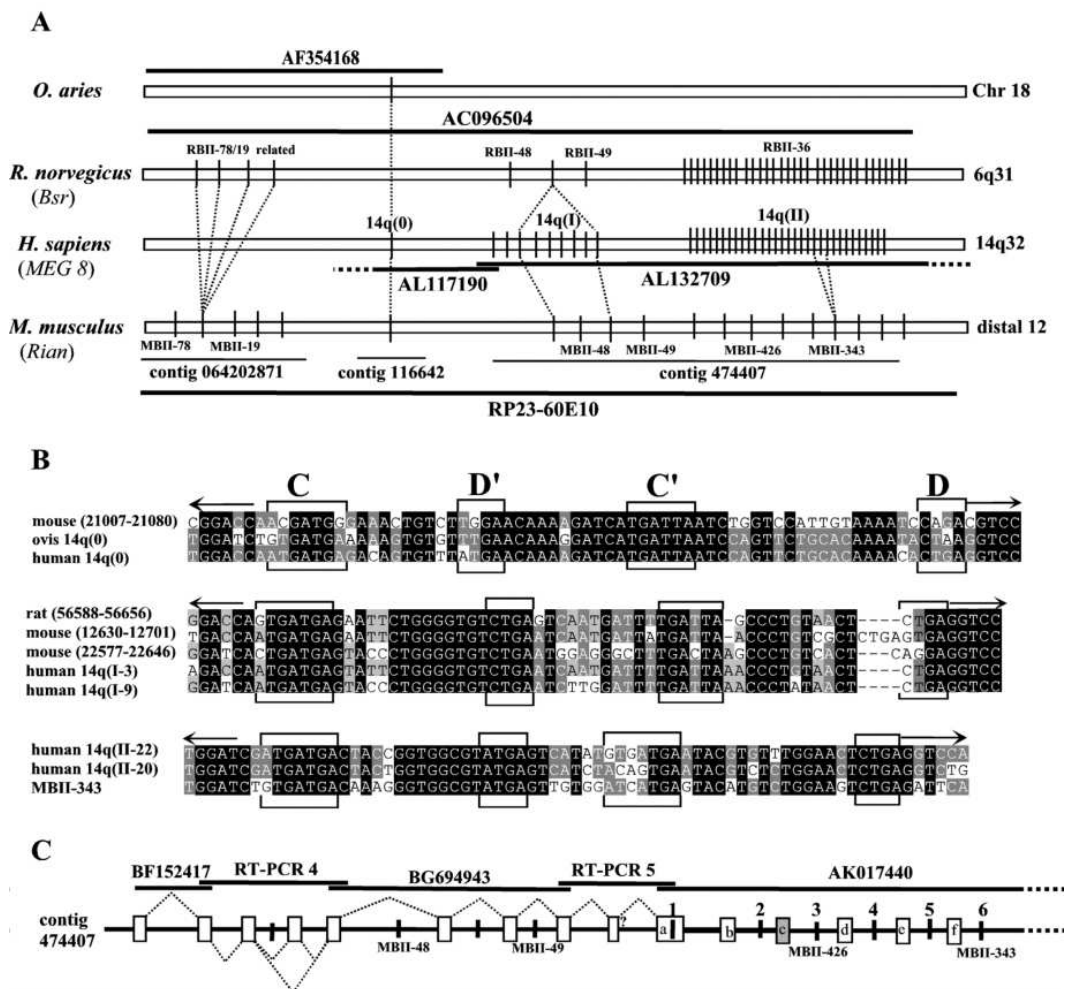


Figure 5. Identification of C/D snoRNA genes at mammalian chromosomal loci orthologous to human 14q32. (A) Genomic location of C/D snoRNAs at several mammalian loci. C/D snoRNA genes are represented by vertical black bars. The relative locations of snoRNAs within BAC AC096504 clone (48 unordered pieces) and the positions of mouse contigs 064202871, 116642 and 474407 have been chosen arbitrarily. Sequence homology between rodents and human C/D snoRNAs are denoted by dotted lines (only several homologies are shown). Within AC096504, RBII-78/19-related snoRNAs are located at 45772–45843, 46637–46708, 77115–77186 and 77980–78049, RBII-49 at 58503–58566 and RBII-48 at 37220–37289. Within contig 064202871, MBII-78/19-related snoRNAs are located at 11698–11764, 12553–12625, 13272–13344, 14137–14209 and 14747–14819. Within contig 474407, MBII-48 snoRNA is located at 17568–17637 and MBII-49 at 24393–24459. The cartoons are not drawn to scale. (B) Sequence alignment of putative mammalian C/D snoRNA homologues of 14q(0) snoRNA (top), 14q(I) snoRNA (middle) and 14q(II) snoRNA (bottom). Symbols are as in Fig. 2A. (C) MBII-48, MBII-49 and other flanking C/D snoRNAs are also intron-encoded within the *Rian* gene. Nucleotide positions of exons within contig 474407: (1867–1954); (6656–6730); (8603–8708); (15831–15923); (16949–17012); (19394–19481); (24120–24207); (24969–25016); (26817–26872 ?); (27404–28312); (30711–30799); (31629–31714); (32570–32656); (33416–33495); (35843–35980); (37406–37491); (38345–38388); (39610–39678); (41280–41364). Underlined exon might represent an alternatively spliced, snoRNA1-containing product. Other symbols are as in Figs 1B and 4A.

snoRNAs, respectively, previously characterized by analysis of a cDNA library (41). As for the other imprinted C/D snoRNAs, we could not detect potential cellular RNA targets based on appropriate complementarity. Interestingly, several copies of the MBII-343/426-related snoRNA cluster exhibit significant homology with several snoRNA copies within human cluster II, suggesting that the *Rian*-encoded snoRNAs might be the rodent counterparts of human 14q (II) snoRNAs (Fig. 5A, B).

Identification of additional C/D snoRNAs at rat 6q31 and mouse distal chromosome 12

To gain more information about the imprinting status of snoRNA genes encoded at human 14q32, orthologous mammalian loci were searched systematically for the presence of putative C/D snoRNAs. During the course of this work, a portion of the rat *Bsr* genomic locus was deposited in GenBank (accession no. AC096504); 33 copies of RBII-36 (Fig. 5A) as well as 7 other C/D snoRNA genes were detected in this portion. Four of these are related to each other and similar to the previously identified mouse MBII-19 and MBII-78 snoRNAs (41), which seem likely to represent variants of the same snoRNA species (data not shown). The three other rat C/D snoRNA genes display strong homology to the mouse C/D snoRNA MBII-48 and MBII-49 (Fig. 5) (41) and also interestingly to the several human 14q(I) snoRNA gene copies (Fig. 5B). Through BLAST searches, we then focused our attention on the genomic organization of the mouse homologues. Interestingly, MBII-48 and MBII-49 snoRNAs were both detected within contig 474407 containing the *Rian* gene (Fig. 5A), in addition to several other C/D snoRNAs that are more particularly related to some of the human 14q(I) snoRNA copies (Fig. 5B). By using a mouse BLAT search at <http://genome.ucsc.edu/>, a putative mouse homologue of the human 14q(0) snoRNA (Fig. 5B) was also detected within contig

116642 embedded within a region of strong homology to human *MEG8*. Finally, five related C/D snoRNAs, including MBII-19 and MBII-78 snoRNAs were also found within mouse contig 064202871. PCR analysis of a partially characterized BAC containing *Gtl2* (RP23-60E10) confirmed that these snoRNA genes are actually encoded by the mouse distal 12 imprinted region (data not shown). RT-PCR experiments and analysis of mouse ESTs overlapping the mouse snoRNA genes showed that MBII-48, MBII-49 and other flanking C/D snoRNAs are also intron-encoded and processed from a transcript physically linked to *Rian* RNA (Fig. 5C). Several of these rodent snoRNA genes contain one or two base substitutions within the C/D motifs (Fig. 5B), and could possibly be unstable or non-functional. Consistent with this, expression of several of them was undetectable on a northern blot (data not shown).

The brain-specific C/D snoRNAs are imprinted in mouse

Northern blot analysis of total RNA extracted from mouse adult tissues shows that all the mouse C/D snoRNAs located at mouse distal chromosome 12 are only expressed within the brain. These include MBII-19, MBII-343 and MBII-426, which were previously reported as ubiquitously expressed, possibly as a result of probe cross-hybridization (Fig. 6A) (41). Expression levels of C/D snoRNAs at mouse distal chromosome 12 have also been studied during brain development. Northern blot analysis performed at various developmental stages (Fig. 6B) showed that all C/D snoRNAs located at distal chromosome 12 are expressed within the embryo (E11–E15), the newborn (postnatal days P0–P7) and the adult. During brain ontogeny, the level of snoRNA expression is roughly constant or slightly increased. This developmental expression pattern is very different from that observed for MBII-52 and MBII-85 snoRNAs (encoded at chromosome 7C), which are very poorly

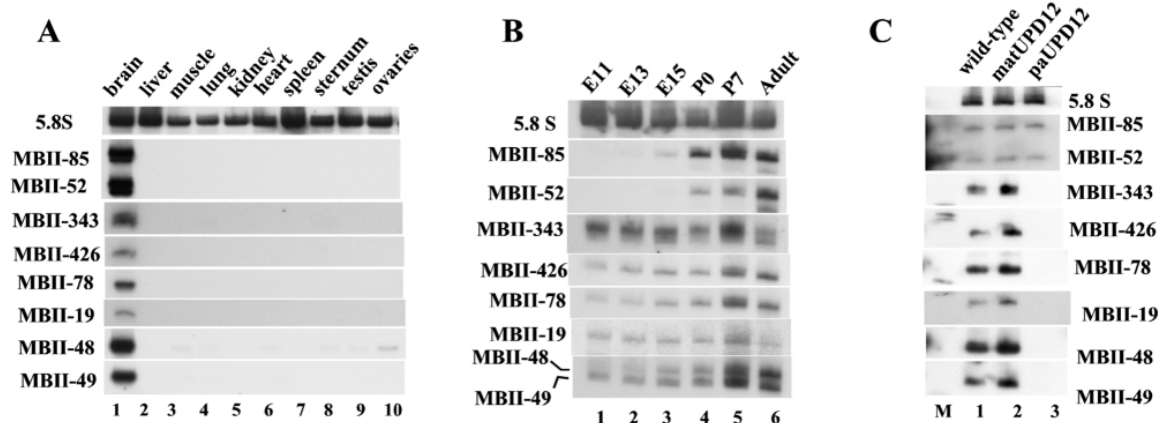


Figure 6. C/D snoRNA genes encoded at mouse distal chromosome 12 are brain-specific and expressed only from the maternal chromosomes. Each lane has been loaded with 10 µg of total RNAs extracted from different mouse adult tissues as specified at the top of each lane in (A), or from developing mouse embryos (E11–E15), newborn (P0–P7) or adult brains in (B), or from embryos (E15.5) with either maternal or paternal uniparental disomy for chromosome 12 (matUPD12 and patUPD12, respectively (55)) in (C). Samples were fractionated on a 6% acrylamide/7 M urea gel and analyzed by northern blot with ³²P-labelled oligonucleotide probes (see Materials and Methods). A 5.8S rRNA probe has been used as gel loading control. Signal intensities (ratio snoRNA/5.8S rRNA signal) were quantified with a Fuji-Bas-1000 imager. M, DNA marker.

1534 Human Molecular Genetics, 2002, Vol. 11, No. 13

expressed between E11 and E15 but are highly upregulated after birth.

We then examined whether the C/D snoRNAs were subjected to genomic imprinting as predicted for genes encoded within this mouse distal chromosome 12 region. As shown in Figure 6C, E15.5 embryos with two maternal chromosomes 12 (matUPD12) express all the C/D snoRNAs analysed. Moreover, for most of them, the expression level is approximately twice that seen for normal embryos as determined by quantification of the hybridization signal (data not shown). Conversely, none of those snoRNAs was detected in total RNA extracted from embryos having two paternal chromosomes 12 (patUPD12), showing that all of these snoRNA genes are imprinted and expressed only from the maternal allele. As a control, we also checked the expression of the two C/D snoRNAs located at 7C. As expected, they were unaffected in UPD12 embryos.

DISCUSSION

Comparative genomic organization and expression of C/D snoRNAs at human 15q11–q13 (mouse 7C) and 14q32 (mouse distal 12)

Imprinted *DLK1/GTL2* and *IGF2/H19* domains at human 14q32 and 11p15.5, respectively, appear to share several genomic and epigenetic features that may be involved in the regulation of these two pairs of reciprocally imprinted genes (3,35,36,39,40,42). In this work, multiple tandemly repeated C/D snoRNA genes at human imprinted 14q32 and at its orthologous region on mouse distal 12 chromosome were

identified. These C/D snoRNAs have a gene organization strikingly reminiscent of that previously reported for the PWS/AS region (14–16) (Fig. 1 and Table 1). Indeed, both imprinted loci are characterized by a complex transcription unit spanning arrays of tandemly encoded C/D snoRNA genes, giving rise to a spectrum of alternatively spliced non-coding RNAs, mostly consisting of short repeated exons that are not conserved between human and rodents. All 14q32-encoded C/D snoRNAs appear to be derived from a unique transcription unit encompassing *MEG8*, similar to the PWS-encoded C/D snoRNA (21) (Fig. 1B). However, despite their related gene organizations, the two loci differ in both snoRNA expression pattern and genomic features around the snoRNA host genes. While the three different snoRNA species encoded in human 14q32 have the same tissue-specific expression pattern, the expression patterns of HBII-52 and HBII-85 snoRNAs at the PWS/AS region differ considerably from each other (Fig. 3). Since the PWS-encoded snoRNAs are produced from the same primary transcript (21), this might reflect brain-specific differential processing of the more distal part of the *SNRPN-SNRUF*-snoRNA host transcript. Unlike human snoRNAs, all mouse distal 12-encoded C/D snoRNAs are mainly detected within the brain, and their level of expression is variable according to the snoRNA species – but always lower than that of 7C-encoded snoRNAs, which are among the most abundant C/D snoRNA within the brain (37,41). Mouse 7C-encoded C/D snoRNAs are predominantly expressed in the adult from the paternal allele (Fig. 6B) (14), while the mouse distal 12 snoRNAs are transcribed from the maternal allele and are expressed during brain development (Fig. 6B, C). The snoRNA host genes at 14q32 and at the PWS/AS domain are dramatically different in their content of common interspersed

Table 1. Imprinted snoRNA genes at human 15q11q13 and 14q32 loci

Locus and snoRNA features	15q11q13		14q32				
snoRNA gene organization:							
C/D snoRNA gene cluster	HBII-85 (27 copies) and HBII-52 (47 copies)		14q(I) (9 copies) and 14q(II) (31 copies)				
Cluster size (kb)	48	99	21.3	44.3			
% identity snoRNA copies/consensus ^a	90.2	93.9	82.3	80.8			
% GC content	50.1	58.36	(41)	34.1			
Repetitive elements (core of snoRNA clusters)	{	SINEs	0.62	0	(13.14)	4.72	6.23
		LINEs	1.43	0	(20.42)	13.76	1.62
		LTR elements	0	1.2	(8.29)	2.53	0.68
		DNA elements	0	0	(2.84)	2.76	1.21
snoRNA mode of biosynthesis:							
Located within repeated introns	Yes		Yes				
A single transcription unit	Yes		Yes				
% identity repeated spliced exons/consensus ^a	59.3	79.9 and 67.5 ^b	47.1	31.2			
Imprinted snoRNA genes in human/mouse	Yes/Yes (paternal)		?/yes (maternal)				
snoRNA expression							
Brain-specific in human	HBII-85 (no) and HBII-52 (yes)		Mainly expressed in brain/heart/mucous uterin				
Brain-specific in rodents	Yes		Yes				
Expression upon embryonic mouse development	No or very low		Yes				
snoRNA features							
Common terminal stem 5'-GG(A/G)TC...GG(C/T)CC-3'	Yes		Yes				
Antisense element conservation	Strong		Moderated				
Identified putative RNA targets	No (5-HT2c mRNA and HBII-52?)		No				

^aAs defined by GeneDoc software.

^bThe repeat unit spanning the HBII-52 snoRNA gene contains two different repeated exons. The complete human genome values (57) are given in parentheses.

repeats and overall base composition (Table 1). In both cases, a sharp transition in G + C content is observed at the borders of the snoRNA clusters, corresponding to a striking increase (14q32) or decrease (15q11q13) in the flanking sequences (Fig. 7). The reciprocal parental expression of the snoRNA gene clusters with regard to their G + C content might reflect the evolutionary maintenance or function of the snoRNA regions in ways that are not yet understood.

C/D snoRNA function and evolutionary divergence

Methylation guides are evolutionary ancient molecules (detected both in Archaea and eukaryotic cells) required to modify a large panel of cellular RNAs, including ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs) (20), tRNAs (43) and possibly mRNAs (14). The C/D snoRNAs identified at the two imprinted loci depart from other known mammalian methylation guides by their tissue-specific expression (Figs 3 and 6A). Moreover, in contrast to most of the other C/D snoRNAs, no statistically significant complementarity to rRNA or U snRNA within snoRNAs of the imprinted loci were identified. HBII-52 was the only exception to this (14). Characterization of the cellular RNAs potentially targeted by these imprinted C/D snoRNAs represents a worthwhile challenge, since lack of HBII-85 snoRNA expression has been proposed to contribute to PWS phenotypes (44). Although the human (*MEG8*), mouse (*Rian*) and rat (*Bsr*) snoRNA host genes share a common physical localization, their DNA sequences are not highly conserved, similar to the human and mouse host genes at the PWS/AS region (14–16). However, in contrast to the PWS-encoded snoRNA gene copies which are relatively well conserved among each other, the *Rian*-encoded mouse snoRNA gene copies, and also to some extent the human 14q(I) and 14q(II) snoRNAs, are much more divergent from each other (Figs 2A and 4B and Table 1). The lack of strong selection pressure on the snoRNA sequence segments upstream from D/D' between mammals or even between snoRNA copies within the same gene cluster might suggest that these imprinted C/D snoRNAs are rapidly evolving molecules with multiple potential RNA targets. Alternatively, this may

strengthen the hypothesis that they confer a genomic regulatory function rather than the alternative previously described snoRNA functions (see below).

Tandemly repeated C/D snoRNAs and imprinting

Interestingly, the 5'–3' terminal stem of most of the tandemly repeated snoRNAs at 15q11q13 and 14q32 loci exhibit an identical sequence (Fig. 2A and Table 1) not shared by any other snoRNA among the scores of ubiquitously expressed, non-repeated C/D snoRNAs reported for mammals (18,41,45). This could indicate that the novel tandemly repeated snoRNAs at the two imprinted loci have all evolved from a common snoRNA ancestor gene during mammalian evolution. Several observations have suggested that C/D snoRNA genes may represent mobile genetic elements (46,47). The tandemly repeated snoRNA arrays may well result from retrotransposition of a snoRNA gene into an intron followed by tandem duplications of the intron and flanking exon. Generation of the repeated snoRNA arrays could have occurred after establishment of the imprinted status of these loci, merely reflecting the fact that they represent 'hot spots' for DNA integration events in the germline, in relation to a particular chromatin structure. Alternatively, their occurrence could have predated imprinting and even played a role in its establishment and maintenance.

Many imprinted loci harbour non-coding RNAs having an imprinted expression opposite to that of linked protein-coding genes on the same chromosomes, raising the possibility of a role for these RNAs in silencing (3,27–29). C/D snoRNA host-gene transcripts could therefore be involved in gene regulation, perhaps in a manner similar to that of other chromosomal RNAs involved in dosage compensation or, like Air RNA, in the regulation of imprinted genes at mouse chromosome 17 (48,49). However, the snoRNA genes at 15q11q13 do not seem to play a central role in the epigenetic process, since a paternal deletion from the *Snrpn* to the *Ube3A* gene in the mouse does not affect the imprinted status of the upstream *Ndn* gene (50). Rather, it has been proposed that transcription of the distal part of the snoRNA host gene might silence in *cis* the *Ube3A* paternal expression, either by RNA antisense mechanisms or by

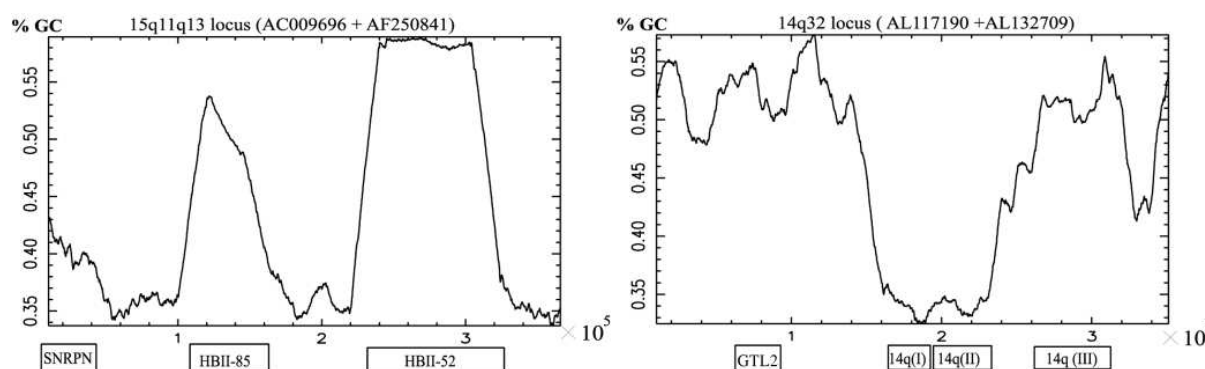


Figure 7. DNA isochores showing differences in G + C content within and around snoRNA gene clusters at the two human imprinted 15q11q13 and 14q32 loci. At the PWS/AS region, the paternally expressed snoRNA genes lie within a (G + C)-rich environment, while at 14q32 the maternally expressed snoRNA genes are located within an (A + T)-rich environment. Accession numbers of the sequences are given in bracket and the relative position of imprinted genes (boxed) is schematized.

altering local chromatin structure (21–23). Whether C/D snoRNA host genes at 14q32 also control the expression of neighbouring imprinted genes is an attractive hypothesis that can be tested using mouse models.

Finally, it is possible that the presence of these snoRNAs reflects a brain-specific function for these imprinted genes. The PWS and AS loci, which cause distinct neurological disorders, are linked to a locus containing snoRNAs. BWS (linked to the imprinted cluster on 11p15.5, where no snoRNAs have been noted to date) is a somatic overgrowth syndrome with no evidence of neurological problems. Aberrant imprinting on chromosome 14 is associated with both growth and neurological anomalies in mUPD14 and pUPD14 individuals. The former exhibit precocious puberty and are growth-retarded, while pUPD14 patients have defects of the axial skeleton and mental retardation (51–53). It is tempting to make functional associations between developmental growth regulators on chromosome 11p and on 14q, and between the neural-specific snoRNA genes on chromosome 15q and on 14q.

In conclusion, we have shown that a second imprinted locus in humans, 14q32, in addition to the previously reported 15q11q13, encodes arrays of tandemly repeated C/D snoRNA genes, and that this outstanding feature of the locus is conserved in human and rodents. Imprinted snoRNAs have only been detected within eutherian mammals so far (15,37). Given that genomic imprinting might be restricted to marsupials and placental mammals (54), a detailed sequence analysis of the orthologous 14q32 and 15q11q13 loci in non-eutherian mammals together with systematic sequence searches of a large set of imprinted loci in mammalian genomes should provide further insight into the potential significance of this intriguing association between genomic imprinting and the presence of these unusual tandemly repeated snoRNA genes.

MATERIALS AND METHODS

Oligonucleotides

These were all synthesised by Y. de Préval (LBME, Toulouse) on a PerSeptive Biosystems Expedite apparatus. 14q(0): 5'-GACCTCAGTGTGTTTGTGCAGAAC-3'; 14q(1): 5'-TGGACCTCAGAGTTCCAGACACGTATTCA-3'; 14(II): TGATTCAGAC/TACCCAG/CG/A G/ATACTCATC-3'; MBII-48: 5'-ATAAGGGTTTAATCACTGTCCT CGGTCA-3'; MBII-49: 5'-AATCCAGTATGTTGTCATCGTCTATG-3'; MBII-19: 5'-CAGACATCTG TTCTCATGGCT-3'; MBII-78: 5'-ACCTCAGATATCTGTTTCATGTCA-3'; MBII-52: 5'-CTGACGTAATCCTATTGAGCAT; MBII-85: 5'-ACA-GAGTTTTCACTCATTTTGTTC-3'; MBII-343: 5'-TCTCAG ACTTCCAGACATGTACT-3'; MBII-426: 5'-TGATCTCA-GAATTAAATTGTGTCG-3'; 5' RT-PCR-2: 5'-CGCGGATCC-GACGAGATTGGATTGGGTGCTATTTCC-3'; 3' RT-PCR-2: 5'-CCGGAATTCAGGCTCCTACCCAGAGGCAACTG-3'; 3' RT-PCR1: 5'-CGCGGATCCTATTGTCTCTATGCTCCT-TACTC-3'; 5' RT-PCR1: 5'-GCGGAATTCTACATGGATCC-CACCTTTGGACAAAG-3'; 5' RT-PCR3: 5'-GCGGAAT-TCTGGATGCAATGAGCTGATCA-3'; 3' RT-PCR3: 5'-CGCGGATCCTGAGGCTCAGAGGACGGCAG-3'; 5' RT-PCR4: 5'-CCGGAATTCGGATGGTTACTGTCTGA-

GACTGAG-3'; 3' RT-PCR4: 5'-CCGGAATTCTCATC-TATCCTCCTGACTCAGGAC-3'; 5' RT-PCR5: 5'-CCGGAATTCTTGCATGCAAGCTCTACAGTTATGC-3'; 3' RT-PCR5: 5'-CCGGAATTCAGCTGCCGAGCTCCATCCA-CATGGT-3'; 5' MBII-19/78-cluster: 5'-GGCTTTGATCCTT-CGGTTTGA-3'; 3' MBII-19/78-cluster: 5'-CCCTTCT GCTCCAAGTTTGCTA-3'.

Search for tandemly repeated C/D snoRNAs and sequence analysis

BAC sequences around DLK1/GTL2 positions have been retrieved from the Human Genome Project Working Draft (<http://genome.ucsc.edu/>), purged for common interspersed repeats by using Repeat Masker (<http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker>) and systematically compared with themselves by dotplot analysis conducted by PipMaker (<http://bio.cse.psu.edu/>). C/D snoRNA-like sequences and other repeats have been subsequently detected by BLAST2 sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>). CpG islands have been defined according to the Human Genome Project Working Draft annotations. Sequence alignment have been obtained by either Clustal W (<http://clustalw.genome.ad.jp/>) or MultiAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>), and conserved nucleotides have been shaded by GeneDoc (<http://www.Cris.com/~ketchup/genedoc.shtml>). Analysis of DNA isochores has been performed using Isochore (EMBOSS) at <http://bioweb.pasteur.fr/seqanal/interfaces/isochose.html>.

Mice with matUPD12 and patUPD12

E15.5 embryos with matUPD12 and patUPD12 and normal control littermates were generated by intercrossing parent animals of two different genetic backgrounds that were double heterozygotes for Robertsonian translocations with monobrachial homology for chromosome 12, as described previously (55).

RNA isolation and northern blot analysis

Total RNA was isolated by the method of Chomczynski and Sacchi (56), adapted to our conditions as follows. Rodent tissues freshly prepared according to French institutional and UK HO guidelines were quickly frozen in liquid nitrogen and stored at -80°C . Frozen samples were then homogenized and resuspended in a 1:1 (vol:vol) mixture of (i) 4 M guanidinium thiocyanate, 25 mM sodium citrate (pH 7.0), 0.5% Sarkosyl 0.1 M 2-mercaptoethanol and (ii) water-saturated phenol. This mixture was then supplemented with sodium acetate (pH 4.0) to 0.1 M. The lysate was mixed with 0.1 volume of chloroform, vigorously vortexed and incubated for 10–15 min at room temperature. RNA was precipitated with 2 volume of ethanol and stored at -20°C . RNAs were fractionated by electrophoresis in 6% acrylamide–7 M urea gels, transferred electrophoretically (120 min in $0.5 \times \text{TBE}$ at 1.0 A) to a nylon membrane (Hybond N+, Amersham) and crosslinked by ultraviolet irradiation (1200 J/cm^2 , Stratalinker, Stratagene, La Jolla, Calif.). Northern blot hybridizations were carried out in the presence of a 5'-end ^{32}P -labelled oligodeoxynucleotide probe (500 000 c.p.m./ml) by overnight incubation at 50°C in

5 × SSPE, 1% SDS, 5 × Denhardt's, 150 µg/ml yeast tRNA. Membranes were washed twice for 15 min at room temperature in 0.1 × SSPE/0.1%SDS.

RT-PCR and cDNA clones

Ten micrograms of total human or mouse brain RNA was reverse-transcribed at 42°C for 120 min, using Superscript II (Gibco BRL) with random hexamer primers and 1/40 of cDNA products amplified by 40 cycles of PCR with *Taq* polymerase (Promega). PCR product cloned into in pGEM-T easy system I vector (Promega) and IMAGE ESTs clones (T85042, AW026953, BF055187, AA451995, AA861571, AA910544, AA680166, AA910544 and AA433836) provided by UK-HGMP RC (<http://www.hgmp.mrc.ac.uk/>) have been sequenced by the CEQ 2000 DNA analysis system (Beckman).

ACKNOWLEDGEMENTS

We thank A. Hüttenhofer for helpful discussions throughout this work and C. Gaspin and P. Thebault for help with the use of sequence analysis soft-wares. We also thank N. Joseph for technical assistance in DNA sequencing and Y. de Préval for oligonucleotide synthesis (IEFG 109). We are grateful to Neil Youngson and Shau-Ping Lin for comments on the manuscript. This work was supported by laboratory funds from the Centre National de la Recherche Scientifique and Université Paul-Sabatier, Toulouse, and by grants from the Toulouse Genopole/Pole Santé (to J.P.B.), from La Ligue Contre le Cancer/Comité de Haute Garonne (to J.C.) and from the UK Medical Research Council (to A.C.F.S.).

REFERENCES

- Tilghman, S.M. (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell*, **96**, 185–193.
- Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.*, **2**, 21–32.
- Ferguson-Smith, A.C. and Surani, M.A. (2001) Imprinting and the epigenetic asymmetry between parental genomes. *Science*, **293**, 1086–1089.
- Paulsen, M. and Ferguson-Smith, A.C. (2001) DNA methylation in genomic imprinting, development, and disease. *J. Pathol.*, **195**, 97–110.
- Szabo, P., Tang, S.H., Rentsendorj, A., Pfeifer, G.P. and Mann, J.R. (2000) Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr. Biol.*, **10**, 607–610.
- Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*, **405**, 482–485.
- Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/*Igf2* locus. *Nature*, **405**, 486–489.
- Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.F., Wölfe, A., Ohlsson, R. and Lobanov, V.V. (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.*, **10**, 853–856.
- Nicholls, R.D. and Knepper, J.L. (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.*, **2**, 153–175.
- Buiting, K., Saitoh, S., Gross, S., Dittich, B., Schwartz, S., Nicholls, R.D. and Horsthemke, B. (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet.*, **9**, 395–400.
- Sutcliffe, J.S., Nakao, M., Christian, S., Orstavik, K.H., Tommerup, N., Ledbetter, D.H. and Beaudet, A.L. (1994) Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nat. Genet.*, **8**, 52–58.
- Shemer, R., Herskko, A.Y., Perk, J., Mostoslavsky, R., Tsuberi, B., Cedar, H., Buiting, K. and Razin, A. (2000) The imprinting box of the Prader-Willi/Angelman syndrome domain. *Nat. Genet.*, **26**, 440–443.
- Ben-Porath, I. and Cedar, H. (2000) Imprinting: focusing on the center. *Curr. Opin. Genet. Dev.*, **10**, 550–554.
- Cavaille, J., Buiting, K., Kieffmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachelier, J.P., Brosius, J. and Hüttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
- de los Santos, T., Schweizer, J., Rees, C.A. and Francke, U. (2000) Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCRI, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *Am. J. Hum. Genet.*, **67**, 1067–1082.
- Meguro, M., Mitsuya, K., Nomura, N., Kohda, M., Kashiwagi, A., Nishigaki, R., Yoshioka, H., Nakao, M., Oishi, M. and Oshimura, M. (2001) Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes. *Hum. Mol. Genet.*, **10**, 383–394.
- Filipowicz, W. (2000) Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *Proc. Natl Acad. Sci. USA*, **97**, 14035–14037.
- Kiss-Laszlo, Z., Henry, Y., Bachelier, J.P., Caizergues-Ferrer, M. and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
- Cavaille, J., Nicoloso, M. and Bachelier, J.P. (1996) Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, **383**, 732–735.
- Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.
- Runte, M., Hüttenhofer, A., Gross, S., Kieffmann, M., Horsthemke, B. and Buiting, K. (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.*, **10**, 2687–2700.
- Chamberlain, S.J. and Brannan, C.I. (2001) The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a. *Genomics*, **73**, 316–322.
- Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L. and Lalande, M. (1998) An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat. Genet.*, **19**, 15–16.
- Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F. and Barlow, D.P. (1997) Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature*, **389**, 745–749.
- Smilnich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W. *et al.* (1999) A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl Acad. Sci. USA*, **96**, 8064–8069.
- Wroe, S.F., Kelsey, G., Skinner, J.A., Bodle, D., Ball, S.T., Beechey, C.V., Peters, J. and Williamson, C.M. (2000) An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. *Proc. Natl Acad. Sci. USA*, **97**, 3342–3346.
- Reik, W. and Walter, J. (2001) Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nat. Genet.*, **27**, 255–256.
- Kelley, R.L. and Kuroda, M.I. (2000) Noncoding RNA genes in dosage compensation and imprinting. *Cell*, **103**, 9–12.
- Slueteels, F., Barlow, D.P. and Lyle, R. (2000) The uniqueness of the imprinting mechanism. *Curr. Opin. Genet. Dev.*, **10**, 229–233.
- Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N. and Georges, M. (2001) The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nat. Genet.*, **27**, 367–369.
- Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., Shay, T., Weissenbach, J., Cockett, N., Gyapay, G. *et al.* (2001) Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (*clpg*) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. *Genome Res.*, **11**, 850–862.
- Hatada, I., Morita, S., Obata, Y., Sotomaru, Y., Shimoda, M. and Kono, T. (2001) Identification of a new imprinted gene, Rian, on mouse chromosome

1538 *Human Molecular Genetics*, 2002, Vol. 11, No. 13

- 12 by fluorescent differential display screening. *J. Biochem. (Tokyo)*, **130**, 187–190.
33. Ishihara, K., Hatano, N., Furuumi, H., Kato, R., Iwaki, T., Miura, K., Jinno, Y. and Sasaki, H. (2000) Comparative genomic sequencing identifies novel tissue-specific enhancers and sequence elements for methylation-sensitive factors implicated in Igf2/H19 imprinting. *Genome Res.*, **10**, 664–671.
 34. Onyango, P., Miller, W., Lehoczky, J., Leung, C.T., Birren, B., Wheelan, S., Dewar, K. and Feinberg, A.P. (2000) Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.*, **10**, 1697–1710.
 35. Paulsen, M., Takada, S., Youngson, N.A., Bensaib, M., Charlier, C., Segers, K., Georges, M. and Ferguson-Smith, A.C. (2001) Comparative sequence analysis of the imprinted Dlk1–Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2–H19 region. *Genome Res.*, **11**, 2085–2094.
 36. Wylie, A.A., Murphy, S.K., Orton, T.C. and Jirtle, R.L. (2000) Novel imprinted DLK1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation. *Genome Res.*, **10**, 1711–1718.
 37. Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A. and Bachellerie, J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.
 38. Komine, Y., Tanaka, N.K., Yano, R., Takai, S., Yuasa, S., Shiroishi, T., Tsuchiya, K. and Yamamori, T. (1999) A novel type of non-coding RNA expressed in the rat brain. *Brain Res. Mol. Brain Res.*, **66**, 1–13.
 39. Schmidt, J.V., Matteson, P.G., Jones, B.K., Guan, X.J. and Tilghman, S.M. (2000) The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. *Genes Dev.*, **14**, 1997–2002.
 40. Takada, S., Tevendale, M., Baker, J., Georgiades, P., Campbell, E., Freeman, T., Johnson, M.H., Paulsen, M. and Ferguson-Smith, A.C. (2000) Delta-like and gtl2 are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.*, **10**, 1135–1138.
 41. Huttenhofer, A., Kiehnmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P. and Brosius, J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
 42. Takada, S., Paulsen, M., Tevendale, M., Tsai, C.E., Kelsey, G., Cattanach, B.M. and Ferguson-Smith, A.C. (2002) Epigenetic analysis of the Dlk1–Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2–H19. *Hum. Mol. Genet.*, **11**, 77–86.
 43. d'Orval, B.C., Bortolin, M.L., Gaspin, C. and Bachellerie, J.P. (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp}. *Nucleic Acids Res.*, **29**, 4518–4529.
 44. Wirth, J., Back, E., Huttenhofer, A., Nothwang, H.G., Lich, C., Gross, S., Menzel, C., Schinzel, A., Kioschis, P., Tommerup, N. *et al.* (2001) A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF–SNRPN transcription unit on chromosome 15. *Hum. Mol. Genet.*, **10**, 201–210.
 45. Bachellerie, J.P. and Cavaille, J. (1997) Guiding ribose methylation of rRNA. *Trends Biochem. Sci.*, **22**, 257–261.
 46. Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897–934.
 47. Bachellerie, J.P., Nicoloso, M., Qu, L.H., Michot, B., Caizergues-Ferrer, M., Cavaille, J. and Renalier, M.H. (1995) Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem. Cell Biol.*, **73**, 835–843.
 48. Park, Y. and Kuroda, M.I. (2001) Epigenetic aspects of X-chromosome dosage compensation. *Science*, **293**, 1083–1085.
 49. Sleutels, F., Zwart, R. and Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 810–813.
 50. Tsai, T.F., Jiang, Y.H., Bressler, J., Armstrong, D. and Beaudet, A.L. (1999) Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader–Willi syndrome. *Hum. Mol. Genet.*, **8**, 1357–1364.
 51. Healey, S., Powell, F., Battersby, M., Chenevix-Trench, G. and McGill, J. (1994) Distinct phenotype in maternal uniparental disomy of chromosome 14. *Am. J. Med. Genet.*, **51**, 147–149.
 52. Georgiades, P., Chierakul, C. and Ferguson-Smith, A.C. (1998) Parental origin effects in human trisomy for chromosome 14q: implications for genomic imprinting. *J. Med. Genet.*, **35**, 821–824.
 53. Cotter, P.D., Kaffé, S., McCurdy, L.D., Jhaveri, M., Willner, J.P. and Hirschhorn, K. (1997) Paternal uniparental disomy for chromosome 14: a case report and review. *Am. J. Med. Genet.*, **70**, 74–79.
 54. Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., MacDonald, R.G. and Jirtle, R.L. (2000) M6P/IGF2R imprinting evolution in mammals. *Mol. Cell*, **5**, 707–716.
 55. Georgiades, P., Watkins, M., Surani, M.A. and Ferguson-Smith, A.C. (2000) Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development*, **127**, 4719–4728.
 56. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
 57. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

2.1.2 Recherche de nouvelles répétitions de gènes d'ARN C/D

Les ARN C/D MBII-48, MBII-78 et RBII-36, dont la localisation subcellulaire a été déterminée par hybridation *in situ*, sont nucléolaires (Jérôme Cavaillé, communication personnelle) : ces ARN sont donc des snoRNA (petits ARN nucléolaires), et il semble probable que les autres ARN C/D exprimés à partir de ce locus le soient aussi.

Il est particulièrement remarquable que trois caractéristiques, inédites jusqu'alors pour des gènes de snoRNA (la répétition en tandem ; l'expression restreinte au cerveau ; l'empreinte génomique parentale), se retrouvent dans deux loci distincts (15q11-13 et 14q32 chez l'Homme) ; l'association de ces particularités reflétait peut-être une relation de causalité entre elles (voir discussion, page 152). Si tel était le cas, d'autres loci des génomes de Mammifères pourraient, à la fois, contenir des gènes de snoRNA répétés, et être soumis à l'empreinte génomique parentale.

J'ai donc écrit un programme, dans le langage fortran, qui recherche les gènes de snoRNA à boîtes C et D dont la structure est la plus canonique :

- il recherche les boîtes C potentielles (de consensus : **RUGAUGA**), en tolérant une déviation ;
- il recherche les boîtes D potentielles (de consensus : **CUGA**), en tolérant une déviation ;
- il assemble les boîtes C et D par paires : lorsqu'une boîte D potentielle suit une boîte C potentielle (à une distance comprise entre deux limites fixées par l'utilisateur), et que les deux boîtes totalisent au maximum une déviation à elles deux, par rapport aux consensus, il poursuit l'analyse sur cette paire de boîtes potentielles ;
- parmi les paires de boîtes potentielles retenues, il sélectionne celles qui sont flanquées de régions complémentaires (*i.e.* : la séquence en amont du nucléotide qui précède la première purine du consensus de la boîte C doit être complémentaire de la séquence immédiatement en aval de la boîte D) ; l'utilisateur fixe la qualité minimale de cette complémentarité.

Un deuxième programme découpe toute la séquence du génome humain en tranches de 25 kb, et recherche les candidats gènes de snoRNA dans chacune de ces tranches¹, avec le premier programme. Le résultat de cette recherche se présente donc sous la forme d'histogrammes, présentant pour chaque tranche de 25 kb le nombre de candidats identifiés ; un fichier annexe contient les séquences de ces candidats.

La plupart des pics de ces histogrammes sont dus à des séquences de faible complexité (souvent des répétitions $(TGA)_n$ peu dégénérées), où le programme trouve de nombreux candidats chevauchants. Ces pics ne représentent donc pas des répétitions de gènes de snoRNA C/D, mais plutôt un unique gène-candidat, possédant plusieurs boîtes C ou D alternatives. Afin d'éliminer ces faux positifs, j'ai comparé les candidats de chacune des fenêtres de 25 kb contenant au moins 10 candidats (en incluant les candidats totalisant une déviation par rapport aux consensus pour leurs boîtes C et D) : après élimination de ces candidats chevauchants, aucun locus ne contient de répétitions de candidats apparentés entre eux, sauf les deux loci déjà connus, 14q32 et 15q11-13 : alors que les loci 14q32 et 15q11-13, sont très riches en gènes-candidats sans déviation sur les boîtes C et D (jusqu'à 15 candidats sans déviation par 25 kb pour le locus 14q32, et jusqu'à 17 pour le locus 15q11-13), aucun autre locus n'en contient plus de 8 par 25 kb.

Aucun autre locus que 14q32 et 15q11-13 ne contient donc de gènes de snoRNA C/D répétés dans les mêmes proportions que ces deux loci.

La figure 2.1 donne, à titre d'exemple, le résultat de l'analyse du chromosome 14 (en orientation sens, selon les conventions de polarité des séquences du NCBI), et la figure 2.2 donne le résultat de l'analyse du chromosome 14 en orientation antisens (les gènes des snoRNA du locus 14q32 sont dans cette orientation).

¹Paramètres utilisés pour cette analyse : distance entre le premier U de la boîte C et le C de la boîte D comprise entre 50 et 120 pb ; la complémentarité entre les 5 nt de séquences flanquantes des boîtes doit être longue d'au moins 4 pb, en tolérant un appariement G-U .

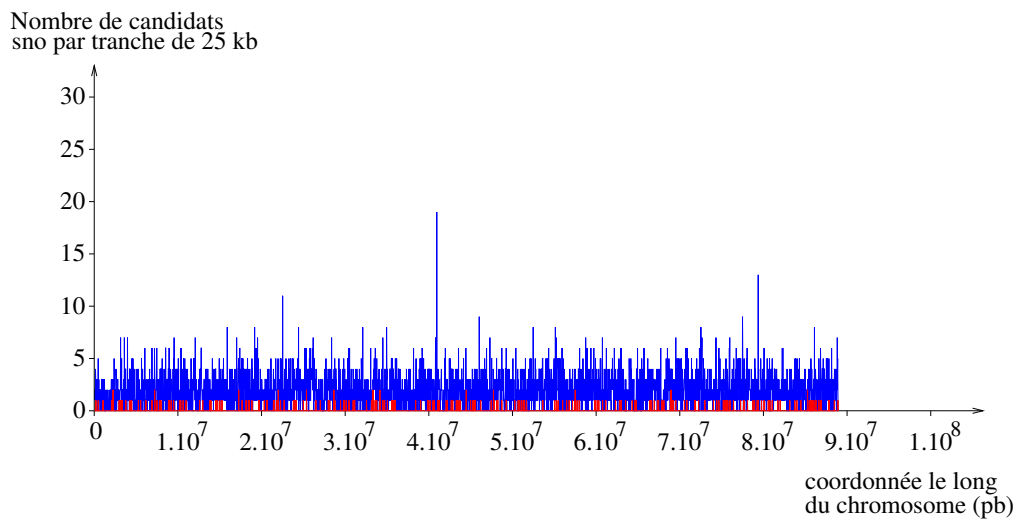


FIG. 2.1 – RECHERCHE DE GÈNES DE SNORNA C/D DANS LE CHROMOSOME 14 HUMAIN (ORIENTATION SENS). Les pics bleus indiquent les nombres de candidats par 25 kb présentant une déviation par rapport aux consensus, pour l'ensemble de leurs deux boîtes. Les pics rouges indiquent les nombres de candidats par 25 kb ne présentant aucune déviation par rapport aux consensus.

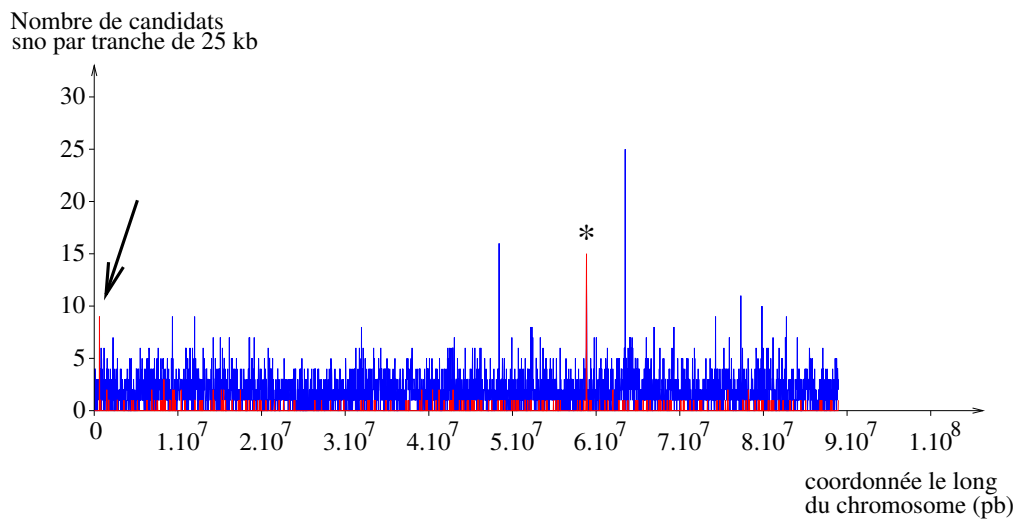


FIG. 2.2 – RECHERCHE DE GÈNES DE SNORNA C/D DANS LE CHROMOSOME 14 HUMAIN (ORIENTATION ANTISENS). Les pics bleus indiquent les nombres de candidats présentant une déviation par rapport aux consensus, pour l'ensemble de leurs deux boîtes. Les pics rouges indiquent les nombres de candidats ne présentant aucune déviation par rapport aux consensus. Le locus 14q32 est signalé par l'astérisque. Le pic de candidats signalé par une flèche est un faux positif, dû à des répétitions $(TGA)_n$, où le programme trouve de nombreux candidats chevauchants.

Par ailleurs, le programme de détection des gènes d'ARN C/D (*i.e.* le premier des deux programmes décrits page 115) a été utilisé dans une autre étude (voir article « Identification of 13 novel human modification guide RNAs », en annexe 1, page 163), où il a permis de découvrir trois nouveaux gènes de snoRNA C/D humains : U96a, U101 et U102 (Vitali *et al.*, 2003).

Une recherche de complémentarités entre les ARN C/D du locus 14q32 humain (ainsi que ceux du locus humain 15q11-13) et le précurseur de 45S de l'ARN ribosomique humain ne met en évidence aucune complémentarité qui ne soit pas statistiquement attendue : cette analyse est présentée en annexe 2, page 173.

2.2 Le locus 14q32/distal 12 contient des gènes de microARN

L'article présenté dans la section précédente, qui décrit les gènes d'ARN C/D répétés au locus 14q32 humain, et dans ses loci orthologues chez les Rongeurs, signale l'existence de séquences répétées (appelées « cluster R » dans l'article) en aval des gènes d'ARN C/D. Dans cette première analyse, deux familles ont été identifiées : la famille A (avec 22 membres chez *Homo sapiens*) et la famille B (6 membres chez *H. sapiens*). Dans chacune de ces familles, le motif répété est une séquence d'environ 80 à 100 pb. Ce motif est lui-même, dans chacune des deux familles, une répétition inversée imparfaite. Donc, une transcription de ces séquences produirait des ARN susceptibles de se replier en tiges-boucles imparfaites de 80 à 100 nt.

À l'occasion d'un travail de clonage de microARN à partir de différents organes de Souris, une autre équipe a découvert 34 nouveaux microARN (Lagos-Quintana *et al.*, 2002). Nous nous sommes intéressés à quatre d'entre eux, exprimés dans le cerveau, dont les gènes sont situés dans le locus distal 12, et les orthologues humains, dans le locus 14q32 ; en particulier, les gènes de deux d'entre eux (miR-127 et miR-136) se trouvent dans un gène prédit, *Rtl1*², et seraient transcrits en orientation antisens de *Rtl1*. Quant aux deux autres gènes de microARN identifiés dans le locus, les gènes de miR-134 et miR-154, ils se trouvent dans les répétitions du « cluster R », et plus précisément, le gène de miR-154 est l'une des répétitions de la famille A (la 13ème chez *M. musculus*). Puisque ces gènes se trouvent dans un locus soumis à l'empreinte génomique parentale, ils pouvaient y être eux-mêmes soumis.

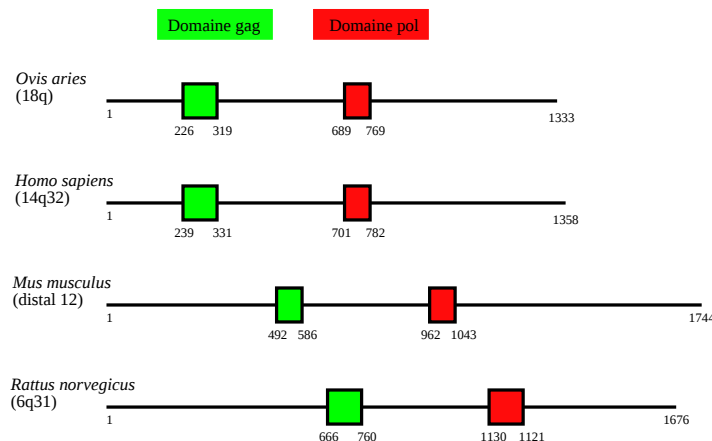


FIG. 2.3 – LE GÈNE *Rtl1* EST UN RÉTROTRANSPOSON CONSERVÉ CHEZ LES MAMMIFÈRES. Ces schémas représentent les gènes *Rtl1* chez le Mouton (*O. aries*), l'Homme (*H. sapiens*), la Souris (*M. musculus*) et le Rat (*R. norvegicus*). Le gène se trouve dans le locus 18q du Mouton, et ses orthologues 14q32 chez l'Homme, distal 12 chez la Souris, et 6q31 chez le Rat. Les nombres sous les schémas indiquent les positions des acides aminés qui délimitent des domaines remarquables. Les domaines gag (en vert) et pol (en rouge), caractéristiques des rétrotransposons et rétrovirus, sont clairement reconnaissables dans le gène *Rtl1* de chaque espèce.

Nous avons donc analysé leur expression dans des embryons de Souris présentant des disomies uniparentales pour le chromosome 12, et montré qu'ils sont exprimés à partir de l'allèle maternel uniquement, alors que *Rtl1* (dont nous montrons qu'il est exprimé ; nous avons appelé ce nouveau gène *Rtl1* pour : *retrotransposon-like gene n°1*) est exprimé à partir de l'allèle paternel uniquement. *Rtl1*, également appelé *PEG11* (Charlier *et al.*, 2001b) et *Shr1* (Lynch et

²Cette séquence est l'orthologue d'un gène caractérisé chez le Mouton, *PEG11* (Charlier *et al.*, 2001b), et elle s'apparente aux rétrotransposons (les domaines gag et pol y sont clairement identifiables : voir figure 2.3).

Tristem, 2003) chez le Mouton, *Hur1* chez l'Homme, *Rar1* chez le Rat, et *Mor1* chez la Souris (Lynch et Tristem, 2003), appartient à la famille *gypsy* de rétrotransposons à longues répétitions terminales (LTR) ; lui-même est dépourvu de LTR. Enfin, puisque miR-127 et miR-136 sont transcrits en antisens de *Rtl1*, ils sont parfaitement complémentaires à son ARN_m (ce sont les premiers microARN de Vertébrés parfaitement complémentaires d'un ARN endogène), et sont donc susceptibles d'en guider la dégradation par RNAi. Leur empreinte réciproque avec *Rtl1* pourrait même signifier qu'ils sont responsables de l'empreinte de ce gène : s'ils pouvaient guider spécifiquement la dégradation de l'ARN_m issu de l'allèle maternel de *Rtl1*, ou s'ils pouvaient guider une modification épigénétique particulière sur leur allèle maternel (voir section 1.4.5, page 80, pour la relation entre petits ARN et modifications épigénétiques, et la discussion, page 154), ils pourraient réprimer l'allèle maternel de *Rtl1*, donc lui imposer le monoallélisme de son expression. Mais il faudrait concevoir une réaction de RNAi (qui est *a priori* cytoplasmique ; en particulier, la maturation finale des microARN semble cytoplasmique chez les Mammifères : voir page 65) qui parvienne à discerner l'origine allélique (maternelle ou paternelle) des ARN_m : dans cette hypothèse, il faudrait par exemple que la réaction de RNAi soit nucléaire, localisée à proximité de l'allèle maternel du locus, que les deux microARN soient complètement maturés localement, au voisinage de leur site de transcription, et que le complexe RISC guidé par miR-127 et miR-136 ne puisse pas diffuser jusqu'à l'allèle paternel. Ces hypothèses sont discutées page 158.

Ces travaux sont présentés dans l'article qui suit, « Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene » (Seitz *et al.*, 2003).

2.2.1 Article n°2

BRIEF COMMUNICATIONS

Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene

Hervé Seitz¹, Neil Youngson², Shau-Ping Lin², Simone Dalbert³, Martina Paulsen³, Jean-Pierre Bachelier¹, Anne C Ferguson-Smith² & Jérôme Cavaille¹

MicroRNAs (miRNAs) are an abundant class of RNAs that are ~21–25 nucleotides (nt) long, interact with mRNAs and trigger either translation repression or RNA cleavage (RNA interference, RNAi) depending on the degree of complementarity with their targets. Here we show that the imprinted mouse distal chromosome 12 locus encodes two miRNA genes expressed from the maternally inherited chromosome and antisense to a retrotransposon-like gene (*Rtl1*) expressed only from the paternal allele.

We mapped four recently reported mouse tissue-specific miRNA genes¹ to an imprinted locus on mouse distal chromosome 12/human 14q32 (Fig. 1). Two of these, *mir-136* and *mir-127*, are located near two CpG islands in a mouse retrotransposon-like gene, *Rtl1*, which is expressed exclusively from the paternal allele (Fig. 2a,b). These miRNA genes are transcribed in an antisense orientation to *Rtl1* and are expressed exclusively from the maternal chromosome (Fig. 2a,b). The mechanism by which *mir-127* and *mir-136* are generated has yet to be determined, but *mir-127* does not seem to be processed from long double-stranded RNAs formed by *Rtl1* and anti-*Rtl1* RNAs, as we did not detect any RNA species ~21–23 nt long corresponding to anti-*mir-127* (data not shown). Moreover, RT-PCR analysis indicates that these miRNAs are generated from a larger host transcript (Fig. 2b). The other two miRNA genes, *mir-154* and *mir-134*, are located in a cluster of conserved repeated sequences² (A-repeats; Fig. 1). The gene *mir-154* (A13) and several copies of related

sequences are intron-encoded in *Mirg* (microRNA-containing gene), a novel maternally expressed gene that lacks an open reading frame (Figs. 1 and 2c). As we could not detect them by northern-blot analysis, their imprinted status and mode of biosynthesis from *Mirg* introns have yet to be resolved.

Although some *Rtl1* RNA is clearly detectable (Fig. 2a), the perfect complementarity between *mir-136*, *mir-127* and the *Rtl1* open reading frame (Fig. 2d) suggests they might function as small interfering RNAs^{3,4} to silence *Rtl1*, presumably in a tissue-specific manner. For instance, such a role could be achieved in the embryo brain or in the placenta, in which *mir-127* and *Rtl1* are co-expressed and reciprocally imprinted (data not shown). Moreover, an involvement of miRNAs in *Rtl1* imprinting cannot be formally ruled out. Although RNA interference is believed to be a *trans*-acting mechanism, preferential nuclear degradation of the *Rtl1* transcript from the maternal allele only, mediated by the sense/antisense gene organization of the *Rtl1* miRNA genes, could account for *Rtl1* imprinting. Alternatively or additionally, miRNAs might mediate chromatin epigenetic changes specifically at the maternal *Rtl1* allele, consistent with the involvement of the RNAi machinery in heterochromatin formation^{5–7}. RNA-directed allele-specific methylation of the two CpG islands in *Rtl1* is unlikely, as they are hypermethylated on both alleles and this occurs even in pUPD12 embryos in which miRNAs

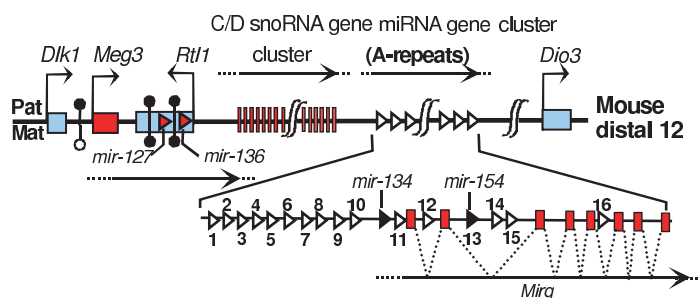


Figure 1 Several microRNA genes map within an imprinted locus on mouse distal 12 chromosome/human 14q32. The position of several imprinted genes is indicated by squares (*Meg3* and protein-coding genes), vertical bars (small nucleolar RNA genes) or triangles (miRNA genes). Experimentally detected¹ or predicted repeated gene copies of miRNAs (with hairpin structure ~70 nt long; data not shown) are indicated by filled (red) or open triangles, respectively. Maternally expressed (red) and paternally expressed (blue) genes indicate the imprinting status. Several methylated regions are represented by circles (filled, hypermethylated; open, hypomethylated). In *Mirg*, spliced exons detected by genomic/cDNA comparison are represented by red boxes with dotted lines denoting splicing events. The maternally expressed antisense transcript to *Rtl1* was initially suggested in sheep¹⁴ (called anti-PEG11). The schematic is not drawn to scale.

¹LBME, Centre National de la Recherche Scientifique (UMR 5099), IFR 109, Université P. Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France.

²Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. ³Universität des Saarlandes, FR Genetik, Postfach 151150, D-66041 Saarbrücken, Germany. Correspondence should be addressed to J.C. (cavaille@bcbg.biotoul.fr).

BRIEF COMMUNICATIONS

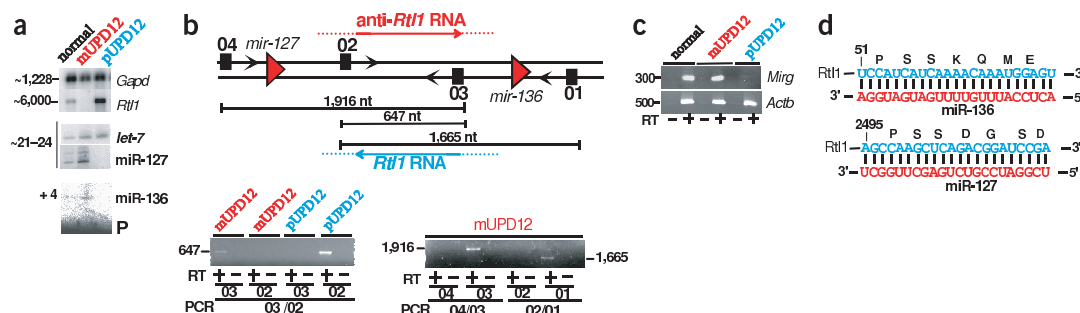


Figure 2 miR-127 and miR-136, two miRNAs expressed from the maternal chromosome, are antisense to *Rtl1*, a retrotransposon-like gene expressed only from the paternally inherited chromosome. (a) Reciprocal imprinting of *Rtl1*, *mir-127* and *mir-136*. Total or poly(A)⁺ RNAs were isolated from normal embryos and embryos with maternal or paternal uniparental disomy for chromosome 12 (mUPD12 and pUPD12, respectively). Paternal allelic expression of *Rtl1* imprinting is evident on northern blots of poly(A)⁺ RNAs using normal and UPD12 embryos (top). Maternal-specific expression of the miRNAs is indicated by northern blots for miR-127 (middle) or by primer extension for miR-136 (bottom). P, ³²P-labeled primer. Probes for *Gapdh* and *let-7* were used as internal gel loading controls. Size is indicated in nucleotides. For unknown reasons, miR-136 could be detected only by primer extension. (b) Strand-specificity of maternally and paternally transcribed transcripts at *Rtl1*. Total RNA from mUPD12 or pUPD12 embryos was subjected to strand-specific RT-PCR with appropriate primers (depicted on top). The antisense RNA was detected only in mUPD12 RNA, whereas the sense RNA was detected only in pUPD12 RNA (bottom left). miR-127 and miR-136 are transcribed as a large precursor (bottom right). In each case, the primers used are indicated under the corresponding lanes (01–04; sequences available on request). RT, reverse transcriptase. (c) *Mirg* is maternally expressed in mouse embryos. Total RNA from normal, mUPD12 or pUPD12 embryos was reverse transcribed with primer *Mirg8R*. Subsequent PCR with primers *Mirg7F* and *Mirg8R* (primer sequences available on request) showed that *Mirg* was detected only in normal and mUPD12 RNA. *Mirg7F* and *Mirg8R* are located in the last two exons of the transcript (Fig. 1). As an additional control, RT-PCR (30 cycles) was done on the biallelically expressed gene *Actb* (encoding β -actin). RT, reverse transcriptase. (d) Schematic representation of miR-127 and miR-136 complementary sites in the *Rtl1* open reading frame (nucleotide numbering begins at the start codon).

are not expressed (N.Y., S.-P.L. and A.C.F.-S., manuscript in preparation). Efforts are underway to decipher whether these miRNAs function in the regulation of *Rtl1* gene expression during mouse development or in adults.

The reciprocal imprinting of protein-coding genes and non-coding RNAs in imprinted clusters is a recurrent theme and suggests a regulatory role for non-coding RNAs in autosomal euchromatic epigenetic control⁸. Except for a direct role of the *Air* antisense RNA in the silencing of imprinted genes on mouse chromosome 17 (ref. 9), the involvement of RNA *per se* has not yet been shown. In contrast to plant miRNAs^{3,10}, none of the metazoan miRNAs reported so far has perfect complementarity to any mRNA. Thus, the imprinted miR-127 and miR-136 are the first miRNAs fully complementary to an endogenous cellular mRNA in animals, suggesting that the RNAi machinery could regulate this gene. RNA interference is considered a defense mechanism against mobile parasitic elements, including retrotransposons and retroviral-like sequences¹¹; several imprinted genes share some features of these types of sequences^{12,13}, notably the mouse gene *Rtl1* identified as imprinted in this study. Thus, these findings identify a potential link between miRNAs, retrotransposon silencing and genomic imprinting.

Animal care. All animal research was conducted in accordance with UK Government Home Office Licensing procedures.

GenBank accession numbers. *Mirg* cDNA sequence, AJ517767; *Rtl1*, BK001261.

ACKNOWLEDGMENTS

This work was supported by grants from the Toulouse Génomique, the Programme Interdisciplinaire du Centre National de la Recherche Scientifique Dynamique et Réactivité des Assemblages Biologiques and La Ligue contre le Cancer/Comité de Haute-Garonne (to J.C.) and by laboratory funds from the Centre National de la Recherche Scientifique and Université Paul-Sabatier, Toulouse (to J.-P.B.). Work in M.P.'s laboratory was supported by the Deutsche Forschungsgemeinschaft. N.Y. is supported by a BBSRC studentship and S.-P. L. by a scholarship from the Taiwanese government. H.S. is supported by a PhD fellowship Allocation de Moniteur Normalien (Ministère de l'Éducation Nationale, de la Recherche et de la Technologie).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 11 March; accepted 18 April 2003

Published online 8 June 2003; doi:10.1038/ng1171

- Lagos-Quintana, M. *et al.* *Curr. Biol.* **12**, 735–739 (2002).
- Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C. & Bachellerie, J.P. *Hum. Mol. Genet.* **11**, 1527–1538 (2002).
- Llave, C., Xie, Z., Kasschau, K.D. & Carrington, J.C. *Science* **297**, 2053–2056 (2002).
- Hutvagner, G. & Zamore, P.D. *Science* **297**, 2056–2060 (2002).
- Reinhart, B.J. & Bartel, D.P. *Science* **297**, 1831 (2002).
- Volpe, T.A. *et al.* *Science* **297**, 1833–1837 (2002).
- Hall, I.M. *et al.* *Science* **297**, 2232–2237 (2002).
- Rougeulle, C. & Heard, E. *Trends Genet.* **18**, 434–437 (2002).
- Sleutels, F., Zwart, R. & Barlow, D.P. *Nature* **415**, 810–813 (2002).
- Rhoades, M.W. *et al.* *Cell* **110**, 513–520 (2002).
- Plasterk, R.H. *Science* **296**, 1263–1265 (2002).
- Neumann, B., Kubicka, P. & Barlow, D.P. *Nat. Genet.* **9**, 12–13 (1995).
- Yoder, J.A., Walsh, C.P. & Bestor, T.H. *Trends Genet.* **13**, 335–340 (1997).
- Charlier, C. *et al.* *Genome Res.* **11**, 850–862 (2001).

2.2.2 Stabilisation de l'ARN_m Rtl1 en l'absence de miR-127 et miR-136

Une région différentiellement méthylée (hypométhylée sur l'allèle maternel, et hyperméthylée sur l'allèle paternel) se trouve entre les gènes *Dlk1* et *Gtl2* (voir Introduction, page 37). Cette séquence acquiert son patron de méthylation au cours de la gamétogenèse (Takada *et al.*, 2002), ce qui en fait un bon candidat pour jouer le rôle de centre d'empreinte dans ce locus. Les membres du laboratoire d'Anne Ferguson-Smith ont donc réalisé une délétion de cette séquence chez la Souris, et caractérisé l'expression de *Dlk1*, *Gtl2*, *Rtl1* et *Dio3* suite à la transmission de cette délétion par le père, et par la mère. Nous avons analysé l'expression de plusieurs gènes d'ARN C/D, ainsi que de miR-127, chez les mêmes mutants.

Lorsque la DMR est mutée sur le chromosome maternel, tous les gènes à expression maternelle deviennent réprimés bialléliquement, et tous les gènes à expression paternelle sont désormais exprimés à partir des deux allèles³ (voir figure 2.4, page 123). Par conséquent, les ARN_m de *Dlk1* et *Dio3* sont sur-accumulés, d'un facteur 2. En revanche, la sur-accumulation des ARN_m de *Rtl1* (d'un facteur 4 à 5) ne peut pas s'expliquer uniquement par la transcription biallélique. L'interprétation la plus simple est que, dans ce mutant, chaque allèle de *Rtl1* est transcrit autant que l'unique allèle exprimé (l'allèle paternel) chez le sauvage, mais que les ARN_m de *Rtl1* sont stabilisés post-transcriptionnellement.

Cette stabilisation s'interprète facilement si les microARN miR-127 et miR-136 guident, chez l'embryon sauvage, une dégradation par RNAi de l'ARN_m Rtl1 : puisque, dans ce contexte génétique, les microARN miR-127 et miR-136 sont réprimés, leur absence devrait se traduire par une stabilisation de l'ARN_m Rtl1. Étant données les valeurs observées, l'expression de *Rtl1* serait alors réprimée d'un facteur 2 à 2,5 par RNAi chez l'embryon sauvage.

Les résultats de ce travail sont présentés dans l'article qui suit, « Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12 » (Lin *et al.*, 2003).

³Ce que démontre l'analyse des ARN d'embryons hétérozygotes pour des sites de polymorphisme de séquence : si l'allèle maternel de leur DMR est délété, on détecte les ARN issus de la transcription des deux allèles.

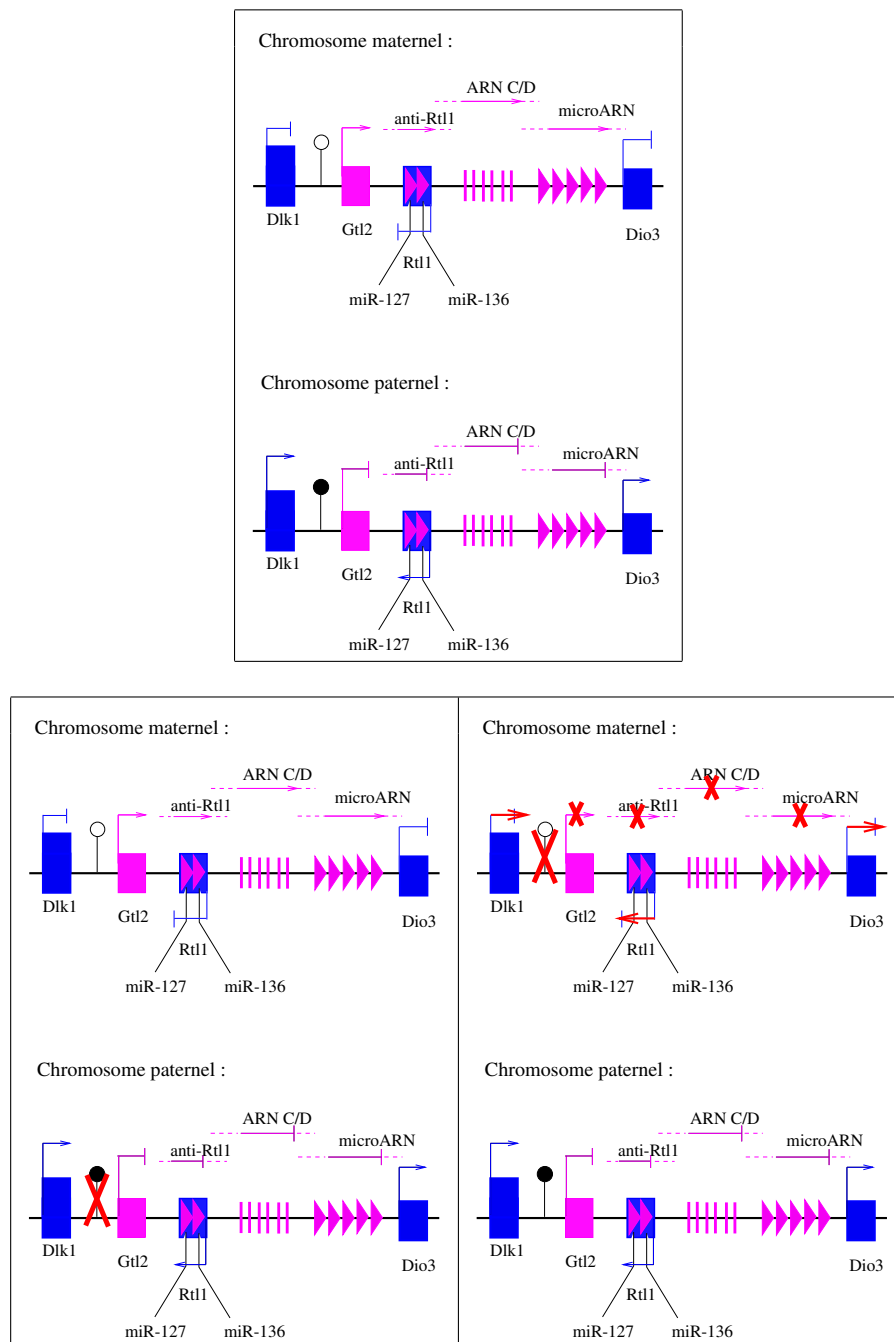


FIG. 2.4 – DÉLÉTION DE LA DMR DU LOCUS *Dlk1/Gtl2*. En haut : situation sauvage (les gènes à expression maternelle sont représentés en rose, et les gènes à expression paternelle, en bleu) ; l'allèle hypométhylé de la DMR est représenté par un disque blanc, et l'allèle hyperméthylé par un disque noir. En bas, à gauche : lorsque la DMR est délétee (croix rouge, sur la figure) sur le chromosome paternel, l'empreinte et le niveau d'expression des gènes du locus ne sont pas affectés. En bas, à droite : lorsque la DMR est délétee sur le chromosome maternel, les gènes habituellement à expression maternelle sont réprimés bi-alléliquement, et les gènes habituellement à expression paternelle sont exprimés bi-alléliquement.

2.2.3 Article n°3

LETTERS

nature
geneticsAsymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12Shau-Ping Lin¹, Neil Youngson¹, Shuji Takada^{1,4}, Hervé Seitz², Wolf Reik³, Martina Paulsen^{1,4}, Jerome Cavaille² & Anne C Ferguson-Smith¹

Genomic imprinting causes parental origin-specific gene expression¹. *Cis*-acting regulatory elements that control imprinting are not fully understood but involve regions that become differentially methylated on the two parental chromosomes during male and female gametogenesis^{2–4}. Understanding properties of maternally and paternally inherited imprints provides insight into the mechanisms and evolution of genomic imprinting. Previously we identified an intergenic germline-derived differentially methylated region (IG-DMR) that is a candidate control element for an imprinted domain on distal mouse chromosome 12 (ref. 5). The 1-Mb cluster contains the paternally expressed protein-coding genes *Dlk1* (refs. 6,7) and *Dio3* (ref. 8,9) and several maternally expressed non-coding RNAs, including *Gtl2* (refs. 6,7,10) and C/D snoRNAs¹¹. A retrotransposon-like gene (*Rtl1*) is expressed from the paternal chromosome and has an antisense transcript expressed from the maternal chromosome containing two microRNAs with full complementarity to *Rtl1* (ref. 12). Here we show that deletion of the IG-DMR from the maternally inherited chromosome causes bidirectional loss of imprinting of all genes in the cluster. When the deletion is transmitted from the father, imprinting is unaltered. These results prove that the IG-DMR is a control element for all imprinted genes on the maternal chromosome only and indicate that the two parental chromosomes control allele-specific gene expression differently.

The organization of the 1-Mb imprinted locus on distal mouse chromosome 12 is illustrated in Figure 1a. To determine the mechanism of imprinting at the *Dlk1-Gtl2* domain, we generated a targeted deletion of the IG-DMR between *Dlk1* and *Gtl2*. We deleted 4.15 kb including the tandem repeats^{5,13} of the IG-DMR (Fig. 1b) in embryonic stem (ES) cells, which we then injected into host blastocysts to generate male chimeras that successfully transmitted the deletion through the

germ line (Fig. 1c–e). Heterozygous IG-DMR-deleted progeny were generated at the expected frequency at embryonic day (E) 16. At later gestational stages, the percentage of live heterozygous embryos fell when the deletion was maternally inherited but remained the same when the deletion was paternally inherited. We found no adult mice carrying the maternally transmitted deletion (in five litters at 4 weeks of age; $n = 32$). Thirteen neonates in seven litters were heterozygous with respect to the maternally inherited deletion, but they were born dead or died within several hours of birth. Therefore, maternal transmission of the IG-DMR deletion is lethal after E16.

We next quantified the expression of *Dlk1* and *Gtl2* transcripts in embryos after maternal and paternal transmission of the deleted IG-DMR. After maternal transmission, *Gtl2* expression was negligible, indicating repression of *Gtl2* from the usually expressed maternal allele (Fig. 2), and *Dlk1* expression was twice that observed in wild-type littermates (Fig. 2). We confirmed that the increase in *Dlk1* expression was caused by activation of the usually silent maternal allele using a single-nucleotide polymorphism in DBA/2 and sequencing of RT-PCR products in embryos from females heterozygous with respect to the deletion (129Ola/C57BL6) crossed to DBA/2 males (Fig. 3a). Therefore, deletion of the unmethylated copy of the IG-DMR on the maternal allele results in biallelic expression of *Dlk1* and biallelic repression of *Gtl2*.

After paternal inheritance of the IG-DMR deletion, overall expression of *Dlk1* was the same as that observed in wild-type littermates and was from the paternal allele (Figs. 2 and 3a). Likewise, expression of *Gtl2* remained from the maternal allele, indicating that the absence of the methylated IG-DMR did not result in loss of imprinting of *Gtl2* on the paternal chromosome. These data show that the methylated IG-DMR on the paternal allele is not involved in *Dlk1* activity and is not necessary for full repression of *Gtl2* on the paternal chromosome.

The *Gtl2* promoter is a DMR that becomes methylated on the paternal allele after fertilization⁶. We carried out methylation analysis to determine whether repression of *Gtl2* after maternal transmission of

¹Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. ²LBME-CNRS (UMR5099), Université P. Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France. ³Laboratory of Developmental Genetics and Imprinting, Developmental Genetics Programme, The Babraham Institute, Cambridge CB2 4AT, UK. ⁴Present addresses: Institute for Molecular Bioscience, The University of Queensland, Brisbane Qld 4072, Australia (S.T.); Universität des Saarlandes, FR Genetik, Postfach 151150, D-66041 Saarbrücken, Germany (M.P.).

Published online 24 August 2003; doi:10.1038/ng1233

LETTERS

the deletion was associated with a change in promoter methylation. The promoter was hypermethylated relative to the methylation observed in embryos with paternal uniparental disomy for chromosome 12 (pUPD12; refs. 5,14; Fig. 4). We also observed slightly less methylation of the *Gtl2* promoter DMR after paternal inheritance of the deletion. This suggests that the maternal allele acquired a paternal epigenotype at *Gtl2* as a consequence of the deletion.

Dlk1 and *Gtl2* are members of a larger cluster of imprinted genes (Fig. 1a). Quantitative northern-blot data (Fig. 2) show that all the genes normally expressed from the maternally inherited chromosome are repressed after maternal inheritance of the deletion. In contrast, there were no significant alterations in the imprinting of any of the genes in the cluster after paternal transmission of the deleted (methylated) IG-DMR (Fig. 2). Therefore, the IG-DMR regulates imprinting of a 1-Mb cluster of genes bidirectionally on the maternal chromosome only.

After maternal inheritance, all the C/D snoRNAs that we tested (MBII-48, MBII-49, MBII-78) as well as miR-127 were expressed at 0–8% of wild-type levels (Fig. 2). *Dio3* expression was double that

seen in wild-type littermates, owing to activation of the normally repressed maternal allele (Fig. 3c). The absence of differential methylation of CpG islands at and around *Dio3* (ref. 8 and data not shown) is consistent with its imprinting being under the control of more distant elements. The usually silent copy of *Rtl1* was activated, and notably, this deregulation resulted in mean expression (450%) that was substantially higher than expected for a double dose (Fig. 2b). This suggests that in the mutant, the gene either is transcribed at higher than normal levels or is abnormally stabilized post-transcriptionally. This finding is notable as it suggests a function for the maternally expressed miRNAs that are no longer transcribed in the maternally inherited knockout.

Because of their 100% complementarity with *Rtl1*, we predicted that miR-127 and miR-136 might act as small interfering RNAs (siRNAs) to trigger cleavage of *Rtl1* transcripts¹². Semi-quantitative, strand-specific, allele-specific analysis indicated that *Rtl1* on the maternal allele was activated to ~80% of the level observed on the paternal allele (Fig. 3b). Although a role for miRNAs in *Rtl1* imprinting cannot be ruled out¹², relatively similar upregulation

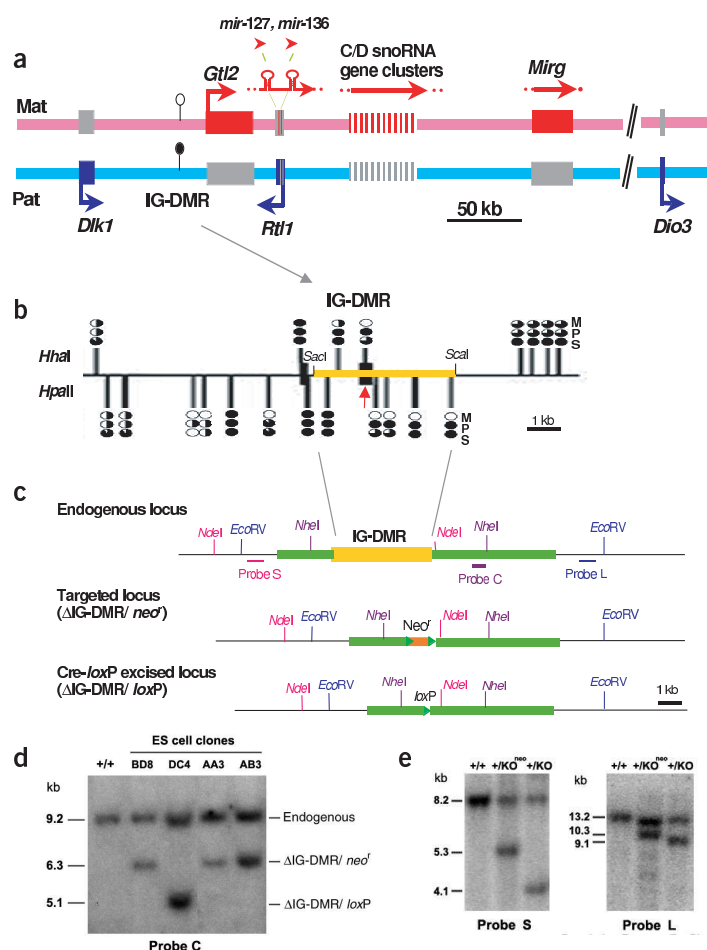


Figure 1 Targeted deletion of the IG-DMR upstream of *Gtl2* on mouse chromosome 12. (a) Schematic representation of the 1-Mb imprinted cluster on mouse distal chromosome 12. Differential methylation at the IG-DMR is shown with filled circles representing methylated alleles and open circles representing unmethylated alleles. (b) The deleted IG-DMR region is shown in yellow. Vertical bars indicate *HhaI* and *HpaII* sites. The methylation status (as described in ref. 5) is illustrated with open and filled circles representing lack of methylation and full methylation, respectively. Partially filled circles represent partial methylation. M, P and S represent the methylation status in mUPD12, pUPD12 and sperm DNA, respectively. The location of the conserved tandem repeat array is indicated by a red arrow. The heavy blocks represent CpG islands. (c) The targeting construct. Dark green bars represent the extent of short- and long-arm homology, and the yellow bar represents the deleted region. (d) Southern-blot hybridization analysis of digested DNA isolated from representative targeted ES cell clones before and after Cre-loxP-mediated excision of *neo*^r. The restriction enzyme (*NheI*) sites and probe C are shown in c. (e) Southern-blot hybridization using endogenous probes outside the regions of the short-arm (probe S) and long-arm (probe L) homologies indicates correct targeting. DNA isolated from mice with the paternally inherited IG-DMR deletion with (+/KO^{neo}) or without (+/KO) *neo*^r and from their wild-type littermates was cut with *NdeI* (for short arm) or *EcoRV* (for long arm).

LETTERS

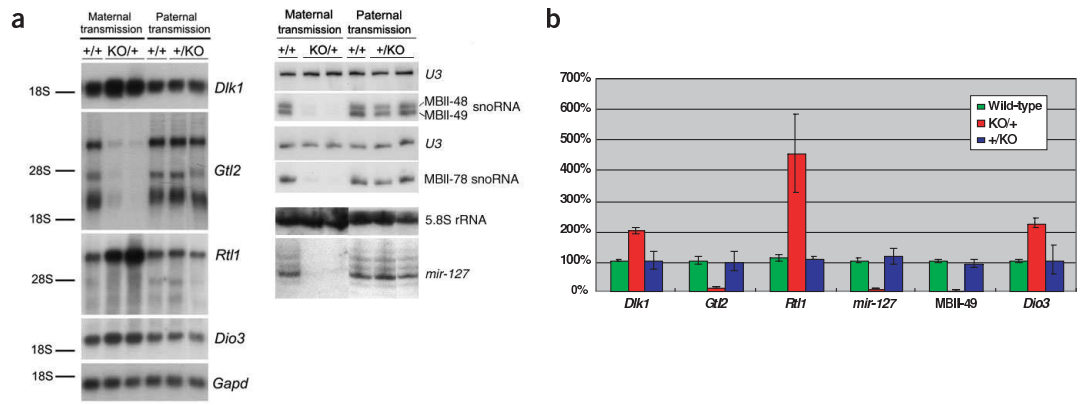


Figure 2 Expression of imprinted genes as a consequence of the IG-DMR deletion. (a) Northern-blot analysis of gene expression at the imprinted domain. *Gapd*, *U3* and 5.8S rRNA were used as unlinked control genes for quantification. (b) Graphical representation of expression comparing E16 embryos carrying the maternally derived IG-DMR deletion (red; KO/+) with those with the paternally derived deletion (blue; +/-KO) and wild-type littermates (green; +/+). Error bars represent s.e.m. Values were calculated from data generated using control and mutant embryos from multiple litters.

from the two chromosomes in the knockout suggests that such post-transcriptional regulation cannot discriminate between *Rtl1* transcripts from either allele. Functional interaction between maternally and paternally expressed transcripts at the orthologous imprinted domain in sheep has been proposed to explain an unusual pattern of inheritance associated with the callipyge (*CLPG*) phenotype¹⁵. The callipyge defect increases expression on both chromosomes but does not affect imprinting¹⁶. The mutant

phenotype (postnatal muscle hyperplasia) is observed in heterozygotes after paternal transmission only, whereas homozygotes with respect to the *CLPG* allele are normal. The normal phenotype in *CLPG/CLPG* sheep suggests that overexpression of a gene (or genes) on the maternal chromosome abrogates the effect of overexpression of a paternally expressed gene (or genes). Interaction between maternally expressed miRNAs and paternally expressed *Rtl1* transcripts is consistent with this model.

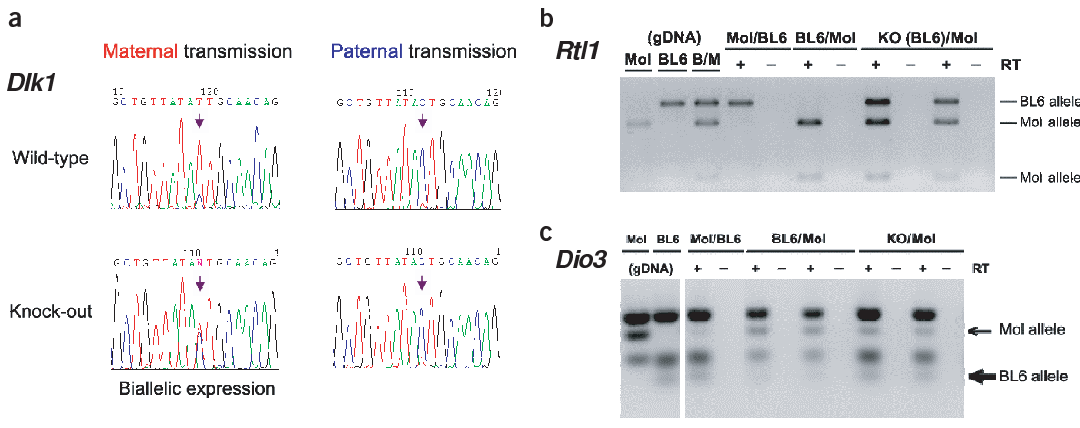


Figure 3 Biallelic expression of *Dlk1*, *Rtl1* and *Dio3* after maternal transmission of IG-DMR deletion. (a) *Dlk1*. Sequence analysis of RT-PCR products from control and heterozygous embryos after maternal and paternal transmission of the IG-DMR deletion (knock-out). Biallelic expression of *Dlk1* was detected only when the IG-DMR deletion was maternally transmitted. (b) *Rtl1*. Strand-specific RT-PCR followed by digestion distinguished the *Rtl1* transcript in *M. musculus molossinus* (Mol) and C57BL/6 (BL6) F₁ hybrid mice. Control hybrids always expressed the paternal allele. Biallelic expression was evident in mice with the maternally inherited deletion (KO (BL6)/Mol). Semi-quantitative analysis indicated that maternal transcript levels were 77.04 ± 0.63% that of the unmanipulated paternal allele (control BL6/Mol hybrid genomic DNA amplification resulted in maternal allelic amplification of 100.62 ± 0.54% compared to paternal). Control RT-PCR reactions lacking reverse transcriptase (RT) are indicated (-). (c) *Dio3*. RT-PCR followed by *TaqI* digestion distinguished *Dio3* transcripts expressed from Mol and BL6 alleles. Control hybrids predominantly expressed the paternal allele, whereas mice with the maternally inherited deletion showed activation of the maternal allele, as evident from the greater intensity of the diagnostic BL6 band (thick arrow) relative to the diagnostic Mol band (thin arrow). Control RT-PCR reactions lacking reverse transcriptase (RT) are indicated (-).

LETTERS

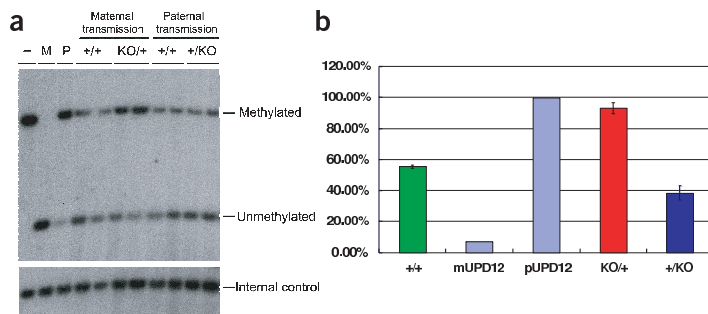


Figure 4 Promoter methylation at *Gtl2* is altered in the absence of the IG-DMR. **(a)** Representative methylation-sensitive Southern-blot hybridization showing the 3-kb methylated band and 1.2-kb unmethylated product in mice that are heterozygous with respect to the IG-DMR deletion after maternal (KO/+) and paternal (+/KO) transmission. Data is shown for *PvuII* (-) and *PvuII* with *SmaI*. M, mUPD12; P, pUPD12. **(b)** Graphical representation of quantitative methylation changes associated with the deletion. Data was generated from embryonic DNA digested with *PvuII* or *HincII* and *HpaII*. The measurement illustrates the percentage of full methylation relative to that observed in pUPD12 mice (normalized at 100% methylation).

Deletion of the IG-DMR on the maternal chromosome results in silencing of the complex array of maternally active non-coding RNAs, including *Gtl2*, which normally expresses multiple alternatively spliced transcripts. *Gtl2* also has little sequence conservation downstream of the first two exons¹³. All maternally expressed imprinted genes are downstream of the IG-DMR, are transcribed in the same orientation and, with the exception of *Gtl2*, have uncharacterized transcriptional start sites¹⁷. Whether the *Gtl2* promoter has a role in the regulation of these downstream RNAs has not yet been determined.

The IG-DMR is believed to be the only germ line-derived DMR in the mouse 125-kb locus that contains both *Dlk1* and *Gtl2* (ref. 5). Because *DLK1* and *GTL2* are imprinted in humans^{10,18} and the region functions in growth control and development, methylation analysis in humans has diagnostic merit, particularly when considering loss of imprinting in human tumors. Using DNA from cells derived from maternal and paternal uniparental disomy for chromosome 14 (UPD14; refs. 19,20), we found that the orthologous IG-DMR region on human chromosome 14 is also differentially methylated on the two parental alleles. The paternally inherited allele is hypermethylated in pUPD14 and hypomethylated in mUPD14 cells relative to wild-type controls (Fig. 5).

Three models of postfertilization regulation of imprinted gene clusters have been suggested¹. First, the insulator model was established from studying the reciprocally imprinted genes *Igf2* and *H19*. In this example, an intergenic germline DMR binds the insulator protein CTCF in a methylation-sensitive manner, blocking access of *Igf2* to shared downstream enhancers on the unmethylated maternal chromosome. The methylated DMR on the paternal chromosome seems to be required for maintaining the reciprocal imprinting of the two genes on that chromosome^{21,22}. Second, in the bipartite imprinting center model as described for the PWS-AS (Prader-Willi and Angelman syndromes) locus on human chromosome 15, two distinct regions (imprinting centers; ICs) that function bidirectionally have been mapped, one on each chromosome. On the maternal chromosome, the AS-IC controls methylation and repression of multiple genes, and on the paternal chromosome, the PWS-IC is required for activity and hypomethylation. The PWS-IC is a DMR, and the AS-IC has not yet been characterized at the molecular level *in vivo*^{1,23}. Third, the antisense transcript mechanism is exemplified by the *Igf2r* domain, where a large non-coding RNA (*Air*) is required for bidirectional repression of a cluster of paternal alleles (including *Igf2r*) on mouse chromosome 17. *Air* is transcribed from an intronic germline DMR in *Igf2r* on the unmethylated paternally inherited copy²⁴. A second imprinted domain on mouse distal chromosome 7, including the

genes *Cdkn1c* and *Kcnq1*, is also regulated bidirectionally by a paternally unmethylated germline DMR, and evidence to date suggests regulation by a mechanism also involving antisense RNA^{25,26}.

Several models could explain the function of the IG-DMR on the maternal chromosome. First, the region might be a transcriptional activator of *Gtl2* and associated transcripts. These non-coding transcripts might then function as *cis*-acting negative regulators of the maternally silent imprinted genes *Dlk1*, *Rtl1* and *Dio3* in a manner reminiscent of the function of *Air* at *Igf2r*²⁴. In a second model, the unmethylated maternal IG-DMR might act directly as a long-range regulator of the whole domain, conferring a particular epigenetic state and chromatin conformation associated with activity and repression over at least a 1-Mb domain. The absence of a significant effect of the deletion on the methylated paternal chromosome indicates that repression of the non-coding RNAs and activity of *Dlk1*, *Rtl1* and *Dio3* are regulated differently on that chromosome.

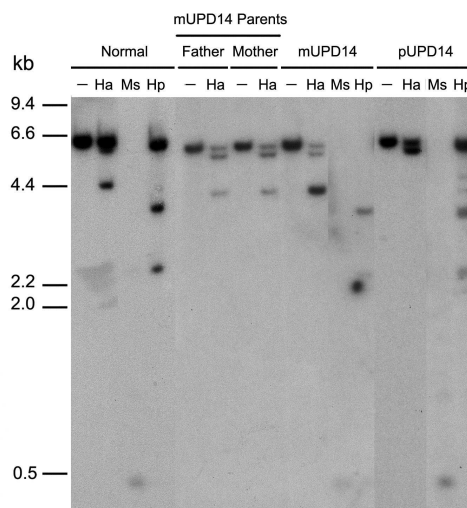


Figure 5 Differential methylation at the IG-DMR is conserved in mouse and human. Comparison of *HpaII* and *HincII* tracks between human mUPD14 and pUPD14 shows that the IG-DMR is hypomethylated on the maternal chromosome and hypermethylated on the paternal chromosome. Ha, *HpaII*; Hp, *HincII*; Ms, *MspI*.

LETTERS

Although we cannot rule out an insulator function at the IG-DMR, it seems different from the *Igf2-H19* domain, where an effect is observed after DMR deletion from each of the parental germ lines²². Rather, our results suggest a model of imprinting control similar to that described at the PWS-AS locus, in which two different regions regulate imprinting on the two parental chromosomes^{1,23}. Alternatively, the methylation and expression status of the paternal chromosome (Fig. 6) might represent an 'ancestral' or 'default' state that was present before the evolution of imprinting on the maternal chromosome at this domain. None of these models is mutually exclusive, and regardless of which model (models) best describes it, imprinting of genes in the 1-Mb domain on chromosome 12 is controlled differently on the two parental chromosomes.

Maternal deletion of the IG-DMR switched the epigenotype and transcriptional behavior of the maternal chromosome to resemble that of the paternal chromosome (Fig. 6). This causes late-gestational lethality. Because multiple imprinted genes are affected, lethality cannot be attributed to any single gene in the cluster. Neither targeted deletion of maternally expressed transcripts nor overexpression of individual paternally expressed genes has been described for this locus. But paternal uniparental disomy 12 (pUPD12) embryos also show fetal lethality, and developmental abnormalities have been described^{14,27}. Comparative phenotypic analysis of pUPD12 embryos with those carrying maternally inherited deletion of the IG-DMR could determine whether the defects associated with pUPD12 can be attributed to the 1-Mb imprinted domain on distal chromosome 12.

METHODS

Targeted deletion of the IG-DMR. We constructed the targeting vector for deleting the IG-DMR (nucleotides 79,883–84,039) using a 2-kb *SacI* fragment (nucleotides 77,841–79,883) and a 4.9-kb *SacI* to *SphI* fragment (nucleotides 84,039–88,936) ligated into the pNeoFlo8 vector (provided by U. Lichtenberg, Cologne) upstream and downstream, respectively, of the neomycin resistance (*neo^r*) gene²⁸. We linearized the targeting vector, electroporated it into the male 129/OlaHsd ES cell line E14Ola²⁹ and selected ES cell clones containing the targeting vector with G418. We confirmed the short-arm and long-arm homologous recombination events of putative targeted ES cell clones by Southern-blot hybridization. We excised *neo^r* in the targeted clones *in vitro* by transient expression of an electroporated Cre recombinase gene (pTurboCre, provided by T. Ley, Washington University, St. Louis, USA). Here, we presented only data from mice lacking *neo^r*, but mice with *neo^r* gave the same results, suggesting that *neo^r* had no effect *in vivo*. The Gene Targeting Facility at The Babraham Research Institute carried out ES cell targeting and blastocyst injection. Mice were maintained on a C57BL6 genetic background unless otherwise indicated. Experiments involving mice were carried out in accordance with the UK Government Home Office licensing procedures.

Expression analysis of imprinted genes in the *Dlk1-Gtl2* imprinted cluster. We isolated total and poly(A)⁺ RNA from embryos with the IG-DMR deleted and control littermates as described previously¹³. We carried out northern-blot hybridization and used probes as described previously for

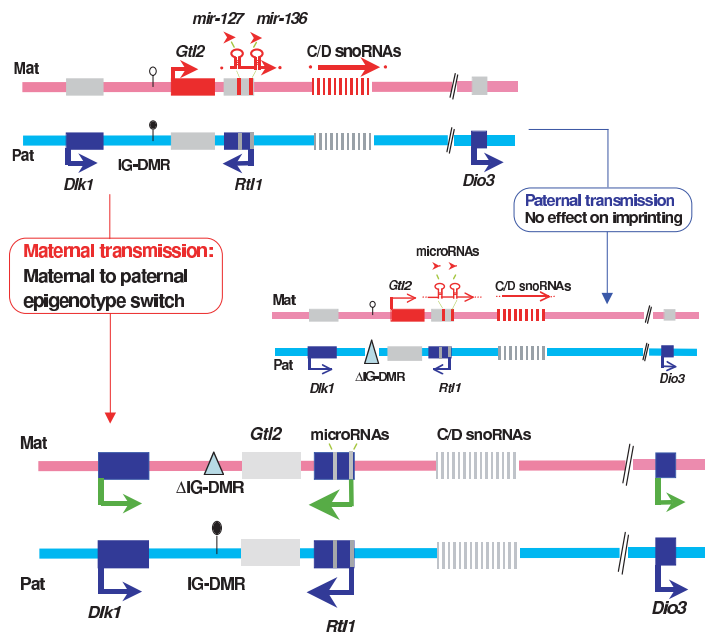


Figure 6 Summary of imprinted gene expression after maternal and paternal transmission of the IG-DMR deletion. When maternally transmitted, deletion of the unmethylated IG-DMR caused bidirectional loss of imprinting of all genes in the domain. No significant differences in imprinting were observed when deletion was paternally transmitted. Schematic is not drawn to scale.

Dlk1 and *Gtl2* (ref. 6), snoRNAs¹¹, *Rtl1* and microRNAs¹², and *Dio3* (ref. 8). We quantified expression using a Storm 860 Phosphorimager (Amersham) and associated ImageQuant software according to manufacturer's instructions. We showed biallelic expression of *Dlk1* from E16 embryos heterozygous with respect to a maternally inherited deletion after we identified a single-nucleotide polymorphism between C57BL6 (also 129/Sv, CBA, Balb/C, C3H, SWR, FVB, AKR) and DBA/2 (G→A substitution). We carried out reciprocal crosses between DBA/2 and knockout stock to generate embryos heterozygous with respect to the maternally or paternally inherited deletion and wild-type littermates and then amplified by RT-PCR and sequenced a 223-bp fragment of RNA isolated from these embryos. We sequenced multiple individual PCR products. We evaluated *Dio3* allelic activity using a *Mus musculus molossinus* restriction-site polymorphism as described previously⁸ and *Rtl1* allelic expression with strand-specific RT-PCR using a *M. musculus molossinus* restriction-site polymorphism (*StuI*). We reverse-transcribed RNA using a strand-specific primer and Superscript III (Invitrogen) according to manufacturer's instructions at 55 °C for 1 h. We carried out subsequent PCR reactions in duplicate over multiple cycles to generate a standard curve. We used hybrid genomic DNA as a control to check for equivalence in allelic amplification efficiency and complete digestion. Amplification reactions proceeded in an exponential manner with equivalent slope up to 40–45 cycles of PCR (data not shown). Therefore, we carried out 38 cycles of amplification before digesting samples with *StuI* and analyzing them by densitometry using Molecular Dynamics ImageQuant software (Amersham). Values were calculated according to their respective molar ratios.

Methylation analysis of the *Gtl2* DMR in IG-DMR knockout embryos. We isolated genomic DNA from embryos using standard techniques⁵. We carried out methylation-sensitive Southern-blot analysis of the *Gtl2* DMR using a 450-bp *KpnI-SmaI* fragment as probe (nucleotides 93,667–94,100) for *SmaI* filters

LETTERS

and probe G1 for *HpaII* filters⁶. We used a fully methylated region in *Rtl1* as an internal control. To make the filters for quantitative methylation analysis, we digested genomic DNA samples with *PvuII* and *SmaI* or *HpaII* or with *HincII* with *HpaII*. A total of ten *HpaII* sites were analyzed. We quantified the intensity of the undigested product relative to the internal control to provide a 'fully methylated' value and plotted the intensity relative to the level observed in mUPD12 or pUPD12 DNA in which the parental chromosomes are hypo- or hypermethylated, respectively. We quantified intensities of diagnostic and control bands on a Storm 860 phosphorimager using Amersham software. Methylation levels are given relative to pUPD12 mice, represented as 100% methylation. We generated mice with maternal and paternal uniparental disomy for chromosome 12 as described previously¹⁴. We confirmed complete digestion of the DNA using a probe for exon 1 of *Dlk1* that is fully unmethylated on both alleles⁶ (data not shown).

Methylation analysis of the IG-DMR orthologous region on human chromosome 14. We obtained lymphocytes derived from an individual with mUPD14 (DD0217), described in ref. 19, and the parents (DD0216, DD0218) from the European Tissue Culture Collection. We obtained fibroblasts and lymphocytes from individual GR, described in ref. 20, from J. Wang (Alfgen, California). We isolated DNA from primary cultures and digested it with *BamHI* and methylation-sensitive restriction enzymes. We generated the 654-bp probe corresponding to the human IG-DMR by PCR amplification of human genomic DNA. We carried out Southern-blot hybridization as described previously⁶. Primer sequences and conditions for their use are available on request.

GenBank accession number. Nucleotide numbers refer to sequence accession number AJ320506 for the *Dlk1-Gli2* cluster.

ACKNOWLEDGMENTS

We thank D. Barlow, G. Kelsey and members of the laboratory of A.C.F.-S. for discussions of this work; P.-W. Teng and A. Woodhouse for technical assistance; M. George and E. Saunders of the Babraham Gene Targeting Facility for ES cell culture and blastocyst injection; R. Voutilainen for assistance with the cell cultures; and M. Constanica for advice with the targeting construct. This work was supported by grants from the UK Medical Research Council and Cancer Research UK. S.-P.L. is funded by a graduate research scholarship from the Taiwanese Government, N.Y. by a BBSRC studentship and H.S. by a PhD fellowship from the Allocation de Moniteur Normalien.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 17 April; accepted 4 August 2003

Published online at <http://www.nature.com/naturegenetics/>

1. Ferguson-Smith, A.C. & Surani, M.A. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293**, 1086–1089 (2001).
2. Bird, A. DNA methylation patterns and epigenetic memory. *Genes. Dev.* **16**, 6–21 (2002).
3. Bourchis, D., Xu, G.L., Lin C.S., Bollman, B. & Bestor, T.H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
4. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983–1993 (2002).
5. Takada, S. *et al.* Epigenetic analysis of the *Dlk1-Gli2* imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with *Igf2-H19*. *Hum. Mol. Genet.* **11**, 77–86 (2002).

6. Takada, S. *et al.* Delta-like and *Gli2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10**, 1135–1138 (2000).
7. Schmidt, J.V., Matteson, P.G., Jones, B.K., Guan, X.J. & Tilghman, S.M. The *Dlk1* and *Gli2* genes are linked and reciprocally imprinted. *Genes. Dev.* **14**, 1997–2002 (2000).
8. Tsai, C.-E. *et al.* Genomic imprinting contributes to thyroid hormone metabolism in the mouse embryo. *Curr. Biol.* **12**, 1221–1226 (2002).
9. Yevdodiyenko, A., Carr, M.S., Patel N. & Schmidt, J. Analysis of candidate imprinted genes linked to *Dlk1-Gli2* using a congenic mouse line. *Mamm. Genome* **13**, 633–638 (2002).
10. Miyoshi, N. *et al.* Identification of an imprinted gene, *Meg3/Gli2* and its human homologue *MEG3*, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes. Cells* **5**, 211–220 (2000).
11. Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C. & Bachevalier, J.P. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* **11**, 1527–1538 (2002).
12. Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* **34**, 261–262 (2003).
13. Paulsen, M. *et al.* Comparative sequence analysis of the imprinted *Dlk1-Gli2* locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the *Igf2-H19* region. *Genome Res.* **11**, 2085–2094 (2001).
14. Georgiades, P., Watkins, M., Surani, M.A. & Ferguson-Smith, A.C. Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* **127**, 4719–4728 (2000).
15. Georges, M., Charlier, C. & Cockett, N. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet.* **19**, 248–252 (2003).
16. Charlier, C. *et al.* The callipyge mutation enhances the expression of coregulated imprinted genes in *cis* without affecting their imprinting status. *Nat. Genet.* **27**, 367–369 (2001).
17. Schuster-Gossler, K., Bilinski, P., Sado, T., Ferguson-Smith, A. & Gossler, A. The mouse *Gli2* gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts and may act as an RNA. *Dev. Dyn.* **212**, 214–228 (1998).
18. Wyllie, A.A., Murphy, S.K., Orton, T.C. & Jirtle, R.L. Novel imprinted *DLK1/GTL2* domain on human chromosome 14 contains motifs that mimic those implicated in *Igf2/H19* regulation. *Genome Res.* **10**, 1711–1718 (2000).
19. Temple, I.K., Cockwell, A., Hassold, T., Pettay, D. & Jacobs, P. Maternal uniparental disomy for chromosome 14. *J. Med. Genet.* **28**, 511–514 (1991).
20. Wang, J., Passage, M., Yen, P., Shapiro, L. & Mohandas, T. Uniparental heterodisomy for chromosome 14 in a phenotypically abnormal familial balanced 13/14 Robertsonian translocation carrier. *Am. J. Hum. Genet.* **48**, 1069–1074 (1991).
21. Arney, K. *H19* and *Igf2*—enhancing the confusion? *Trends Genet.* **19**, 17–23 (2003).
22. Thorvaldsen, J.L., Duranm K.L. & Bartolomei, M.S. Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes. Dev.* **12**, 3693–3702 (1998).
23. Buiting, K. *et al.* Epimutation in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. *Am. J. Hum. Genet.* **72**, 571–577 (2003).
24. Sleutels, F., Zwart, R. & Barlow, D. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
25. Fitzpatrick, G., Soloway, P.D. & Higgins, M.J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of *KvDMR1*. *Nat. Genet.* **32**, 426–431 (2002).
26. Mancini-DiNardo, D., Steele, S.J., Ingram, R.S. & Tilghman, S.M. A differentially methylated region within the gene *Kcnq1* functions as an imprinted promoter and silencer. *Hum. Mol. Genet.* **12**, 283–294 (2003).
27. Georgiades, P., Watkins, M., Burton, G.J. & Ferguson-Smith, A.C. Roles for genomic imprinting and the zygotic genome in placental development. *Proc. Natl. Acad. Sci. USA* **98**, 4522–4527 (2001).
28. Constanica, M. *et al.* Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat. Genet.* **26**, 203–206 (2000).
29. Mundt, C., Licence, S., Shimizu, T., Melchers, F. & Martenson, I.L. Loss of precursor B cell expansion but not allelic exclusion in VpreB1/VpreB2 double-deficient mice. *J. Exp. Med.* **193**, 435–445 (2001).

2.3 Une concentration de gènes de microARN

L'article n°2, « Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene » (Seitz *et al.*, 2003), présente les gènes de quatre microARN murins dans le locus distal 12 (ces quatre microARN avaient été clonés par une autre équipe), et suggère que les répétitions du « cluster R » décrit dans l'article n°1 (Cavaillé *et al.*, 2002) constituent autant de nouveaux gènes de microARN. En effet, la transcription de toutes ces séquences produirait des ARN capables de se replier en tiges-boucles imparfaites de 80 à 100 nt, et la plupart (les membres de la famille A) sont apparentées au gène de l'un des microARN avérés, miR-154. Enfin, il est connu que les gènes de microARN peuvent être regroupés (Lagos-Quintana *et al.*, 2001 ; Lau *et al.*, 2001 ; Mourelatos *et al.*, 2002 ; Reinhart et Bartel, 2002 ; Dostie *et al.*, 2003 ; Lagos-Quintana *et al.*, 2003).

Toutes ces observations nous ont incités à rechercher de nouveaux gènes de microARN dans le locus 14q32 (*H. sapiens*) / distal 12 (*M. musculus*). Cette recherche a commencé par la prédiction des gènes de microARN dans le locus humain, par la recherche systématique de séquences présentant les caractéristiques connues des gènes de microARN ; ensuite ces candidats (tous conservés chez la Souris, puisque leur conservation entre Homme et Souris faisait partie des propriétés que nous imposions dans la sélection des candidats) ont été vérifiés expérimentalement chez la Souris, par Northern blot (pour quelques-uns) et par extension d'amorce (pour chacun des candidats testés). La similarité, en séquence, de nombreux candidats ne nous a pas permis d'analyser l'expression de chacun sans ambiguïté. Parmi les 40 candidats identifiés, nous avons donc testé l'expression de 16 (les candidats appelés A22, A24, B3, B4, B6, C2, C3, D, G, H, I, K, L, M, N et Q), et parmi ces 16 candidats, 10 se sont avérés être d'authentiques gènes de microARN, donnant lieu à l'expression d'ARN détectables par Northern blot ou extension d'amorce (ce sont les candidats appelés A22, A24, B4, B6, C2, C3, D, K, M et N). Cette dernière technique nous est apparue plus sensible que le Northern blot (de nombreux ARN détectés par extension d'amorce ne l'étaient pas par Northern blot). Si la plupart de ces gènes de microARN sont localisés dans le cluster R, quelques-uns sont en-dehors (parmi les gènes d'ARN C/D répétés, ou encore en amont).

Parmi les 6 candidats testés pour lesquels nous n'avons pas détecté de microARN, se trouvent probablement des faux positifs (*i.e.* : des séquences transcrites en ARN qui se replient en tige-boucle, mais qui ne sont pas maturées en microARN) ; il est également possible que certains soient des gènes de microARN dont le produit est trop peu abondant, mais que des travaux futurs (par exemple, des travaux de clonage de microARN) mettent malgré tout en évidence des microARN exprimés à partir de ces candidats.

Tous les nouveaux gènes de microARN analysés sont exprimés à partir de l'allèle maternel uniquement, sont contrôlés par la région différenciellement méthylée étudiée dans l'article n°3 (Lin *et al.*, 2003), et leur expression est régulée au cours du développement, chez la Souris : exprimés principalement dans le cerveau adulte, ils sont toutefois exprimés à la fois dans la tête et le tronc de l'embryon. L'expression dans le tronc persiste encore au moins une semaine après la naissance : il est donc possible qu'un organe situé dans le tronc, donc distinct du cerveau, exprime ces microARN chez l'adulte. La plupart des gènes de ces microARN sont introniques, et plusieurs sont cotranscrits, sous la forme de précurseurs communs de plus grande taille : il se pourrait que tous ces microARN, et éventuellement les ARN C/D (voir article n°1) et l'ARN non-codant Glt2 avec eux, soient transcrits sous la forme d'un long précurseur commun, dont l'épissage libérerait à la fois un ARN non-codant, issu du raboutage des exons, et des précurseurs d'ARN C/D et de microARN. Cette hypothèse est discutée page 159.

Ces microARN n'ont pas d'ARN-cible évident ; cependant, un modèle a été proposé pour justifier des observations étranges, chez le Mouton, et qui suppose qu'un (ou : des) gène(s) à

expression maternelle de l'orthologue ovin de ce locus réprime(nt) post-transcriptionnellement un (ou : des) gène(s) à expression paternelle du même locus (Georges *et al.*, 2003). Les microARN que nous avons décrits sont des candidats naturels pour jouer ce rôle de répresseurs à expression maternelle, puisque les rôles démontrés des microARN sont des répressions post-transcriptionnelles (par dégradation de l'ARN-cible, ou inhibition de sa traduction si c'est un ARN_m ; voir Introduction, page 68), et que tous ces nouveaux microARN semblent exprimés à partir de l'allèle maternel uniquement.

Ces travaux sont exposés dans l'article « A large imprinted microRNA gene cluster at the mouse *Dlk1-Gtl2* domain » (Seitz *et al.*, 2004), présenté ci-après.

2.3.1 Article n°4

Letter

A Large Imprinted microRNA Gene Cluster at the Mouse Dlk1-Gtl2 Domain

Hervé Seitz,¹ Hélène Royo,¹ Marie-Line Bortolin,¹ Shau-Ping Lin,²
Anne C. Ferguson-Smith,² and Jérôme Cavaillé^{1,3}

¹LBME-CNRS (UMR 5099), IFR-109, Université P. Sabatier, 31062 Toulouse Cedex, France; ²Department of Anatomy, University of Cambridge, Cambridge CB2 3DY, United Kingdom

microRNAs (or miRNAs) are small noncoding RNAs (21 to 25 nucleotides) that are processed from longer hairpin RNA precursors and are believed to be involved in a wide range of developmental and cellular processes, by either repressing translation or triggering mRNA degradation (RNA interference). By using a computer-assisted approach, we have identified 46 potential miRNA genes located in the human imprinted 14q32 domain, 40 of which are organized as a large cluster. Although some of these clustered miRNA genes appear to be encoded by a single-copy DNA sequence, most of them are arranged in tandem arrays of closely related sequences. In the mouse, this miRNA gene cluster is conserved at the homologous distal 12 region. In vivo all the miRNAs that we have detected are expressed in the developing embryo (both in the head and in the trunk) and in the placenta, whereas in the adult their expression is mainly restricted to the brain. We also show that the miRNA genes are only expressed from the maternally inherited chromosome and that their imprinted expression is regulated by an intergenic germline-derived differentially methylated region (IG-DMR) located ~200 kb upstream from the miRNA cluster. The functions of these miRNAs, which seem only conserved in mammals, are discussed both in terms of epigenetic control and gene regulation during development.

[Supplemental material is available online at www.genome.org.]

Over 200 distinct genes encoding 21- to 25-nucleotide (nt)-long noncoding RNAs, the microRNAs (miRNAs), have been identified through either computer-assisted approaches or cDNA cloning strategies in many organisms, including worm, plants, flies, mouse, and human (Lai 2003; Bartel 2004). They are encoded within irregular hairpins and generated by a two-step mechanism involving Drosha and Dicer, two distinct dsRNA-specific ribonucleases belonging to the RNase III family. First, Drosha cleaves a larger RNA precursor (the pri-miRNA) in the nucleus to release an irregular ~70-nt-long hairpin (the pre-miRNA; Lee et al. 2003) from which the miRNA is excised from one strand within the cytoplasm through the activity of Dicer (Bernstein et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001). Then, the mature 21- to 23-nt-long miRNAs incorporate into the RISC/miRNP complexes containing proteins belonging to the Argonaute family (Schwarz and Zamore 2002) and base-pair with mRNAs to trigger either RNA degradation (RNA interference or RNAi) or translation repression (Hutvagner and Zamore 2002; Llave et al. 2002). Indeed, miRNAs perfectly complementary to the RNA target guide sequence-specific RNA cleavages (they function like small interfering RNAs or siRNAs), whereas miRNAs imperfectly complementary to the target repress its translation by an as-yet-unknown mechanism.

In plants, miRNAs exhibit perfect or near-perfect complementarities with their targets, and they can either guide mRNA degradation (Llave et al. 2002; Palatnik et al. 2003; Tang et al. 2003; Xie et al. 2003) or inhibit their translation (Aukerman and Sakai 2003; Chen 2004). Remarkably, most of predicted target mRNAs encode for transcription factors known to regulate leaf

and flower development, embryonic patterning, and timing transition to flowering (Rhoades et al. 2002). By contrast, animal miRNAs seem to act predominantly as translational repressors, and they have been shown to play a key role in the establishment of temporal and spatial gene expression pattern (Lee et al. 1993; Reinhart et al. 2000; Johnston and Hobert 2003), in the regulation of tissue growth, apoptosis, and fat metabolism (Brennecke et al. 2003; Xu et al. 2003) and in the differentiation of hematopoietic lineages (Chen et al. 2004). However, although many computationally predicted mRNA targets have been recently reported in flies and in mammals (Enright et al. 2003; Lewis et al. 2003; Stark et al. 2003), only a very few miRNA targets have been experimentally tested in vivo, and other RNA-mediated gene silencing mechanisms can also be envisaged. In this regard, RNAi-related mechanisms are also required for the silencing of heterochromatin in *Schizosaccharomyces pombe* (Hall et al. 2002; Reinhart and Bartel 2002; Volpe et al. 2002; Schramke and Allshire 2003; Verdell et al. 2004), for the programmed chromosomal rearrangements in *Tetrahymena* (Yao et al. 2003) and for the transcriptional gene silencing in plants and in *Drosophila* (Mette et al. 2000; Hamilton et al. 2002; Pal-Bhadra et al. 2002, 2004; Zilberman et al. 2003). Thus, 21- to 25-nt-long RNAs are involved in a range of eukaryotic gene regulation phenomena, and their potential as modulators of gene expression is just beginning to be appreciated.

Genomic imprinting refers to an epigenetic phenomenon by which a subset of mammalian autosomal genes, the so-called imprinted genes, are expressed in a parent-of-origin specific manner (only one of the two parental alleles is expressed). Most of these imprinted genes are clustered within large chromosomal domains harboring both paternally and maternally expressed imprinted genes. Molecular mechanisms underlying genomic imprinting are not fully understood but include allele-specific DNA methylation and core histone modifications, *cis*-acting elements

³Corresponding author.

E-MAIL cavaill@ibcg.biotoul.fr; FAX 33-5-61-33-58-86.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2743304>. Article published online before print in August 2004.

Seitz et al.

functioning as imprinting control centers, chromatin boundaries or insulators, and antisense noncoding RNAs (Verona et al. 2003). We have previously shown that the imprinted mouse distal 12 chromosome (human 14q32) encodes four miRNA genes—*mir-134*, *mir-154*, *mir-136*, and *mir-127*—of which the two latter are only expressed from the maternal allele and are transcribed in an antisense orientation to a retrotransposon-like gene (*Rtl1*; Fig. 1). *Rtl1* is transcribed from the paternal chromosome only, raising the possibility that the RNAi machinery might negatively regulate the *Rtl1* gene and/or contribute to its imprinting status (Lin et al. 2003; Seitz et al. 2003).

In the present work, through a large and systematic computer-assisted approach within the ~1-Mb human imprinted 14q32 locus/mouse distal 12 domain, we report the identification of 46 potential miRNA genes. Remarkably, most of these genes map within a ~40-kb interval that represents, to our knowledge, the largest miRNA gene cluster described so far. All the miRNA genes tested exhibit a tissue-specific expression pattern, and they are only expressed from the maternally inherited allele with imprinted expression controlled by a differentially methylated region (DMR) located ~200 kb away. Although their functions are still unknown, the unexpected characterization of these novel miRNA genes considerably improves our knowledge of gene organization at this imprinted domain by adding a novel layer of complexity in terms of noncoding RNA genes and their potential involvement in epigenetic controls.

RESULTS

Identification of miRNA Genes at Human 14q32 Domain

The tendency of miRNA genes to occur in clusters (Lagos-Quintana et al. 2001, 2003; Lau et al. 2001; Mourelatos et al. 2002; Reinhart et al. 2002; Dostie et al. 2003) prompted us to search systematically for additional miRNA genes encoded within the whole 1-Mb human imprinted 14q32 domain. This was achieved by identifying conserved sequences between mouse and human that have the potential to fold into imperfect hairpins of similar length to known miRNA genes. The sequences were filtered by using the MiRscan software recently developed by Lim and colleagues (2003), and a total of 35 conserved hairpins were eventually retained (Fig. 2A). Remarkably, most of them (29) are clustered within a 40-kb-long region that maps ~25 kb downstream of the C/D snoRNA gene cluster (Cavaille et al. 2002). The remaining selected hairpins are encoded either around or within the *Rtl1* gene (including the previously reported *mir-127* and *mir-136* [Lagos-Quintana et al. 2002] and three novel ones, *mir-L*, *mir-M*, *mir-N*) or upstream from the *Dio3* gene (*mir-Q*; Figs. 1, 2A, Supplemental data S1).

Multiple sequence alignments revealed that 15 predicted miRNA genes have sequences related to the previously described *mir-154* gene precursor (Lagos-Quintana et al. 2002; Seitz et al. 2003), and a subsequent BLAST search with lower stringency conditions allowed us to identify eight additional human *mir-154*-related sequences (Figs. 2B, 3A), although six do not have clear mouse counterparts. These *mir-154*-related sequences have been grouped into the A-type family (A1 to A24 gene copies). Interestingly, within the same 40-kb interval other hairpins could also be ranked together into families of closely related sequences: the B-type family (six and three gene copies in human and mouse, respectively) and the C-type family (four gene copies both in human and in mouse; Figs. 2B, 3B,C). Thus, overall a total of 40 conserved in silico-predicted human miRNA genes, all of them mapping in the same relative orientation, have been isolated downstream of the C/D snoRNA gene cluster (Fig. 2B), and to our knowledge, they represent the largest cluster of miRNA genes described to date. A-, B-, and C-type pre-miRNA sequences appear to be related to each other (particularly at their terminal 5' and 3' ends; data not shown), and this might suggest that the three miRNA families have evolved through duplications from a common miRNA gene ancestor.

Experimental Detection of the In Silico-Predicted miRNAs

The expression of all the predicted miRNA species was first monitored in mouse by Northern blot analysis with oligonucleotide probes complementary to either the 5' or the 3' arm of the predicted fold-back structures. We failed to detect convincing signals for most of them, including *mir-154* and *mir-134*, which were originally cloned from cDNA libraries (Lagos-Quintana et al. 2002; data not shown). Indeed, only *mir-C3*, *mir-A24*, and B-type miRNAs gave positive signals in total RNA isolated from whole embryo or adult brain (data not shown). Nevertheless, based on our current knowledge of the relative position of the mature miRNA within the pre-miRNA (Lim et al. 2003), we designed a novel specific set of primers for several miRNAs distributed along the cluster (Supplemental data S1) and performed a more sensitive primer extension analysis. By using this approach, we mapped the 5' end of small RNA species encoded within pre-miR-A22 (3' arm), pre-miR-A24 (3' arm), pre-miR-B4 (3' arm), pre-miR-B6 (3' arm), pre-miR-C2 (3' arm), pre-miR-C3 (3' arm), pre-miR-154 gene (5' arm), pre-miR-134 (5' arm), pre-miR-K (3' arm), pre-miR-M (3' probe), and pre-miR-N (3' probe; Table 1; data not shown). Because of potential cross-hybridization between related miRNA sequences, we have been unable to test unambiguously each miRNA gene candidate, hence the possibility that some of the in silico-predicted miRNA variants might not be expressed cannot be formally ruled out. However, three re-

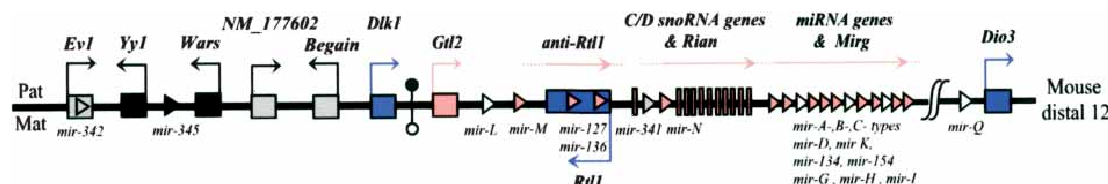


Figure 1 Schematic representation of the imprinted mouse distal 12 domain (human 14q32). The position of several imprinted genes is indicated by squares (*Gtl2* and protein-coding genes), vertical bars (snoRNA genes), or triangles (miRNA genes). Maternally expressed and paternally expressed genes are filled in pink and blue, respectively. Biallelically expressed genes and genes with undetermined imprinted status are colored in black and grey, respectively. Open triangles represent in silico-predicted miRNA gene that we failed to experimentally detect (i.e., pre-miR-L, pre-miR-H, pre-miR-G, pre-miR-I, pre-miR-Q). Two recently described miRNA genes—*mir-342* and *mir-345* (Kim et al. 2004)—map within an intron of *Ev1* and between *Yy1* and *Wars* genes, respectively. Imprinting status of *mir-342* remains unsolved as we failed to detect it in embryos. An intergenic germ-line derived differentially methylated region (IG-DMR) located between *Dkl1* and *Gtl2* genes is represented by circles (filled indicates hypermethylated; open, hypomethylated). Mat indicates maternal chromosome; Pat, paternal chromosome. The figure is not drawn to scale.

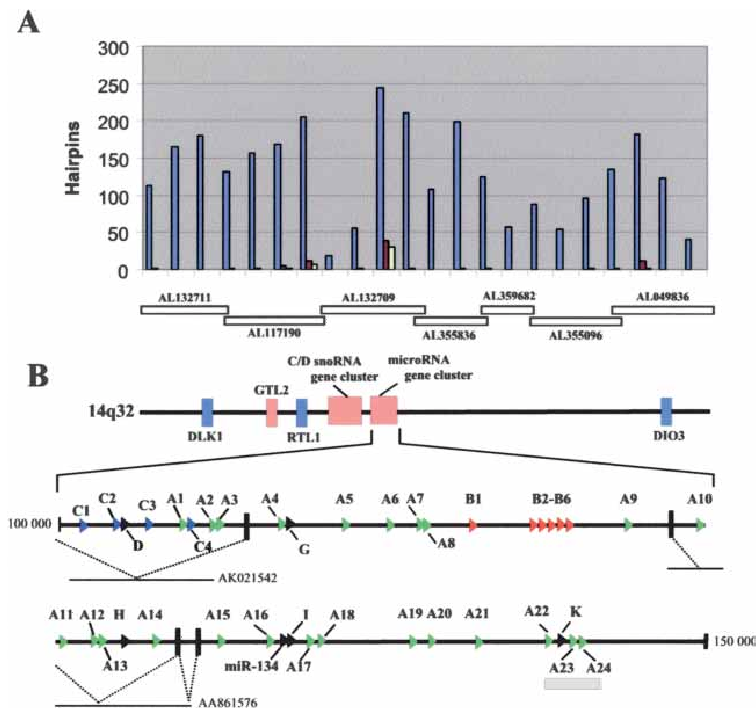


Figure 2 Identification of computationally predicted miRNA genes at the human imprinted 14q32 domain. (A) Distribution of the conserved hairpins between human and mouse along the 1-Mb-long imprinted 14q32 domain. Vertical bars indicate the number of in silico-predicted hairpins found within each 50-kb-long region. Blue bars indicate the number of human hairpins at the 14q32 domain; red bars, the number of conserved hairpins between mouse and human; and yellow bars, the number of conserved hairpins giving rise to a miRscan score >10 (Lim et al. 2003). The relative positions of the overlapping human BACs covering the analyzed domain are indicated below the histogram. (B) A large cluster of miRNA genes mapping downstream from the C/D snoRNA gene cluster. miRNA genes belonging to the A-, B-, and C-types are represented by green, red, and blue arrows, respectively, and miRNAs encoded by a single-copy gene (miR-D, G, H, I, K) are indicated as black arrows. Note that four miRNA genes—miR-A22, miR-K, miR-A23, and miR-A24—are embedded within a conserved CpG island (indicated by a gray rectangle) from which the most conserved sequences lie within the pre-miRNA genes (data not shown). Dotted lines denote splicing events identified by ESTs analysis. Numbering indicates the relative position of the miRNA gene cluster (in nucleotides) within the BAC AL132709 sequence. The picture is drawn to scale.

cently cloned miRNAs—miR-323, miR-329, and miR-300—are encoded within the 3' strand of the A1, A2, and A9 pre-miRNA gene precursors (Houbaviy et al. 2003; Kim et al. 2004), respectively, validating our in silico approach and suggesting that others, if not all the A copies, might also give rise to mature miRNAs. Our failure to experimentally detect miRNAs within miR-D, miR-G, miR-H, miR-I, miR-L, and miR-Q predicted hairpins might reflect expression of these miRNA genes at levels below the classical experimental detection. Consistent with this, miR-300 embedded within the in silico-predicted miR-D hairpin (5' arm) has been cloned, but its transcripts have not been detected by Houbaviy et al. (2003).

Several miRNAs described here have been also detected experimentally in bovine and rat but not in chicken brain (data not shown). BLAST searches also failed to detect significant hits in *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Fugu rubripes* genomes. Taken together, these observations strongly suggest that these miRNA sequences are specific to mammalian species. Based on conservation criteria between human and fish *F. rubripes*, a recent study argued that the number of miRNAs in hu-

man genome should not exceed ~250, with ~40 remaining to be determined (Lim et al. 2003). Our study clearly shows that for mammalian systems, this number may be an underestimate. Reinforcing this notion, poorly conserved embryonic stem-cell-specific miRNA genes, also organized in a tandem array, have been recently described and proposed to play a key role in the regulation of early mammalian development (Houbaviy et al. 2003).

Tissue-specific Expression Pattern of the Distal 12 miRNA Genes

Many miRNA genes are expressed in a temporal and/or tissue-specific manner in line with their well-documented functions in development (Carrington and Ambros 2003). To gain insight into the miRNA gene expression pattern, primer extensions have been performed for several miRNAs by using total RNA extracted from several adult mouse tissues, embryos (head or trunk), and placenta. As shown in Figure 4A, in adults all tested miRNAs exhibit nearly the same tissue-specific expression pattern with signals detected only in the brain. This is consistent with miR-A1/miR-323 and miR-A2/miR-329 being originally isolated from rat neurones (Kim et al. 2004). Furthermore, within the developing embryos (E12.5, E14.5, and E17.5) all the tested miRNA genes are expressed at roughly the same levels not only in the head but also in the trunk and the placenta, indicating that other tissues distinct from the brain express them as well (Fig. 4A; data not shown). In neonates (P0, P2, P7), miRNAs are still detected in the trunk, even though we noticed a reproducible decrease in the expression level for most of them after birth (Fig. 4B). Curiously, although each primer extension product generally appears as a band doublet for a given miRNA, for samples extracted from the head or the trunk in neonates, several miRNAs exhibit a different intensity ratio

between the two vicinal cDNA bands. The biological relevance of this difference, if any, remains unclear (see also Table 1). Further studies are needed to identify cell type-specific expression patterns as well as mechanisms of miRNA gene regulation.

Maternal Expression of the miRNA Gene Cluster Is Regulated by a Distal Intergenic, Germ-Line-Derived DMR (IG-DMR).

Because of their location within a previously characterized imprinted domain, we next examined whether the novel miRNA genes are subjected to genomic imprinting. As shown in Figure 4C, E15.5 embryos with a maternal uniparental disomy of chromosome 12 (matUPD12) express all the miRNAs that we have tested, whereas embryos having two paternal chromosome 12 copies (patUPD12) do not, showing that these miRNA genes are only expressed from the maternally inherited chromosome. We noticed that miRNA signals detected in the matUPD embryos are not twice as intense as those seen in the wild-type embryos. Al-

Seitz et al.

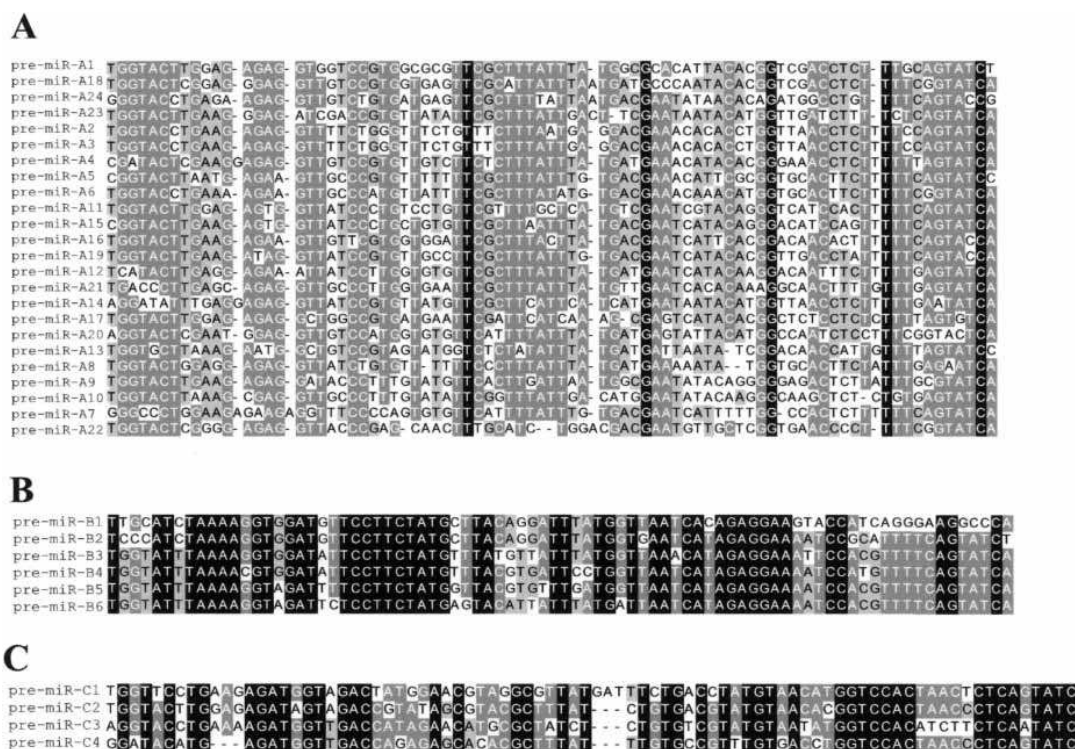


Figure 3 Multiple sequence alignment of the three human miRNA families. The names of the pre-miRNA variants belonging to the A-, B-, and C-types are indicated at the left in A, B, and C, respectively. Conservation of each nucleotide position is denoted by shading according to GeneDoc software: black (100%), dark gray (80% to 99%), and clear gray (60% to 79%). Sequence alignment of the conserved murine pre-miRNAs belonging to the A-, B-, and C-types is shown in Supplemental data S2.

though we can not exclude that these observations have a biological significance, we believe that our primer extension assay is not quantitative enough to reproducibly observe the expected twofold increase in the *matUPD* embryos. An IG-DMR located between the *Dlk1* and *Gtl2* genes has been recently shown to play a key role in controlling imprinted expression within the whole locus (Fig. 1). Deletion of the unmethylated IG-DMR from the maternally inherited chromosome results in the silencing of all the maternally expressed imprinted genes together with loss of imprinting (biallelic expression) of the maternally repressed genes. When inherited from the paternal allele, the deletion of the methylated IG-DMR does not affect imprinted gene expression (Lin et al. 2003). All the miRNAs that we have checked were absent from total RNA extracted from embryos having a maternal deletion of the IG-DMR, whereas those with the paternally inherited deletion had miRNA expression similar to that of the control unmanipulated littermates (Fig. 4D). We therefore conclude that, similar to the other maternally expressed noncoding RNA genes (anti-Rtl1, *Gtl2*, and the C/D snoRNA genes; Lin et al. 2003), the expression of the miRNA gene cluster is controlled by this distal *cis*-acting element.

The miRNAs Are Processed From Large Noncoding RNAs

All the miRNA genes described in this study share the same tissue-specific expression pattern with strongest expression in the adult brain. A similar gene expression pattern has been previously reported for the other neighboring maternally expressed

noncoding RNA genes (*Gtl2* and C/D RNA genes; Schmidt et al. 2000; Cavaille et al. 2002). Very little is known about miRNA transcription and its regulation in mammals (Bartel 2004). Although a few of these clustered miRNA genes are embedded within introns or exons of *Mirg*, a maternally expressed gene lacking a conserved protein-coding potential (Seitz et al. 2003), the genomic organization of most of them is still unclear (Figs. 2B, 5). Each pre-miRNA gene might be transcribed from its own tissue-specific promoter or alternatively it might be processed from (some) polycistronic-like RNA(s) as already reported for other clustered miRNAs (Lee et al. 2002). To address this second possibility, we have used RT-PCR to detect primary precursors encompassing several miRNAs relatively close to each other. As shown in Figure 5, transcripts encompassing several neighboring miRNAs are readily detected. Although we cannot rule out that some miRNA genes are transcribed from their own promoter, it seems more likely that most, if not all, pre-miRNA genes are generated through RNA processing either from polycistronic RNA and/or from intronic sequences rather than being individually expressed as ~70-nt-long primary transcripts. Thus, the novel clustered miRNA genes share a number of common properties with the upstream C/D snoRNA genes (Cavaille et al. 2002). Both classes of small noncoding RNA genes are (1) relatively weakly conserved during evolution, (2) organized as repeated arrays and processed from larger noncoding RNA genes, (3) transcribed in the same direction from the maternal chromosome only, (4) regulated by the IG-DMR, and (5) expressed in a tissue-specific manner with strongest expression in the adult

Table 1. List of Experimentally Detected or Cloned miRNAs at the Mouse Distal 12 Domain

miRNA ^a	Sequence (5' to 3') ^b	Maternally expressed miRNAs	References
miR-342	UCUCACACAGAAUCCGACCCGUC	n.d.	Kim et al. 2004; this study
miR-345	UGCUGACCCUAGUCCAGUGC	—	Kim et al. 2004; this study
miR-127	UCGGAUCCGUCUGAGCUUGGCU	+	Lagos-Quintana et al. 2002; Seitz et al. 2003
miR-136	ACUCCAUUUGUUUUGAUGUGGA	+	Lagos-Quintana et al. 2002; Seitz et al. 2003
miR-134	UGUGACUGGUUGACACAGGG	+	Lagos-Quintana et al. 2002; this study
miR-154/miR-A19-5' ^c	UAGGUUAUCCGUGUUGCCUUCG	+	Lagos-Quintana et al. 2002; this study
miR-323/miR-A1-3'	GCACAUUACAGGUCGACCUUCU	n.d.	Kim et al. 2004; this study
miR-329/miR-A2-3'	AACACACCCAGCUAACCUUUUU	n.d.	Kim et al. 2004; this study
miR-300/miR-A9-3'	UAUGCAAGGGCAAGCUCUCUUC	n.d.	Houbaviy et al. 2003; this study
miR-409/miR-A22-3'	GAAUGUUGCUCGGUGAACCCUU	n.d.	this study
miR-410/miR-A24-3'	AAUAUAACACAGAUGGCCUGUU	+	this study
miR-376b/miR-B4	AUCAUAGAGGAACAUCACUUU	+	this study
miR-376/miR-B6	AUCGUAGAGGAAAUCCAUGUU	+	this study
miR-411/miR-C2	AACACGGUCCACUAACCCUACAU	+	this study
miR-380-3p/miR-C3	UAUGUAGUAGGUCCACAUUUU	+	this study
miR-299/miR-D	UGGUUUUACCGUCCCAUACAU	n.d.	Houbaviy et al. 2003; this study
miR-412/miR-K	ACUUCACCGUGGUCACUAGCCGU	n.d.	this study
miR-337/miR-M	UCAGUCCUUAUAGUUGCCUUUC	+	Kim et al. 2004; this study
miR-370/miR-N ^d	GCCUGCUGGGUGGAACCCUGUU	+	this study
miR-341 ^e	UCGAUCGGUCGUCGGUCAGU	n.d.	Kim et al. 2004

^aNovel miRNA sequences have been submitted to the miRNA registry (<http://www.sanger.ac.uk/cgi-bin/Rfam/mima/browse.pl>) and they have been given new names in accordance to the miRNA registry numbering (e.g. miR-A22-3' is also called miR-409 in the miRNA registry).

^bThe exact length of the miRNAs which have been detected solely by primer extension is not known. Thus, the two last nucleotides at the 3'-termini are in italics as they are only predicted based on a 23 nt theoretical long RNA species. Primer extension assay generates two cDNA products with a size differing by one nucleotide. Based on the miR-134 and miR-154 sequences obtained independently by cloning strategies (Lagos-Quintana et al. 2002), we have considered the shorter cDNA product as the correct 5' termini while the longer one might correspond to in vivo processing heterogeneity and/or addition of an extra nucleotide in a matrix-independent manner by the AMV reverse transcriptase (Promega technical service, pers. comm.).

^cmiR-154/A19 are processed from the 5' side of A19 (Lagos-Quintana et al. 2002) while miR-323/A1, miR-329/A2, miR-A9, miR-A22 and miR-A24 are processed from the 3' strand of the A1, A2, A9, A22 and A24 gene copies, respectively suggesting that both strands of A-type can potentially be converted to miRNAs.

^dDepending upon RNA samples, we could also detect a ladder-like pattern superimposed to the mature 5' end of miR-N. Thus, this miRNA might not fulfill the stringent criteria described in (Ambros et al. 2003).

^emiR-341 was not found in our in silico search as it is not conserved at the human imprinted 14q32 interval.

n.d. indicates not determined.

brain. Thus, it is tempting to speculate that C/D RNAs and miRNAs are processed from a single, large primary transcript that might initiate at the Gtl2 promoter. Future work is now needed to fully characterize this atypical and complex transcription unit.

DISCUSSION

Most imprinted genes play important roles in embryonic and/or placental growth (Tycko and Morison 2002) and developmental abnormalities in mouse and human have been previously associated with the lack of expression of the maternally inherited genes at distal 12 and 14q32, respectively (Georgiades et al. 2000; Kurosawa et al. 2002). Given the known roles of miRNAs in development (Carrington and Ambros 2003), it is likely that these novel imprinted miRNAs mediate a function during mouse development. In contrast to the *mir-127* and *mir-136* genes encoded in the antisense strand of the *Rtl1* gene (Seitz et al. 2003), none of these novel maternally expressed miRNAs exhibit full complementarity to a cellular mRNA, suggesting they probably inhibit mRNA translation rather than triggering mRNA degradation. Indeed, a role in translation control is supported by the association of miR-A1/miR-323 with the polyribosomes in rat primary cortical cultures (Kim et al. 2004). So far, despite the identification of >150 miRNAs from mouse and human, their biological functions remain highly elusive, mainly due to the difficulty in finding convincing RNA target(s). Through computer searches, cross genome comparison (i.e., pufferfish versus mammalian) are commonly used to predict conserved mRNA targets (Lim et al. 2003). Because the miRNAs described in this study seem to be restricted

to mammalian species, assigning mRNA target(s) for these imprinted miRNAs might be a difficult task. Moreover, given that the sequences of most of clustered pre-miRNA genes are related to some extent, we could envision that (1) they regulate the same mRNA (or a common class of mRNAs) by interacting at various binding sites along the 3' untranslated region (3' UTR); or (2) conversely, the relative lack of strong selection pressure on several miRNAs (see Supplemental data S2) might also allow them to target multiple and distinct mRNA targets. In the sheep, deregulated gene expression at the *Dlk1-Gtl2* domain (Charlier et al. 2001) is associated with the *Clpg* phenotype, a muscular hypertrophy manifest only in heterozygous individuals who have inherited the *Clpg* mutation from their paternal chromosome, but not observed in *Clpg/Clpg* homozygotes. One mechanism postulated to explain this unexpected mode of inheritance is a *trans* interaction between the reciprocally imprinted genes (e.g., a *trans*-acting repressor expressed from the maternal chromosome could negatively regulate some paternally expressed genes; for a review, see Georges et al. 2003). The miRNA genes described in this study are obvious candidates for playing such a role.

RNA-mediated gene silencing mechanisms are thought to operate at the imprinted *Igf2r-Air* domain as well as in X-chromosome inactivation in female mammalian cells (Plath et al. 2002; Sleutels et al. 2002). Moreover RNA-mediated formation of heterochromatin has been observed in *S. pombe* (Hall et al. 2002; Volpe et al. 2002; Schramke and Allshire 2003; Verdel et al. 2004), and small RNAs play a role in the DNA methylation and

Seitz et al.

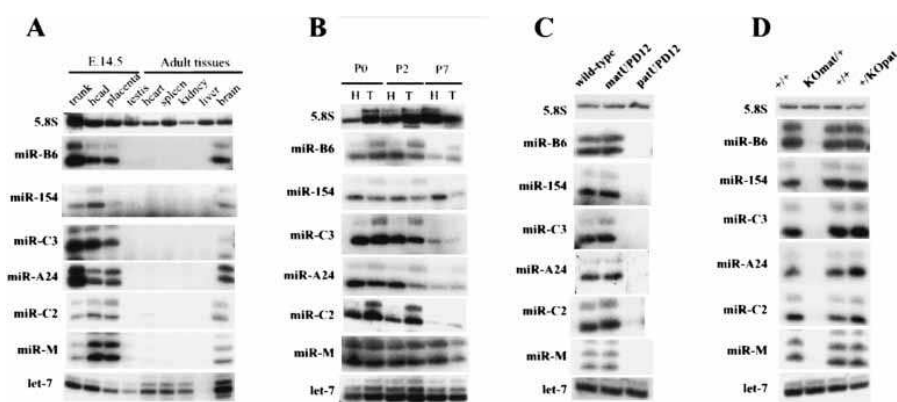


Figure 4 Expression of in silico-predicted miRNA genes. Tissue-specific (A) and developmental gene expression pattern (B) of the miRNAs analyzed by primer extension assay. H indicates head; T, trunk. 5.8S and let-7 probes have been used as gel loading controls (Northern blot analysis). (C) The miRNA genes are only expressed from the maternally inherited chromosome. Ten micrograms of total RNA extracted from wild-type embryos, embryos with maternal uniparental disomy for chromosome 12 (matUPD12), or embryos with paternal uniparental disomy for chromosome 12 (patUPD12) have been subjected to primer extension assay. (D) The miRNA imprinted gene expression is controlled by the intergenic germline-derived differentially methylated region (IG-DMR). Ten micrograms of total RNA extracted from embryo with a deletion of the IG-DMR from either the maternally or the paternally inherited chromosomes (as specified at the top of the picture) have been subjected to primer extension analysis.

silencing of target sequences in plants (Aufsatz et al. 2002). Thus, an emerging theme is the idea that noncoding RNAs have key functions in genome modification. Whether the clustered miRNAs can control gene expression at the transcriptional level is still an open and challenging question. We noticed that four miRNA genes—miR-A22, miR-A23, miR-A24, and miR-K—are embedded within a conserved CpG island located at the 3' end of the miRNA gene cluster. The methylation status of this CpG-rich domain is under investigation.

Did this large cluster of miRNAs emerge within this imprinted domain before or after the onset of imprinting at this region? If the former, did it confer some aspect of imprinting control, or if the latter, is the imprinted expression of these small hairpin RNAs a mere consequence of more widespread expres-

sion of noncoding RNAs from the maternal chromosome? One thought-provoking idea pertaining to the evolution of the imprinting mechanism suggests that intragenomic parasitic sequences, including repeated sequences, retrotransposons, and other related elements, are heritably and variably silenced in order that they are rendered transcriptionally inactive, "hidden" by heterochromatinization, and/or immobilized (Barlow 1993; Bird 1997; Yoder et al. 1997; Martienssen 1998; Whitelaw and Martin 2001). Although this hypothesis was based on a function for DNA methylation, it could equally be applied to a small RNA-mediated mechanism. Thus, one can speculate that these repeated small RNA genes might have integrated into this chromosomal domain prior to the emergence of its imprinting, and their subsequent gene amplification may have been perceived by the

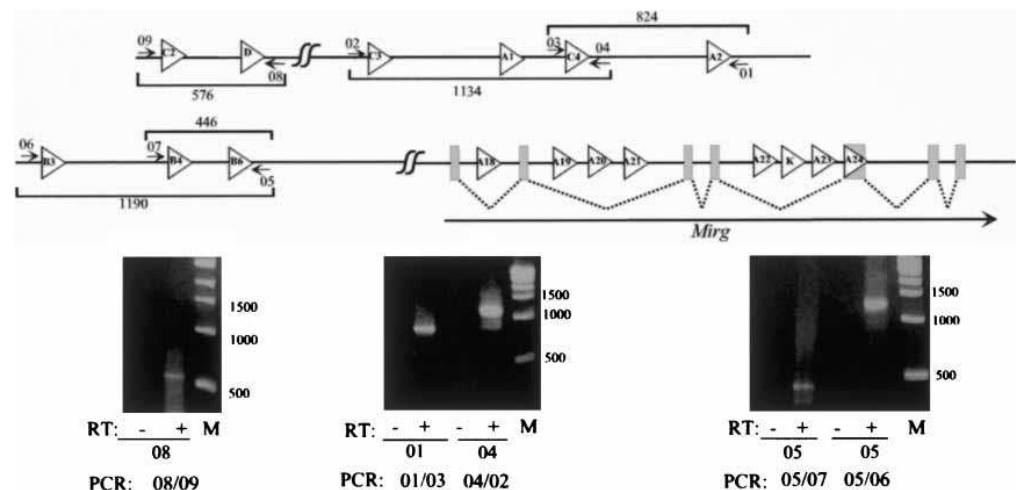


Figure 5 Detection of large transcripts overlapping several miRNA genes. Ten micrograms of DNase I-, proteinase K-treated total RNAs extracted from the head of mouse embryo (E17.5) was subjected to RT-PCR with appropriate primers (numbered arrows) mapping outside from the ~70-nt pre-miRNA genes. The location of the primers and expected sizes of the amplification products are shown on the top part of the figure (not drawn to scale). Primer sequences are available on request. RT indicates reverse transcriptase (Superscript II, Invitrogen).

host as "foreign DNA," resulting in imposition of epigenetic regulation to the surrounding protein-coding genes. Conversely, the miRNA gene cluster might have been generated after the establishment of the imprinting, perhaps due to a specific chromatin architecture prone to gene duplication and/or to the different behavior of the two parental chromosomes, which is permissive for nondisruptive insertion. It remains to be determined whether the parental origin specific nature of the miRNA expression contributes to, or is a consequence of, genomic imprinting.

An exhaustive characterization of all the epigenetic marks along this large imprinted domain together with genetically engineered mouse models is now required to fully appreciate the potential involvement of these imprinted miRNA genes in mouse development and/or epigenetic control.

METHODS

Unless otherwise noted, all techniques for manipulating nucleic acids were performed according to standard protocols.

Oligodeoxynucleotides

They were all synthesised by Y. de Préal (LBME) on a PerSeptive Biosystems Expedite apparatus. Primer sequences are available on request

Searching for miRNA Genes at Human Imprinted 14q32 Domain

BAC sequences around DLK1/GTL2 positions were retrieved from the Human Genome Project Working Draft (<http://genome.ucsc.edu/>) and 100-nt-long stem-loop structures with a $\Delta G < -30$ kcal/mole were systematically identified within the ~1-Mb-long imprinted 14q32 region by using RNAfold software from <http://www.tbi.univie.ac.at/~ivo/RNA/>. We then retained the conserved hairpin structures between the murine and human orthologous imprinted loci (according to BLAT search at <http://genome.ucsc.edu/>) that give rise to a miRscan score >10 by using the software developed by D.P. Bartel and his colleagues (<http://genes.mit.edu/mirscan/>). Sequence alignment of the miRNA genes were obtained by MultiAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>), and conserved nucleotides were shaded by GeneDoc (<http://www.Cris.com/~ketchup/genedoc.shtml>).

Experimental Detection of the miRNAs

In silico-predicted miRNA genes were validated either by Northern blot analysis or by primer extension. Total RNA was isolated by the guanidium thiocyanate method from mouse tissues or from whole embryos freshly prepared according to French institutional and United Kingdom Home Office guidelines. For Northern blot analyses, 15 to 20 μg of DNase I-, protease K-treated total RNA was separated on a denaturing 15% acrylamide/7M urea gel. Electrophoresis in 0.5 $\times$ TBE buffer was performed onto nylon membranes (Hybond N+) followed by an UV light irradiation of the membrane. Hybridizations were carried out with 5' (^{32}P)-labeled oligodeoxynucleotide probes, through an overnight incubation in 5 $\times$ SSPE, 1% SDS, 150 $\mu\text{g}/\text{mL}$ yeast tRNA, and 5 $\times$ Denhardt's. Membranes were washed twice with 2 $\times$ SSPE and 0.1 % SDS at room temperature before autoradiography. For primer extension assay, after a heat denaturation step (5 min at 75°C), 10 μg of total RNA was annealed with a (^{32}P)-labeled and gel-purified specific primer (see Supplemental data) and reverse transcribed during 2 h at 42°C by AMV reverse transcriptase (Promega), and the resulting cDNA products were analyzed on a 15% acrylamide /7 M urea gel.

ACKNOWLEDGMENTS

This work has been initiated in the Dr. Jean-Pierre Bachelier's laboratory, and we thank him very much for his permanent support and for the critical reading of the manuscript. We are also grateful to D. Morello for her help with mouse manipulation. Work in the laboratory of J.C. is supported by the Ministère de

l'éducation Nationale, de la Recherche et de la technologie (Action Concertée Incitative 2003: Biologie cellulaire moléculaire et structurale), and work in the laboratory of A.F.-S. is funded by a CR-UK grant. S.-P.L. is funded by a graduate research scholarship from the Taiwanese Government; H.S., by a PhD fellowship Allocation de Moniteur Normalien (Ministère de l'Éducation Nationale, de la Recherche et de la technologie). The authors declare that they have no competing financial interests.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

NOTE ADDED IN PROOF

Human miRNAs derived from our computationally predicted pre-miR-B3 and pre-miR-A23 have been experimentally cloned by Suh et al. (2004). They have been named miR-368 and miR-369, respectively.

REFERENCES

- Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. 2002. RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **99**: 16499–16506.
- Aukerman, M.J. and Sakai, H. 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730–2741.
- Barlow, D.P. 1993. Methylation and imprinting: From host defense to gene regulation? *Science* **260**: 309–310.
- Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.
- Bird, A. 1997. Does DNA methylation control transposition of selfish elements in the germline? *Trends Genet.* **13**: 469–472.
- Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**: 25–36.
- Carrington, J.C. and Ambros, V. 2003. Role of microRNAs in plant and animal development. *Science* **301**: 336–338.
- Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C., and Bachelier, J.P. 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* **11**: 1527–1538.
- Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., and Georges, M. 2001. The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nat. Genet.* **27**: 367–369.
- Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83–86.
- Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* **303**: 2022–2025.
- Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. 2003. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* **9**: 180–186.
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. 2003. MicroRNA targets in *Drosophila*. *Genome Biol.* **5**: R1.
- Georges, M., Charlier, C., and Cockett, N. 2003. The callipyge locus: Evidence for the *trans* interaction of reciprocally imprinted genes. *Trends Genet.* **19**: 248–252.
- Georgiades, P., Watkins, M., Surani, M.A., and Ferguson-Smith, A.C. 2000. Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* **127**: 4719–4728.
- Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**: 2232–2237.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**: 4671–4679.
- Houbaviy, H.B., Murray, M.F., and Sharp, P.A. 2003. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5**: 351–358.
- Hutvagner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056–2060.
- Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. 2001. A cellular function for the RNA-interference

Seitz et al.

- enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.
- Johnston, R.J. and Hobert, O. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845–849.
- Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & Dev.* **15**: 2654–2659.
- Kim, J., Krichevsky, A.M., Grad, Y., Hayes, G.D., Kosik, K.S., and Church, G.M. 2004. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl. Acad. Sci.* **101**: 360–365.
- Kurosawa, K., Sasaki, H., Sato, Y., Yamanaka, M., Shimizu, M., Ito, T., Okuyama, T., Matsuo, M., Imaizumi, K., Kuroki, Y., et al. 2002. Paternal UPD14 is responsible for a distinctive malformation complex. *Am. J. Med. Genet.* **110**: 268–272.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**: 735–739.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. 2003. New microRNAs from mouse and human. *RNA* **9**: 175–179.
- Lai, E.C. 2003. microRNAs: Runt of the genome assert themselves. *Curr. Biol.* **13**: R925–R936.
- Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–862.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–854.
- Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. 2002. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO J.* **21**: 4663–4670.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415–419.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.
- Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. 2003. Vertebrate microRNA genes. *Science* **299**: 1540.
- Lin, S.P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. 2003. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat. Genet.* **35**: 97–102.
- Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**: 2053–2056.
- Martienssen, R. 1998. Transposons, DNA methylation and gene control. *Trends Genet.* **14**: 263–264.
- Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**: 5194–5201.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. 2002. miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes & Dev.* **16**: 720–728.
- Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. 2003. Control of leaf morphogenesis by microRNAs. *Nature* **425**: 257–263.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. 2002. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**: 315–327.
- Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. 2004. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**: 669–672.
- Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. 2002. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**: 233–278.
- Reinhart, B.J. and Bartel, D.P. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831.
- Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–906.
- Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes & Dev.* **16**: 1616–1626.
- Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. 2002. Prediction of plant microRNA targets. *Cell* **110**: 513–520.
- Schmidt, J.V., Matteson, P.G., Jones, B.K., Guan, X.J., and Tilghman, S.M. 2000. The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. *Genes & Dev.* **14**: 1997–2002.
- Schramke, V. and Allshire, R. 2003. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301**: 1069–1074.
- Schwarz, D.S. and Zamore, P.D. 2002. Why do miRNAs live in the miRNP? *Genes & Dev.* **16**: 1025–1031.
- Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., and Cavaille, J. 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* **34**: 261–262.
- Sleutels, F., Zwart, R., and Barlow, D.P. 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**: 810–813.
- Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. 2003. Identification of *Drosophila* MicroRNA targets. *PLoS Biol.* **1**: E60.
- Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., et al. 2004. Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* **270**: 488–498.
- Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. 2003. A biochemical framework for RNA silencing in plants. *Genes & Dev.* **17**: 49–63.
- Tycko, B. and Morison, I.M. 2002. Physiological functions of imprinted genes. *J. Cell. Physiol.* **192**: 245–258.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672–676.
- Verona, R.L., Mann, M.R., and Bartolomei, M.S. 2003. Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu. Rev. Cell. Dev. Biol.* **19**: 237–259.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–1837.
- Whitelaw, E. and Martin, D.I. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**: 361–365.
- Xie, Z., Kasschau, K.D., and Carrington, J.C. 2003. Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.* **13**: 784–789.
- Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. 2003. The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**: 790–795.
- Yao, M.C., Fuller, P., and Xi, X. 2003. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* **300**: 1581–1584.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335–340.
- Zilberman, D., Cao, X., and Jacobsen, S.E. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716–719.

WEB SITE REFERENCES

<http://genome.ucsc.edu/>; Human Genome Project Working Draft.
<http://www.tbi.univie.ac.at/ivo/RNA/>; RNA fold software.
<http://genes.mit.edu/mirscan/>; software developed by D.P. Bartel and colleagues.
<http://prodes.toulouse.inra.fr/multalin/multalin.html>; MultiAlin.
<http://www.Cris.com/ketchup/genedoc.shtml>; GeneDoc.

Received April 1, 2004; accepted in revised form June 16, 2004.

2.3.2 Six gènes de microARN additionnels

En raison de la similarité, en séquence, de nombreux candidats (en particulier, la vingtaine de membres de la « famille A », relativement apparentés, n'auraient pas pu être distingués chacun par extension d'amorce ou Northern blot), nous n'avons pas pu vérifier l'expression de plusieurs gènes de microARN prédits. Mais simultanément, plusieurs laboratoires (qui réalisaient des banques de microARN murins ou humains) ont cloné les produits de plusieurs de ces gènes prédits, démontrant qu'il s'agissait d'authentiques gènes de microARN.

Ainsi, les microARN humains miR-A23 (appelé : miR-369) et miR-B3 (appelé : miR-368) ont été clonés à partir de cellules-souches embryonnaires humaines (Suh *et al.*, 2004), et les microARN miR-A10 (alors nommé miR-381), miR-A16 (nommé miR-382), miR-A21 (nommé miR-377) et miR-C1 (nommé miR-379) ont été clonés à partir d'une lignée cellulaire murine (S. Griffiths-Jones, communication personnelle; <http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml>).

2.3.3 Le phénotype callipyge du Mouton : implication des microARN ?

Le phénotype appelé « callipyge » (pour : « belles fesses », en grec), chez le Mouton, se caractérise par une hypertrophie musculaire généralisée, mais principalement visible sur l'arrière-train (Koochmaraie *et al.*, 1995 ; Cockett *et al.*, 2001). Ce caractère est monogénique, et dû à une mutation portée par le chromosome 18 ovin (Cockett *et al.*, 1994), sur un locus orthologue au locus humain 14q32 (Berghmans *et al.*, 2001) ; ce locus est également soumis à l'empreinte génomique parentale chez le Mouton (Charlier *et al.*, 2001b). Au niveau histologique, le phénotype callipyge apparaît comme une hypertrophie des fibres musculaires rapides, qui se manifeste à partir de la troisième semaine après la naissance (Cockett *et al.*, 2001).

Le mode de transmission de ce phénotype est très intrigant : les homozygotes mutants ne présentent pas le phénotype (la règle de transmission de ce caractère est donc un cas de « surdominance »), et parmi les hétérozygotes, seuls ceux qui ont hérité de la mutation par leur père sont de phénotype mutant, les autres sont de phénotype sauvage (la surdominance est dite « polaire ») (Cockett *et al.*, 1996) (le tableau 2.1 résume les règles de transmission de ce phénotype).

Génotype (mat./pat.)	phénotype
$+/+$	[sauvage]
$clpg/+$	[sauvage]
$+/clpg$	[clpg]
$clpg/clpg$	[sauvage]

TAB. 2.1 – RÈGLES DE TRANSMISSION DU PHÉNOTYPE CALLIPYGE CHEZ LE MOUTON. « mat. » : allèle maternel ; « pat. » : allèle paternel. « clpg » : callipyge.

La mutation *callipyge* a été identifiée : c'est une transition A→G, dans une séquence non-codante entre les gènes *Dlk1* et *Gtl2* (Freking *et al.*, 2002 ; Smit *et al.*, 2003). Elle n'affecte pas l'empreinte des gènes du locus, mais active, en *cis*, leur expression (du moins : l'accumulation de leurs ARN) (Charlier *et al.*, 2001a) : lorsque la mutation touche le chromosome paternel (chez les homozygotes, et chez les hétérozygotes ayant hérité de la mutation par leur père), les ARN de *DLK1* et *RTL1/PEG11*⁴ sont sur-accumulés ; lorsque la mutation affecte le chromosome maternel (chez les homozygotes, et chez les hétérozygotes ayant hérité de la mutation par leur mère),

⁴N.B. : à l'époque, il n'était pas encore établi que *Dio3* est également soumis à l'empreinte génomique parentale (Tsai *et al.*, 2002 ; Yevtodiyenko *et al.*, 2002).

les ARN de *GTL2* et *MEG8*⁵ sont sur-accumulés. Outre cet effet *cis*, la mutation a également un effet *trans* : portée par le chromosome maternel, elle réprime *DLK1* et *RTL1/PEG11*, alors que, portée par le chromosome paternel, elle active *Gtl2* (Charlier *et al.*, 2001a). Ces observations sont résumées par la figure 2.5.

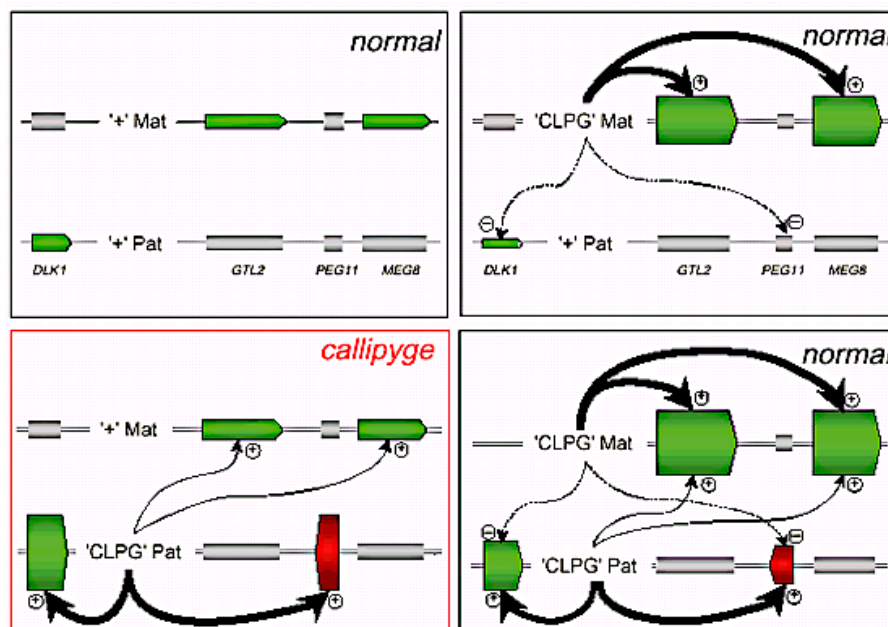


FIG. 2.5 – LA MUTATION CALLIPYGE DÉRÈGULE LES GÈNES DU LOCUS 18q OVIN. Figure tirée de Charlier *et al.*, 2001a. Le phénotype des animaux est indiqué en haut à droite de chaque schéma. Les gènes transcrits du côté « centromère » au côté « télomère » sont représentés en vert, et le gène transcrit du côté « télomère » au côté « centromère » est représenté en rouge. L'effet *cis* de la mutation callipyge est représenté par des flèches épaisses, et l'effet *trans*, par des flèches fines.

Finalement, le génotype associé au phénotype callipyge (le génotype $+^{mat.}/clpg^{pat.}$) se caractérise par la plus forte activation de *DLK1* et *RTL1/PEG11*, et l'absence d'activation des gènes à expression maternelle. Le phénotype pourrait donc être dû à la surexpression de *DLK1* ou *RTL1/PEG11* ; tous les gènes étudiés du locus sont exprimés dans le muscle (Charlier *et al.*, 2001b), donc chacun pourrait être impliqué dans l'apparition du phénotype. La surdominance polaire pourrait alors révéler une interaction en *trans* entre des gènes du locus : la répression d'un gène exprimé uniquement à partir d'un allèle parental, par un gène exprimé uniquement à partir de l'autre allèle parental, expliquerait la surdominance (l'activation des deux antagonistes, par les deux mutations *callipyge*, s'annulerait), et l'empreinte génomique parentale expliquerait la polarité de la surdominance (si le phénotype est dû à la surexpression d'un gène à expression paternelle, ou, plus subtilement, à la surexpression d'un gène à expression paternelle couplée à une expression modérée d'un gène à expression maternelle, seuls les hétérozygotes ayant hérité de la mutation par leur père présenteraient le phénotype). Il a donc été proposé que la surexpression de *DLK1* ou de *RTL1/PEG11* était responsable du phénotype callipyge, et que *GTL2* ou *MEG8* était un répresseur du gène à expression paternelle responsable du phénotype (Georges *et al.*, 2003).

Les microARN que nous avons décrits sont des candidats crédibles pour jouer ce rôle de

⁵Le gène *MEG8* avait été décrit comme un gène non-codant à expression maternelle, avant la découverte des gènes d'ARN C/D répétés ; c'est en réalité l'extrémité 5' du gène-hôte des ARN C/D (Cavaillé *et al.*, 2002).

répresseurs à expression maternelle : tous sont exprimés à partir de l'allèle maternel uniquement, et les microARN sont connus pour réprimer des gènes-cibles spécifiques, par RNAi ou répression traductionnelle (voir Introduction, partie 1.4.3, page 68). Ils pourraient donc subir le contrôle de la mutation *callipyge* (en particulier, ils pourraient être cotranscrits avec *MEG8*, gène-hôte des ARN C/D : voir Discussion, page 159), comme les gènes en amont.

Dans le cadre de cette hypothèse, un gène à expression paternelle se présente comme une cible évidente des microARN : *RTL1/PEG11*, parfaitement complémentaire de deux microARN du locus, et dont des données génétiques suggèrent effectivement qu'il serait réprimé par des gènes à expression maternelle (voir pages 118 et 122, et les articles n°2 et 3). En particulier, *RTL1/PEG11* est largement surexprimé, chez les moutons qui présentent le phénotype, dans les muscles hypertrophiés ; la corrélation des abondances de l'ARN *RTL1/PEG11* et de son ARN antisens (précurseur de miR-127 et miR-136) dans les différents génotypes s'accorderait bien avec un modèle dans lequel *PEG11* est le gène à expression paternelle responsable du phénotype, et les microARN, ses répresseurs à expression maternelle (Bidwell *et al.*, 2004). En revanche, aucun indice ne permet d'imaginer par quels moyens la surexpression de *RTL1/PEG11* pourrait aboutir à une hypertrophie musculaire : ce gène appartient à une famille de rétrotransposons, et n'a donc pas de fonction cellulaire *a priori* ; son ARN est polyadénylé, mais sa traduction n'a pas été démontrée. Cependant, la conservation de sa phase ouverte de lecture chez tous les Mammifères analysés suggère qu'il a acquis une fonction profitable à l'organisme qui l'héberge (Lynch et Tristem, 2003).

Un deuxième candidat, *DLK1*, pourrait être responsable du phénotype *callipyge*, et réprimé par des microARN. En effet, la surexpression par des souris transgéniques d'une isoforme de *DLK1* ovin provoque une hypertrophie musculaire qui évoque le phénotype ovin (Davis *et al.*, 2004).

Puisqu'aucun microARN n'est parfaitement complémentaire de l'ARN_m *DLK1*, si *DLK1* était le gène dont la surexpression provoque le phénotype, son répresseur pourrait être un microARN imparfaitement complémentaire, qui en inhiberait la traduction. Cette régulation expliquerait pourquoi la protéine *DLK1* n'est pas détectée dans les muscles des animaux homozygotes mutants, alors qu'elle est détectée dans les muscles des hétérozygotes *+^{mat.}//clpg^{pat.}* (Georges *et al.*, 2003). Et puisque seules les séquences des microARN murins sont connues (ou de leurs orthologues humains dans quelques cas : Suh *et al.*, 2004), il est encore impossible de rechercher des complémentarités entre les microARN ovins et l'ARN_m *DLK1* ovin. En posant l'hypothèse que les gènes de microARN du locus sont conservés chez le Mouton⁶, il est en revanche possible de chercher des complémentarités entre la séquence ovine de *DLK1* et les microARN murins, et de mettre en évidence des complémentarités que l'ARN_m *DLK1* ovin pourrait présenter aux microARN ovins.

J'ai donc recherché des complémentarités entre l'ARN_m *DLK1* ovin et les microARN exprimés à partir du locus distal 12 murin, par trois programmes décrits dans la littérature (Lewis *et al.*, 2003 ; Enright *et al.*, 2003 ; Kiriakidou *et al.*, 2004), et par un quatrième, écrit pendant ma thèse, qui ne tient compte que de la complémentarité de l'extrémité 5' du microARN à l'ARN_m, et recherche celles qui satisfont les règles établies par Kiriakidou *et al.*, 2004 (les effets de mutations sur l'extrémité 3' du microARN n'ont pas été aussi bien caractérisés). Ces différents programmes donnent des résultats nettement différents, et de manière générale, ne prédisent

⁶Pour l'instant, la séquence génomique ovine du locus complet n'est pas disponible ; la partie proximale du cluster R est toutefois séquencée (C. Charlier, non publié), et les gènes de miR-370/miR-N, miR-299/miR-D, miR-411/miR-C2 et miR-380-3p/miR-C3 y sont reconnaissables : ils sont conservés entre la Souris et le Mouton, ce qui suggère que le Mouton exprime un microARN pratiquement identique à miR-370/miR-N murin (avec une unique transition C→U), un microARN identique à miR-299/miR-D murin, un microARN identique à miR-411/miR-C2 murin, et un microARN qui présenterait trois transitions (A→G et G→A) par rapport au microARN miR-380-3p/miR-C3 murin (J. Cavaillé, communication personnelle).

pas plus de complémentarités fonctionnelles entre les microARN et l'ARN_m DLK1, qu'entre les microARN et un échantillon de gènes murins de taille comparable à celle de la séquence Dlk1 utilisée, et de composition nucléotidique similaire, choisis aléatoirement⁷. Les programmes actuels de recherche de complémentarités fonctionnelles ne prédisent donc pas d'interactions significatives entre les séquences murines des microARN, et la séquence ovine de l'ARN_m DLK1. Il est toutefois possible qu'un (ou : des) microARN spécifiquement ovin(s) puisse(nt) réprimer *DLK1* ; il est également probable que nos connaissances des déterminants de la répression traductionnelle guidée par les microARN soient encore trop imparfaites pour en tirer des conclusions définitives (voir Introduction, partie 1.4.3, page 71).

Enfin, les gènes à expression paternelle *BEGAIN* (dont l'empreinte vient d'être démontrée chez le Mouton) et *DIO3* (dont l'empreinte n'était pas connue à la conception du modèle d'interaction en *trans* par Georges *et al.*, 2003) ne semblent pas impliqués dans l'apparition du phénotype callipyge, puisque leur expression n'est pas affectée par la mutation *callipyge* (Davis *et al.*, 2004).

⁷Les séquences des gènes Fanca, n° d'accension : AF208116 ; Ipo11, n° d'accension : BC029746 ; Jak1, n° d'accension : S63728 ; NFasc, n° d'accension : AJ543322 ; Pkd2, n° d'accension : AF014010 ; Rev11, n° d'accension : BC058093 ; Scn7a, n° d'accension : L36179 ; Setdb1, n° d'accension : AF091628.

Chapitre 3

Conclusion

C'est pas la taille qui compte.
(dicton populaire)

3.1 Bilan

Notre attention a été attirée sur le locus 14q32 (*Homo sapiens*) / distal 12 (*Mus musculus*) par la présence dans le locus orthologue, chez le Rat, d'un gène de snoRNA C/D répété et à expression restreinte au cerveau chez l'adulte, RBII-36 (Cavaillé *et al.*, 2001). Cette situation rappelait celle du locus 15q11-13 (*H. sapiens*) / 7C (*M. musculus*), d'autant que le locus 14q32 / distal 12 était lui aussi connu pour héberger des gènes soumis à l'empreinte génomique parentale (Miyoshi *et al.*, 2000 ; Schmidt *et al.*, 2000 ; Takada *et al.*, 2000).

Une recherche systématique de séquences répétées dans ce locus a révélé l'existence de gènes d'ARN C/D répétés, suivis de répétitions d'un autre type, le « cluster R » (Cavaillé *et al.*, 2002). Le clonage, par un autre laboratoire, de microARN dont les gènes sont situés dans le cluster R (Lagos-Quintana *et al.*, 2002) nous a amenés à rechercher les gènes de microARN dans ce locus, et à identifier plusieurs dizaines de candidats (Seitz *et al.*, 2004), dont plusieurs ont été confirmés expérimentalement, par nous-mêmes ou par des équipes indépendantes (Seitz *et al.*, 2004 ; Suh *et al.*, 2004 ; S. Griffiths-Jones, communication personnelle). Ce locus héberge donc une grande concentration de gènes de microARN, la plus grande connue à ce jour.

D'autre part, parmi les gènes des microARN clonés en 2002 par Lagos-Quintana et ses collaborateurs, deux sont également situés dans le locus 14q32 / distal 12, mais hors du cluster R : ces deux microARN, miR-127 et miR-136, sont transcrits en orientation antisens d'un rétrotransposon, *Rtl1*, et sont donc parfaitement complémentaires de deux séquences de son ARN_m ; ils sont donc susceptibles d'en guider la dégradation par interférence à ARN (Seitz *et al.*, 2003). Cette hypothèse est confortée par l'analyse de l'expression des gènes du locus chez des mutants dont l'empreinte est affectée dans le domaine distal 12 : une stabilisation de l'ARN *Rtl1* semble corrélée avec l'absence de miR-127 et miR-136 (Lin *et al.*, 2003).

Enfin, les gènes de tous les microARN étudiés ici, ainsi que ceux des ARN C/D étudiés ici, sont soumis à l'empreinte génomique parentale, et exprimés uniquement à partir de l'allèle maternel. Ils sont exprimés, chez la Souris, au cours du développement embryonnaire, puis leur expression semble se réduire, chez l'adulte, au cerveau. Tous ces gènes seraient d'évolution récente, puisqu'ils paraissent propres aux Mammifères.

Un article de revue est présenté en Annexe 3 (page 195) ; il résume nos travaux sur le locus 14q32 humain, et ses orthologues, et le compare au locus 15q11-13 humain (et ses orthologues).

3.2 Discussion

Les gènes de tous ces petits ARN ont été découverts par des techniques de recherche systématique, par clonage de petits ARN (Cavaillé *et al.*, 2001 ; Lagos-Quintana *et al.*, 2002) et par analyse de séquences génomiques (Cavaillé *et al.*, 2002 ; Seitz *et al.*, 2004). Ces deux approches ne donnent toutefois aucune information sur les fonctions éventuelles des gènes révélés.

3.2.1 Ces petits ARN non-codants ont-ils une fonction ?

Il est tout à fait possible que ces nouveaux gènes d'ARN C/D et de microARN n'aient aucun rôle physiologique. Ce pourraient être, par exemple, des éléments génétiques mobiles, dont la multiplication ne conférerait aucun avantage sélectif au génome-hôte. En effet, les gènes d'ARN C/D peuvent se comporter en éléments génétiques mobiles : plusieurs exemples de gènes mobiles d'ARN C/D et de rétrogènes d'ARN C/D ont été rapportés (ces ARN stables et nucléaires seraient des substrats privilégiés pour la reverse-transcription et l'insertion dans le génome) (Bachellerie *et al.*, 1995 ; Maxwell et Fournier, 1995 ; Renalier *et al.*, 1996 ; Bachellerie *et al.*, 2002 ; Cervelli *et al.*, 2002) ; quant aux gènes de microARN, ils pourraient également se déplacer et se multiplier dans les génomes (par rétrotransposition des précurseurs repliés en tige-boucle ?),

puisque plusieurs gènes de microARN possèdent des isoformes dispersées dans les génomes (ainsi, on dénombre actuellement une douzaine d'isoformes du gène *let-7* chez les Mammifères), et des transposons de Plantes présentent des similarités frappantes avec les gènes de microARN (en particulier, leurs transcrits, repliés en tige-boucle, sont digérés par une RNase III pour libérer de petits ARN d'une vingtaine de nucléotides) (Mette *et al.*, 2002). L'insertion de quelques gènes d'ARN C/D, et de quelques gènes de microARN, aurait été suivie de duplications, à l'origine de ces nombreuses répétitions de séquences apparentées. Les loci soumis à l'empreinte génomique parentale connaissent des fréquences de recombinaison beaucoup plus élevées que le reste du génome (spécialement au cours de la méiose de la lignée germinale femelle) (Lercher et Hurst, 2003); ce phénomène pourrait être dû à la conformation particulière de la chromatine de ces loci pendant la gamétogenèse, donc en particulier pendant la méiose. Des événements de recombinaison entre chromosomes homologues auraient pu conduire à la multiplication de ces gènes non-codants. Ainsi, quelques éléments transposables de types « ARN C/D » et « microARN », initialement insérés dans un locus qui se trouvait soumis à l'empreinte génomique parentale, auraient été dupliqués à plusieurs reprises, pour atteindre aujourd'hui plusieurs dizaines de copies. Cette amplification serait donc une conséquence de l'empreinte de leur locus, ce qui expliquerait qu'on ne connaît pas, à l'heure actuelle, d'autre concentration de gènes de microARN de cette ampleur, et que le seul autre locus contenant une telle concentration de gènes d'ARN C/D, le locus 15q11-13, soit lui aussi soumis à l'empreinte génomique parentale (Cavaillé *et al.*, 2000; voir Résultats, page 115). La figure 3.1, page 153, résume la séquence d'événements qui aurait ainsi pu aboutir à la multiplication de rétrogènes dans le locus. Par dérive génétique, un allèle constitué de nombreuses copies de rétrogènes serait alors devenu majoritaire dans une population.

Ce modèle exige une série de coïncidences, pour que le nombre de copies dans une population augmente progressivement par dérive génétique, sans pression de sélection. Si ces petits ARN jouaient une fonction dans la cellule, leur amplification progressive au cours de l'évolution s'expliquerait plus facilement.

3.2.2 Un rôle des petits ARN non-codants dans le développement ?

Les embryons de Souris portant une délétion maternelle de la DMR dérivée de la lignée germinale du locus distal 12 présentent de graves défauts développementaux, et meurent avant la fin de la gestation ou peu après la naissance (Lin *et al.*, 2003) (voir Résultats, page 122). Or, cette manipulation n'affecte directement que les gènes soumis à l'empreinte génomique parentale dans le locus distal 12 : ces défauts sont donc dus à la dérégulation de gènes soumis à l'empreinte dans ce locus.

Ces embryons meurent avant la fin de la gestation (la létalité se manifeste à partir du seizième jour de gestation ; la gestation chez la Souris dure 19 à 20 jours) ; dans chaque muscle squelettique, les fibres musculaires sont surabondantes et hypertrophiées à partir du seizième jour, et leur différenciation est retardée ; au dix-neuvième jour, des défauts du cartilage déforment la cage thoracique, et plusieurs régions du squelette sont hypo-ossifiées ; enfin, la morphologie du foie est affectée (Anne Ferguson-Smith, communication personnelle). Ces défauts peuvent être attribués à la surexpression de gènes à expression paternelle (*Dlk1*, connu pour inhiber la différenciation des préadipocytes en adipocytes : Moon *et al.*, 2002 ; le rétrotransposon *Rtl1* ; *Dio3*, qui inactive les hormones thyroïdiennes T3 et T4 : Hernandez *et al.*, 1999), ou à la répression de gènes non-codants à expression maternelle (*Gtl2*, les gènes d'ARN C/D, les gènes de microARN).

De la même manière, les défauts du développement des embryons présentant une disomie uniparentale pour le chromosome 12 sont probablement dus aux dérégulations de l'expression des gènes du locus distal 12, puisqu'aucun autre locus soumis à l'empreinte génomique parentale n'est

connu sur ce chromosome¹. Les embryons à disomie uniparentale paternelle (donc : déficients pour l'expression du gène *Gtl2*, et des gènes d'ARN C/D et de microARN, et qui surexpriment *Dlk1*, *Rtl1* et *Dio3*) présentent exactement les mêmes défauts que les embryons portant une délétion maternelle de la DMR dérivée de la lignée germinale (voir ci-dessus), avec une anomalie supplémentaire : le placenta est hypertrophié le dernier jour de la gestation. Les embryons à disomie uniparentale maternelle (donc : déficients pour l'expression des gènes *Dlk1*, *Rtl1* et *Dio3*, et qui surexpriment *Gtl2* et les gènes d'ARN C/D et de microARN) meurent, pour la plupart, le dernier jour de gestation, et les survivants meurent peu après la naissance ; ils présentent un défaut de croissance net à partir du seizième jour de gestation, et leur placenta est lui-même visiblement réduit à partir du seizième jour ; les fibres musculaires sont réduites (en nombre, et en taille) à partir du seizième jour ; le squelette ne présente pas de défaut visible, à l'exception de l'hypo-ossification de deux os de l'oreille interne (Georgiades *et al.*, 2000).

Plusieurs de ces troubles pourraient être dus à la dérégulation de *Dlk1*, impliqué notamment dans la différenciation des adipocytes. Des embryons déficients pour *Dlk1* présentent un retard de croissance qui pourrait expliquer celui des embryons à disomie maternelle, et des défauts du squelette de la cage thoracique ; la moitié de ces souris meurent peu après la naissance ; à ces défauts s'ajoute, chez les souris adultes survivantes, une hypertrophie des adipocytes, et une hyperadiposité générale (Moon *et al.*, 2002). Cependant, tous les défauts des souris à disomie uniparentale maternelle ne peuvent pas être attribués à la répression de *Dlk1*, puisque la moitié des souris *Dlk1*^{-/-} atteignent l'âge adulte, alors que la moitié des embryons à disomie uniparentale maternelle meurent avant la naissance, et l'autre moitié, immédiatement après (Georgiades *et al.*, 2000). Quant à l'expression biallélique de *Dlk1* (simultanée à l'expression biallélique d'une forme d'épissage de *Gtl2*), suite à un défaut naturel d'établissement de l'empreinte du locus chez la Souris, elle n'a aucun effet sur la viabilité des embryons : l'expression biallélique de *Dlk1* n'est donc pas responsable des défauts de viabilité des embryons présentant une disomie uniparentale paternelle (Croteau *et al.*, 2003).

Les défauts développementaux des embryons à disomie uniparentale pour le chromosome 12 qui ne peuvent pas être attribués à la dérégulation de *Dlk1*, pourraient être dus à la dérégulation des gènes d'ARN C/D, ou plus probablement, des gènes de microARN (ces ARN C/D et microARN sont tous exprimés au cours du développement embryonnaire). En effet, de manière générale, les microARN pourraient jouer un rôle central dans les divisions asymétriques, base cellulaire de nombreux processus développementaux : en neutralisant les ARN_m hérités de la cellule-mère non différenciée (par RNAi ou répression traductionnelle), ils permettraient aux cellules-filles de s'engager dans des voies de différenciation (Rhoades *et al.*, 2002 ; Bartel et Chen, 2004). Cette hypothèse a le mérite d'expliquer l'observation surprenante selon laquelle toutes les fonctions démontrées pour des microARN ont trait au contrôle du développement (voir Introduction, partie 1.4.4, page 75) : la fonction générique des microARN pourrait être de curer chaque cellule d'ARN_m spécifiques, hérités de la cellule-mère, et d'établir un contexte de régulation post-transcriptionnelle qui influencerait l'expression des gènes, et donc, participerait à la définition de l'identité cellulaire (Bartel et Chen, 2004 ; Hobert, 2004) (voir Introduction, page 78). Les gènes de microARN du locus distal 12 pourraient ainsi jouer un rôle dans la différenciation des cellules musculaires, un processus affecté par les disomies uniparentales du chromosome 12, mais apparemment pas par l'inactivation ou la surexpression de *Dlk1* (voir plus haut).

¹De plus, les phénotypes des embryons à disomie uniparentale paternelle (et la disomie uniparentale affecterait tous les gènes soumis à l'empreinte sur le chromosome 12), et celui des embryons qui portent une délétion maternelle de la DMR dérivée de la lignée germinale (laquelle n'affecte *a priori* que les gènes du locus distal 12), sont très similaires ; seul le phénotype du placenta diffère (voir texte). Il est donc probable que le chromosome 12 ne contienne pas d'autre locus soumis à l'empreinte dont la dérégulation perturbe le développement, et la différence de phénotype du placenta semble due à un contrôle particulier du locus dans le placenta, où la DMR ne jouerait pas le même rôle que dans l'embryon (S.-P. Lin *et al.*, manuscrit en préparation).

3.2.3 Des ARN-cibles ?

Si les gènes d'ARN C/D ou de microARN du locus ont une fonction (en particulier dans le contrôle du développement, *cf* ci-dessus), leur mode d'action, au niveau moléculaire, pourrait impliquer l'appariement de ces petits ARN avec des ARN-cibles, comme tous les ARN C/D et microARN dont le rôle moléculaire est documenté ; dans ce cas, la compréhension de la fonction biologique de ces nouveaux gènes passera par l'identification de leurs ARN-cibles.

Conservation de la séquence nucléotidique

Un trait remarquable des gènes d'ARN C/D du locus 14q32 et de ses orthologues, est la grande divergence de séquence entre des gènes visiblement apparentés, dans une même espèce, et entre espèces différentes (Cavaillé *et al.*, 2002) (article n°1) ; dans une moindre mesure, les gènes de microARN sont également divergents (les séquences humaine et murine de miR-329/miR-A2 diffèrent de 4 nucléotides ; celles de miR-300/miR-A9 diffèrent de 8 nucléotides ; celles de miR-376b/miR-B4 diffèrent de 2 nucléotides ; celles de miR-376/miR-B6 diffèrent de 1 nucléotide), alors que la plupart des microARN sont absolument conservés entre l'Homme et la Souris. Ces gènes semblent de plus restreints aux Mammifères.

Il est frappant de constater que ces caractéristiques, relativement nouvelles pour des microARN, sont partagées par de nombreux microARN exprimés dans les cellules-souches embryonnaires de Souris, et dont certains pourraient être cotranscrits sous la forme de précurseurs communs (Houbaviy *et al.*, 2003).

Les microARN que nous avons décrits semblent donc faire partie d'une classe de molécules spécifiques des Mammifères. Ils ont pu passer inaperçus jusqu'à présent en raison, justement, de leur faible conservation : il avait été décrété que les gènes de microARN sont conservés entre différentes espèces, qui pouvaient être éloignées (en particulier, les gènes de microARN de Mammifères devaient être conservés chez les Poissons), et les recherches bio-informatiques de gènes de microARN ont utilisé ce critère pour sélectionner les séquences candidates (Ambros *et al.*, 2003 ; Grad *et al.*, 2003 ; Lai *et al.*, 2003 ; Lim *et al.*, 2003a ; Lim *et al.*, 2003b) ; certains auteurs avaient ainsi prédit, par extrapolation à partir de l'erreur d'estimation de leur programme sur les prédictions de gènes de microARN déjà connus, que les génomes de Vertébrés ne contenaient pas plus de 255 gènes de microARN (Lim *et al.*, 2003a).

Après la découverte de nombreux gènes de microARN propres aux Mammifères, il faut sans doute rectifier cette estimation : les génomes de Vertébrés ne contiennent probablement pas tous le même nombre de gènes de microARN, et un nombre inconnu de gènes de microARN spécifiques d'espèces s'additionnent aux ≈ 250 gènes conservés prédits. Les gènes de microARN spécifiques des Mammifères pourraient jouer un rôle particulier dans l'embryogenèse, puisque les microARN que nous avons décrits sont exprimés dans l'embryon (Seitz *et al.*, 2004), et que la plupart des autres sont exprimés principalement dans les cellules-souches embryonnaires non différenciées (Houbaviy *et al.*, 2003 ; Suh *et al.*, 2004).

Le manque de conservation des séquences de ces petits ARN pourrait donner des indications sur leur fonction. Par exemple, si les microARN guident l'hétérochromatinisation de leur locus en s'appariant à l'ADN de leurs gènes, ou à des transcrits naissants complémentaires (voir page 154), des mutations pourraient s'accumuler dans la séquence du microARN sans que sa fonction en soit affectée : puisque le microARN sera toujours homologue à son propre gène, il pourrait échapper, dans une certaine mesure, à la pression qui impose la conservation des séquences fonctionnelles (il serait toutefois sensible aux mutations qui perturberaient excessivement la structure de son précurseur en tige-boucle, et qui empêcheraient sa maturation).

Si ces petits ARN s'appariaient à des ARN-cibles (pour en guider la 2'-O-méthylation pour les ARN C/D, ou pour en diriger le clivage ou la répression traductionnelle, pour les microARN), leur

faible conservation pourrait s'expliquer par une redondance fonctionnelle, entre les nombreuses copies de ces gènes : si un nombre suffisant de gènes garde la capacité de s'apparier à sa cible, les autres échapperont à la pression de sélection. En particulier, les séquences responsables de leur spécificité de cible (le segment situé en amont de la boîte D ou D', pour les ARN C/D ; l'extrémité 5', pour les microARN) ne sont pas plus conservées que le reste de leurs séquences (voir articles n°1 et 4).

Dans une espèce donnée, les différentes copies de ces gènes, relativement divergentes, pourraient alors avoir acquis de nouvelles fonctions, et réguler de nouveaux ARN-cibles.

Les nouveaux ARN C/D : des guides de méthylation ?

Les ARN C/D exprimés à partir des gènes du locus distal 12 semblent nucléolaires (du moins, ceux dont la localisation a été analysée ; voir page 115). Ils s'accumulent donc, comme tous les snoRNA impliqués dans la maturation des ARN_r, au voisinage des sites de transcription des ARN_{pré-r}, ce qui pourrait signifier qu'ils participent eux-mêmes à la maturation des ARN_r. Cette perspective est très intéressante, puisque, chez l'adulte, ces ARN C/D ne semblent exprimés que dans le cerveau : ils pourraient alors se rendre responsables de maturations des ARN_r qui soient spécifiques du cerveau ; le ribosome du cerveau serait différent du ribosome des autres tissus, ce qui pourrait avoir de nombreuses conséquences fonctionnelles. Ces maturations pourraient par exemple jouer un rôle dans la traduction localisée au voisinage des synapses : chez les Vertébrés, des polyribosomes et des ARN_m spécifiques se localisent dans les dendrites de nombreux neurones post-synaptiques, et la synthèse protéique qui s'y déroule (sous le contrôle de stimuli donnés) modifie la sensibilité de la synapse, et pourrait contribuer à la plasticité synaptique (voir Steward et Schuman, 2003, pour revue). De la même manière, si ces nouveaux ARN C/D participaient à la maturation des snRNA, impliqués dans l'épissage, leur expression spécifique du cerveau pourrait jouer un rôle dans un contrôle de l'épissage spécifique de tissu.

Cependant, ces nouveaux ARN C/D ne semblent pas présenter de complémentarités fonctionnelles à l'ARN_{pré-r} ou aux snRNA en amont de leurs boîtes D et D' : les complémentarités identifiées sont courtes, et statistiquement attendues (*cf* Annexe 2, page 173). Ils ne semblent donc pas en mesure de guider de 2'-O-méthylation sur l'ARN_{pré-r} ou les snRNA. Plus généralement, ils sont dépourvus de longue complémentarité (à une position quelconque sur l'ARN C/D) à ces ARN, donc ne semblent pas pouvoir influencer leur repliement (voir Annexe 2, page 173).

Ces ARN C/D pourraient aussi guider la modification post-transcriptionnelle d'autres ARN cellulaires : l'exemple du snoRNA MBII-52/HBII-52 (dont le gène est porté par un autre locus soumis à l'empreinte, le locus 15q11-13 : voir page 53) frappe l'imagination. Ce snoRNA présente une complémentarité de 18 nt à un ARN_m spécifique du cerveau, l'ARN_m du récepteur 5-HT_{2C} à la sérotonine, et cette complémentarité se trouve immédiatement en amont de la boîte D du snoRNA MBII-52/HBII-52 ; le site prédit de 2'-O-méthylation est un nucléotide connu pour être édité, et des données *in vitro* indiquent que la 2'-O-méthylation inhibe l'édition catalysée par l'une des deux enzymes responsables de l'édition de ce site, ADAR-2 (Yi-Brunozzi *et al.*, 1999) (voir page 57). *Ex vivo*, une construction contenant ce site d'édition, transcrite dans le nucléole par l'ARN-polymérase I, est éditée par ADAR-2, et cette édition est inhibée par un guide de méthylation dirigé contre le site d'édition (P. Vitali, communication personnelle). Ainsi, des ARN_m pourraient être modifiés post-transcriptionnellement dans le nucléole ; cette hypothèse s'accorderait bien avec l'observation de la localisation nucléolaire de plusieurs ARN_m (voir Pederson, 1998, pour une revue).

L'identification d'ARN 2'-O-méthylés par l'action de ces nouveaux ARN C/D est plus difficile qu'il n'y paraît : une complémentarité de 12 nt entre la séquence de l'ARN C/D immédiatement en amont de la boîte D, et l'ARN-cible, suffit à guider une méthylation totale de la cible (Cavaillé et Bachellerie, 1998). Or, le chromosome X humain (dont la séquence ne représente que 4 % de

la séquence totale du génome humain) présente 8 séquences parfaitement complémentaires aux 12 nt immédiatement en amont de la boîte D de MBII-52/HBII-52 (dont la séquence de 5-HT_{2C}, qui est portée par ce chromosome) (résultats non montrés). D'autres informations (pas toutes disponibles dans les banques de données), telles que le patron d'expression des séquences transcrites (à comparer au patron d'expression des ARN C/D), sont nécessaires pour identifier un nombre raisonnable d'ARN-cibles candidats ; la démonstration expérimentale de la 2'-O-méthylation *in vivo*, quant à elle, n'est pas chose aisée non plus, puisque les protocoles de détection des 2'-O-méthylations d'ARN, développés pour l'étude de la méthylation des ARN ribosomiques (Maden, 1990 ; Maden *et al.*, 1995), sont peu adaptés à l'analyse d'ARN_m faiblement exprimés (résultats non montrés).

Outre ces rôles de guidage de la 2'-O-méthylation ou d'aide au repliement, les ARN C/D que nous avons décrits pourraient jouer un rôle original. Il est en effet remarquable que la tige terminale de la plupart de ces ARN C/D, et de ceux des gènes du locus 15q11-13, sont très semblables : elles se rapprochent du consensus 5' GG(A|G)TC...GG(C|T)CC 3', alors que les séquences des tiges terminales des snoRNA exprimés ubiquitairement n'obéissent pas à ce consensus (Cavaillé *et al.*, 2002). Cette particularité pourrait avoir une signification fonctionnelle : par exemple, une protéine pourrait reconnaître spécifiquement les ARN C/D présentant cette tige terminale, et leur conférer des propriétés inhabituelles. Il reste donc possible que ces nouveaux ARN C/D jouent un rôle biologique, mais sans interagir avec des ARN-cibles.

Les nouveaux microARN : gènes-cibles potentiels

L'identification des ARN-cibles potentiels des microARN exprimés à partir du locus distal 12 pose un problème similaire : hormis miR-127 et miR-136 (voir ci-dessous), aucun de ces microARN ne présente de complémentarité parfaite à un ARN_m (ils ne semblent donc pas guider de dégradation par RNAi).

Des données génétiques, chez le Mouton, suggèrent qu'un (ou : des) gène(s) à expression maternelle, dans l'orthologue ovin du locus humain 14q32, réprime(nt) un (ou : des) gène(s) à expression paternelle, du même locus (voir page 140). Le gène *DLK1* pourrait être ce gène à expression paternelle, puisqu'il est connu pour participer à des processus développementaux (Georges *et al.*, 2003). Les microARN du locus ovin seraient de bons candidats pour réprimer *DLK1* ; cependant, chez la Souris, aucun ne présente de complémentarité parfaite à l'ARN_m *Dlk1* (ils semblent donc incapables d'en guider la dégradation par RNAi), et les programmes de recherche de complémentarités fonctionnelles pour la répression traductionnelle ne mettent pas en évidence d'interaction privilégiée entre les microARN du locus, et l'ARN_m *Dlk1* (voir Résultats, page 140). Ces programmes ne sont probablement pas encore parfaitement efficaces, et il est possible que l'interaction en *trans* entre gènes du locus soit propre au Mouton ; cependant, il convient aussi de ne pas négliger un autre gène à expression paternelle, *RTL1/PEG11*, que les microARN (à expression maternelle) miR-127 et miR-136 semblent capables de réprimer post-transcriptionnellement, par RNAi (voir articles n°2 et 3).

Bien qu'il ne soit pas mobile (les génomes de Mammifères ne contiennent pas d'autre séquence identique), *Rtl1* semble être un rétrotransposon, avec lesquels sa séquence l'apparente clairement. Son immobilité pourrait signifier qu'un système de défense du génome l'a maîtrisé, au moins dans la lignée germinale. Cette répression de *Rtl1* pourrait faire intervenir une dégradation par RNAi, guidé par miR-127 ou miR-136. Il semble en effet que le RNAi réprime les éléments génétiques mobiles dans de nombreuses espèces (Schramke et Allshire, 2003 ; Sijen et Plasterk, 2003 ; Chicas *et al.*, 2004 ; Shi *et al.*, 2004 ; Svoboda *et al.*, 2004a), et cette tendance pourrait s'appliquer à *Rtl1*.

En première analyse, il pourrait sembler paradoxal qu'un rétrotransposon contienne les séquences qui le répriment (puisque les gènes de miR-127 et miR-136, transcrits en antisens

de *Rtl1*, en sont indissociables) ; cependant, la répression d'un rétrotransposon, probablement bénéfique au génome-hôte, l'est donc aussi pour son parasite : un rétrotransposon actif, en menaçant l'intégrité de son génome-hôte, risque de disparaître avec lui. Un rétrotransposon réprimé, immobilisé, ne peut plus se multiplier, mais il peut sans doute se maintenir plus sûrement qu'un rétrotransposon actif. De ce point de vue, *Rtl1* serait tombé dans un piège évolutif (incapable de rétrotransposer, il pourrait accumuler des mutations sur son unique copie), mais ce piège est aussi une protection.

Il est intéressant de remarquer que la phase ouverte de lecture de *Rtl1* est conservée chez les différents Mammifères où il a été étudié. Il est donc possible que *Rtl1* ait acquis une fonction bénéfique à son génome-hôte, et qu'il remplisse désormais une fonction déterminée dans la biologie des Mammifères (Lynch et Tristem, 2003).

3.2.4 Gènes de petits ARN répétés et empreinte génomique parentale : une association étonnante

Les deux loci contenant des gènes d'ARN C/D répétés sont soumis à l'empreinte génomique parentale ; étant donné que l'empreinte est un phénomène rare (on estime qu'elle affecte moins de 1 % des gènes de Mammifères : voir page 24), et qu'aucun autre locus humain ne semble contenir de gènes d'ARN C/D répétés (voir Résultats, page 115), la coïncidence est troublante. De la même manière, aucun autre locus ne semble aussi riche en gènes de microARN que le locus 14q32 (Seitz *et al.*, 2004). De surcroît, les gènes d'ARN C/D du locus 15q11-13 (de los Santos *et al.*, 2000 ; Cavaillé *et al.*, 2001), et au moins les gènes de microARN du locus 14q32 (résultats non montrés) semblent propres aux Mammifères placentaires (les petits ARN ne sont pas détectés dans le cerveau d'autres espèces, et leurs gènes sont absents des génomes du Poisson *Fugu rubripes* et du Poulet, et du génome de plusieurs Invertébrés) : la présence de ces gènes répétés de petits ARN corrèle donc avec l'empreinte génomique parentale². Ces observations surprenantes s'expliqueraient si la nature des loci soumis à l'empreinte génomique parentale favorisait la multiplication de ce type de gènes (la répétition des gènes de petits ARN serait alors une conséquence de l'empreinte de ces loci), ou si la répétition des gènes de petits ARN provoquait l'empreinte de leur locus (la répétition des gènes de petits ARN serait alors une cause de l'empreinte de ces loci).

Dans la première hypothèse, une caractéristique particulière des loci soumis à l'empreinte génomique parentale aurait permis la multiplication d'un nombre initialement réduit de gènes d'ARN C/D (dans les loci 15q11-13 et 14q32) et de microARN (dans le locus 14q32). Par exemple, puisqu'il est connu que les loci soumis à l'empreinte recombinent plus fréquemment que les autres (Lercher et Hurst, 2003), la multiplication de ces gènes pourrait résulter de recombinaisons multiples entre les chromosomes homologues : cette hypothèse est présentée en figure 3.1. Si l'augmentation du nombre de copies de ces gènes conférait un avantage sélectif (ainsi, s'ils ont acquis une fonction régulatrice — par exemple au cours du développement — et que leur surexpression présentait un avantage), l'allèle le plus riche en copies de gènes de petits ARN serait avantagé, après chaque événement de recombinaison. L'empreinte génomique parentale des loci 15q11-13 et 14q32 aurait aussi pu favoriser l'apparition initiale des gènes de petits ARN dans ces loci : le locus 15q11-13 semble être un point chaud d'intégration des rétrogènes, puisque son orthologue murin a intégré très récemment (notamment, depuis la divergence entre les genres *Mus* et *Rattus*) plusieurs rétrogènes (Chai *et al.*, 2001). Si cette particularité était reliée à l'empreinte génomique parentale (par exemple, si elle dépend de la structure chromatinienne du locus lors de la gamétogenèse), elle pourrait s'appliquer également au locus 14q32.

²À l'exception des Marsupiaux, apparemment dépourvus des gènes de petits ARN répétés, mais chez lesquels des cas de gènes soumis à l'empreinte ont été rapportés ; cependant, on ne sait pas si ces deux loci sont aussi soumis à l'empreinte chez les Marsupiaux (voir page 162).

Or, l'un des gènes d'ARN C/D du locus 14q32 et de ses orthologues, le gène RBII-36 (spécifique du Rat) possède de nombreux rétrogènes, dispersés dans le génome (J. Cavaillé, communication personnelle) : l'un des rétrogènes de RBII-36 aurait ainsi pu s'intégrer, par hasard, dans le locus 6q31-32 du Rat, et l'empreinte de ce locus lui aurait alors permis de se multiplier.

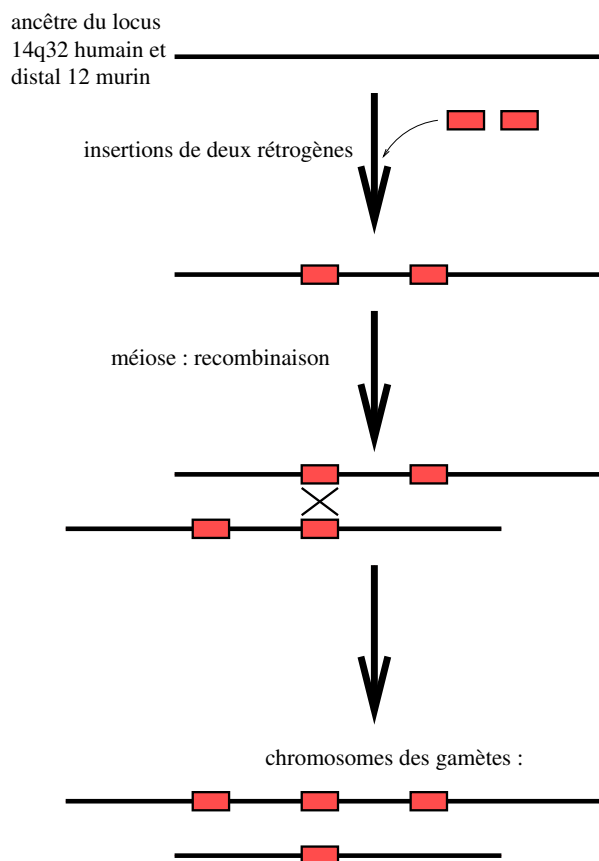


FIG. 3.1 – MODÈLE D'INVASION DU LOCUS PAR DES RÉTROGÈNES. L'insertion successive de deux rétrogènes identiques (représentés par des rectangles rouges) dans le locus crée un substrat propre à s'amplifier par recombinaison : après recombinaison, l'un des allèles s'est enrichi en copies du rétrogène.

Dans la deuxième hypothèse, la répétition des gènes de petits ARN (par un mécanisme inconnu) dans des loci initialement non soumis à l'empreinte génomique parentale, aurait provoqué l'empreinte du locus. En effet, les loci soumis à l'empreinte sont riches en séquences répétées et en rétrogènes, et il a été proposé que l'apparition, dans un locus donné, de telles séquences pouvait aboutir à son empreinte, comme une conséquence d'une extinction incomplète de séquences traitées comme des parasites génomiques (voir Introduction, page 38) : la répression de ces gènes, par des modifications épigénétiques apposées lors de la gamétogenèse, ne serait efficace que dans une des deux lignées germinales. Elle s'étendrait aux gènes avoisinants, forçant ainsi le mono-allélisme de leur expression. Dans l'exemple du locus 15q11-13, la multiplication de gènes d'ARN C/D serait alors la cause de l'empreinte du locus ; dans le cas du locus 14q32, des gènes d'ARN C/D et des gènes de microARN se seraient multipliés simultanément, et ces deux familles de répétitions seraient responsables de l'empreinte de leur locus.

La principale difficulté de ce modèle est qu'il explique mal l'existence de gènes à expression paternelle dans le locus 14q32 : si une extinction inefficace des gènes de petits ARN non-codants (restreinte à l'allèle paternel) s'étend aux séquences du voisinage (*Gtl2* dans ce cas, s'il n'est

pas cotranscrit avec les petits ARN), pourquoi *Dlk1*, *Rtl1* et *Dio3* seraient-ils réprimés sur l'allèle maternel ? Cette difficulté se résout si les gènes à expression maternelle ont une activité de répression des séquences voisines, en *cis*. En d'autres termes, si ces gènes ont un rôle dans le mécanisme de l'empreinte de *Dlk1*, *Rtl1* et *Dio3*. Puisque cet effet doit être exclusivement en *cis* (de manière à n'affecter que l'allèle maternel de ces trois gènes), il est probablement transcriptionnel. Chacune des deux familles de petits ARN, ARN C/D et microARN, ainsi que les ARN constitués des exons non-codants de leurs gènes-hôtes, raboutés après épissage, pourraient éventuellement jouer ce rôle :

- Les ARN non-codants des gènes-hôtes (si les précurseurs des petits ARN étaient cotranscrits, il n'y aurait qu'un unique gène-hôte) pourraient avoir des propriétés de répresseurs en *cis*, comme cela semble être le cas de l'ARN Air, nécessaire à la répression de l'allèle paternel de trois gènes de son locus (Sleutels *et al.*, 2002 ; Sleutels *et al.*, 2003). Le mécanisme de cette répression n'est pas encore connu. Ainsi, chez le Rat, l'ARN du gène-hôte des ARN C/D (appelé *Bsr*) se localise exclusivement dans le noyau (Cavaillé *et al.*, 2001), et en particulier, sur le locus de son gène (É. Royo, résultats non publiés). De plus, dans l'autre locus contenant des gènes d'ARN C/D répétés (le locus humain 15q11-13 et ses orthologues), l'extrémité distale de l'ARN du gène-hôte des ARN C/D est antisens à l'ARN_m *Ube3a*, dont le gène est soumis à une empreinte réciproque à celle des gènes d'ARN C/D (voir Introduction, page 36).
- Aucun exemple de répression transcriptionnelle par des ARN C/D n'est connu : si les ARN C/D exprimés à partir de l'allèle maternel du locus 14q32 doivent réprimer la transcription de trois gènes du voisinage, il faut que ce soit par une propriété que ne partagent pas les ARN C/D dont les gènes ne sont pas soumis à l'empreinte. Cette propriété pourrait être la répétition de leurs gènes : seuls les gènes d'ARN C/D répétés pourraient alors réprimer la transcription en *cis*. La répétition de ces gènes, transcrits en multiples copies de précurseurs d'ARN C/D, pourrait aboutir à une concentration locale élevée en protéines impliquées dans la maturation des ARN C/D ; parmi elles, les protéines Tip48 et Tip49 (voir page 47), également connues sous les noms de Tip49b, RUVBL2, Rvb2, Tih2p, p50, TAP54 β et Reptin52 (pour Tip48), et Tip49a, RUVBL1, NMP238, Rvb1, TAP54 α , Tih1p, p55, TAP54 et Pontin52 (pour Tip49), participent à plusieurs complexes de modification de la chromatine (référencés dans Ogura et Wilkinson, 2001). Tip49 interagit avec la matrice nucléaire (Holzmann *et al.*, 1998). Ainsi, si l'association des précurseurs d'ARN C/D à ces deux protéines ne masque pas leurs sites d'interaction avec les complexes de remodelage de la chromatine, ou avec la matrice nucléaire, la transcription de nombreux gènes d'ARN C/D à partir de l'allèle maternel pourrait recruter sur cet allèle des activités de modification de la chromatine, ou associer spécifiquement cet allèle à la matrice nucléaire, ce qui pourrait réprimer la transcription de *Dlk1*, *Rtl1* et *Dio3* sur cet allèle (un exemple d'association à la matrice nucléaire spécifique d'un allèle d'un locus soumis à l'empreinte génomique parentale est déjà connu : Weber *et al.*, 2003).
- Les microARN ne sont pas connus pour jouer un rôle dans la modification de la chromatine ; cependant, des siRNA issus de la digestion de longs ARN doubles brins (bimoléculaires ou monomoléculaires) semblent impliqués dans le guidage de l'hétérochromatinisation chez de nombreuses espèces (voir Introduction, page 80) : des protéines requises pour la production des siRNA ou leur fonction sont nécessaires à la mise en place ou au maintien de l'hétérochromatine de loci spécifiques chez la levure à fission *Schizosaccharomyces pombe* (et de petits ARN homologues des séquences hétérochromatinisées, apparemment des siRNA, sont détectés) (Volpe *et al.*, 2002 ; Hall *et al.*, 2002 ; Schramke et Allshire, 2003), chez la Crucifère *Arabidopsis thaliana* (Chan *et al.*, 2004) et chez la Drosophile (Pal-Bhadra *et al.*, 2004), et des ARN présentant les caractéristiques des siRNA pourraient jouer un rôle dans l'hétérochromatinisation de séquences homologues chez un Cilié (Liu *et al.*, 2004c), et

chez des Vertébrés (Fukagawa *et al.*, 2004 ; Kawasaki et Taira, 2004 ; Morris *et al.*, 2004). La spécificité de séquence de ces processus d'hétérochromatinisation repose certainement sur l'appariement entre les siRNA et un acide nucléique complémentaire ; on ne sait pas encore si cet acide nucléique est de l'ADN (le siRNA n'y aurait alors accès qu'à condition que la double hélice d'ADN soit ouverte), ou un ARN, qui pourrait être un transcrit naissant, donc encore à proximité de la chromatine de son gène. Les microARN, dont la biosynthèse s'apparente à celle des siRNA, et qui s'associent à plusieurs partenaires protéiques des siRNA (voir Introduction, page 62), pourraient jouer un rôle semblable. Si les microARN des gènes du locus 14q32 étaient responsables de l'empreinte de *Dlk1*, *Rtl1* et *Dio3*, il faudrait que leur effet se restreigne à l'allèle maternel, à proximité de leur site de transcription, alors que les effets des siRNA sur la chromatine chez les Vertébrés sont des effets *trans* (Kawasaki et Taira, 2004 ; Morris *et al.*, 2004) ; un modèle impliquant les microARN dans l'hétérochromatinisation de l'allèle maternel du locus devra donc expliquer cette spécificité (voir ci-dessous).

Les données disponibles sur la répression transcriptionnelle guidée par des siRNA chez les Mammifères sont très récentes, et semblent encore contradictoires : alors que la transcription d'un ARN qui se replie en tige-boucle, dans les ovocytes de Souris, ne guide pas de méthylation de l'ADN d'une séquence homologue en *trans* (Svoboda *et al.*, 2004b), l'introduction de siRNA synthétiques dans des cellules humaines 293FT en culture provoque la méthylation d'une séquence homologue aux siRNA à condition que le noyau des cellules soit perméabilisé (Morris *et al.*, 2004). En revanche, la transfection de cellules épithéliales humaines, saines ou tumorales, par des siRNA homologues de séquences d'un promoteur aboutit à la méthylation des cytosines des dinucléotides CpG et des trinucéotides CpNpG (appelées « cytosines symétriques » ; les cytosines asymétriques n'ont pas été analysées) des séquences ciblées (Kawasaki et Taira, 2004) ; la méthylation peut se propager aux cytosines symétriques du voisinage (des cytosines situées à une centaine de pb du site de complémentarité sont méthylées, mais pas des cytosines situées à environ 200 pb). Il est possible que les types cellulaires utilisés dans ces trois études ne sont pas également compétents pour la méthylation d'ADN dirigée par les siRNA ; par exemple, les siRNA pourraient guider des modifications post-traductionnelles des histones, lesquelles aboutiraient ou non, selon le type cellulaire, à la méthylation de l'ADN qui leur est associé ; ou encore, puisque l'une des études montre une dépendance à la perméabilisation de l'enveloppe nucléaire, il est possible que les petits ARN soient exclus du noyau des ovocytes de Souris. Étant donnée leur petite taille, les siRNA devraient traverser les pores nucléaires par simple diffusion ; cependant, le système d'export des pre-miRNA, par l'exportine 5 (voir Introduction, page 65) (Yi *et al.*, 2003 ; Lund *et al.*, 2004), pourrait chasser activement du noyau les siRNA qui y sont entrés, puisque cette exportine présente une grande affinité pour les produits de Dicer (Gwizdek *et al.*, 2003 ; Lund *et al.*, 2004). En outre, les siRNA, même transfectés, s'associent probablement rapidement aux protéines du complexe RISC, pour former une particule de grande taille (Pham *et al.*, 2004) (voir aussi Introduction, page 71) ; la taille de ce complexe (80S, une taille comparable à celle du ribosome) lui interdirait alors de diffuser jusqu'à la chromatine.

Dans le cas qui nous intéresse, si les microARN des gènes du locus 14q32 pouvaient guider l'hétérochromatinisation de séquences en *cis* sur l'allèle maternel, le phénomène aurait probablement lieu dans les ovocytes, où l'empreinte maternelle se met en place (voir Introduction, page 26) ; dans ce cas, l'étude menée sur les ovocytes de Souris (Svoboda *et al.*, 2004b) est probablement la plus pertinente pour notre sujet. Ce travail a montré que la transcription d'ARN qui se replie en tige-boucle (le segment bicaténaire est long d'environ 500 pb : Svoboda *et al.*, 2001 ; ce n'est donc pas un pre-miRNA, plutôt un précurseur de siRNA) ne provoque pas la méthylation de l'ADN portant une séquence homologue, en *trans*. Les auteurs n'ont pas pu analyser la méthylation du gène de cet ARN bicaténaire, pour des raisons techniques (Svoboda *et al.*, 2004b). Puisque les cellules de Mammifères semblent avoir la propriété de méthylar les

séquences d'ADN complémentaires des siRNA (Kawasaki et Taira, 2004 ; Morris *et al.*, 2004), les observations réalisées sur l'ovocyte de Souris pourraient signifier que, alors que ces cellules en ont la capacité intrinsèque, elles ne méthylent pas les séquences d'ADN homologues des siRNA pour une raison qui n'est pas forcément en rapport avec la machinerie de méthylation de l'ADN : par exemple, les siRNA pourraient être exportés du noyau des ovocytes de Souris avec une très grande efficacité, et n'auraient pas le temps de guider la méthylation de séquences homologues en *trans*. Puisque la production de siRNA à partir d'ARN repliés en longue tige-boucle peut se dérouler intégralement dans le noyau (Shinagawa et Ishii, 2003), allonger le temps de résidence des siRNA dans le noyau (par exemple, en inhibant leur export) pourrait alors leur permettre de guider la méthylation de séquences en *trans*. Dans le cadre de cette hypothèse, les siRNA pourraient peut-être guider une méthylation d'ADN en *cis* dans des conditions physiologiques : transcrits à proximité de leur cible, ils pourraient l'atteindre rapidement, et en diriger la méthylation avant d'être exportés ; des précurseurs de microARN, transcrits dans les ovocytes, pourraient eux-mêmes être maturés intégralement dans le noyau, et guider la méthylation de leur propre gène. Cette méthylation pourrait alors se propager aux séquences avoisinantes, notamment les promoteurs de *Dlk1*, *Rtl1* et *Dio3* ; ce processus pourrait être d'autant plus efficace que les gènes de microARN sont répétés, et la transcription des gènes de microARN dispersés dans le génome n'aurait donc pas d'effet notable sur l'expression des gènes voisins.

Ce modèle pourrait ainsi expliquer l'empreinte des gènes *Dlk1*, *Rtl1* et *Dio3* par l'expression monoallélique des gènes de microARN. Il faudrait toutefois qu'il explique, en plus, pourquoi le(s) promoteur(s) des gènes à expression maternelle (*Gtl2*, les gènes d'ARN C/D, et les gènes des microARN eux-mêmes) échappe(nt) à cette hétérochromatinisation : il faudrait qu'il(s) soi(en)t protégé(s), d'une manière ou d'une autre, contre l'effet des microARN. Si un facteur spécifique de l'ovogenèse protège la DMR dérivée de la lignée germinale contre la méthylation (absent des spermatogonies, il serait responsable de la différence de méthylation entre les deux gamètes : un tel mécanisme est postulé pour l'établissement de la méthylation différentielle de la DMR dérivée de la lignée germinale du locus *Igf2/H19* : voir Introduction, page 27), il pourrait protéger également les promoteurs des gènes à expression maternelle.

Ces différents modèles sont présentés en figure 3.2, page 157.

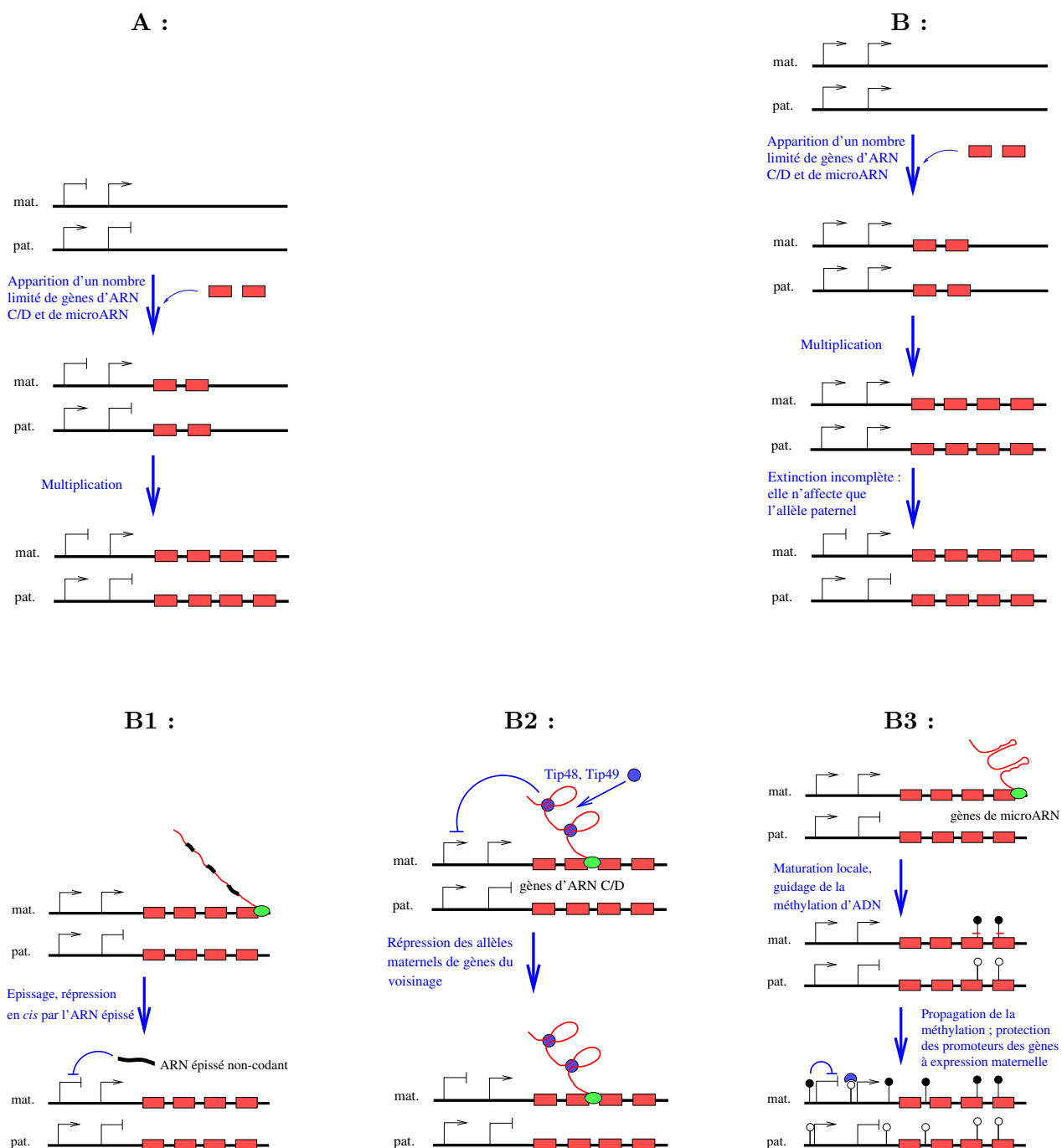


FIG. 3.2 – EMPREINTE GÉNOMIQUE PARENTALE ET RÉPÉTITIONS DE GÈNES DE PETITS ARN NON-CODANTS. « mat. » : chromosome maternel ; « pat. » : chromosome paternel. **A** : L'empreinte génomique parentale du locus préexistait à l'apparition des gènes répétés de petits ARN ; elle aurait pu favoriser leur multiplication (voir figure 3.1). **B** : La multiplication des gènes de petits ARN dans le locus provoque son hétérochromatinisation, comme pour la plupart des séquences répétées du génome. Cette hétérochromatinisation n'est efficace que dans la lignée germinale mâle, d'où l'empreinte des gènes de petits ARN. **B1** : L'ARN non-codant issu du raboutage des exons réprime en *cis* la transcription de gènes du voisinage (tels que *Dlk1*, *Rtl1* et *Dio3*), à la manière de l'ARN Air. **B2** : La transcription de nombreux précurseurs d'ARN C/D recrute des protéines de remodelage de la chromatine (telles Tip48 et Tip49), qui répriment la transcription des gènes du voisinage sur l'allèle maternel. **B3** : Les microARN, maturés à proximité de leur site de transcription, en provoquant la méthylation, qui s'étend ensuite aux séquences adjacentes (les promoteurs des gènes à expression maternelle seraient protégés contre cette méthylation), provoquant la répression transcriptionnelle des gènes du voisinage.

Outre cet effet transcriptionnel potentiel sur les trois gènes à expression paternelle, les petits ARN non-codants du locus pourraient participer à l’empreinte de *Rtl1* au niveau post-transcriptionnel, par RNAi spécifique d’allèle guidé par miR-127 ou miR-136. Ces deux microARN, exprimés à partir de l’allèle maternel uniquement, sont parfaitement complémentaires de l’ARN_m de *Rtl1*, et sont donc susceptibles d’en guider la dégradation par RNAi (voir Résultats, page 118). Dans ce cas, le gène *Rtl1* pourrait être transcrit à partir des deux allèles, mais une dégradation de ses ARN_m issus de la transcription de l’allèle maternel n’autoriserait l’expression que de l’allèle paternel. Il est difficilement concevable qu’une réaction de RNAi *a priori* cytoplasmique (Fire *et al.*, 1998 ; Martinez *et al.*, 2002) puisse distinguer l’origine allélique des ARN_m *Rtl1* ; plus probablement, si miR-127 ou miR-136 dirige le clivage de cet ARN, la réaction se déroule à proximité du locus maternel, dans le noyau.

Deux groupes indépendants ont montré que la réaction de RNAi pouvait se dérouler dans le noyau : chez le Nématode, il est possible de diriger la dégradation d’un ARNpré-m par RNAi³ (Bosher *et al.*, 1999) ; et un extrait nucléaire de cellules humaines HeLa guide le clivage d’ARN-cibles présentant des complémentarités parfaites à deux microARN endogènes, miR-21 et let-7a (Meister *et al.*, 2004).

En plus d’une localisation nucléaire de la réaction de RNAi, l’hypothèse d’un RNAi spécifique d’allèle, dirigé par miR-127 ou miR-136, nécessite aussi que la réaction de clivage ne puisse pas avoir lieu sur l’allèle paternel ; ce serait le cas si le complexe RISC, assemblé autour de miR-127 ou miR-136, ne se forme qu’à proximité de l’allèle maternel, et ne diffuse pas vers l’allèle paternel. Il faudrait alors que ces deux microARN soient absents de l’environnement nucléoplasmique de l’allèle paternel de *Rtl1* : par exemple, ces microARN pourraient être maturés sur leur site de transcription, sur le chromosome maternel, où ils dirigeraient le clivage de l’ARN_m *Rtl1* ; ils seraient retenus dans cette région du noyau, et n’en sortiraient éventuellement que pour être exportés vers le cytoplasme.

Une étude *in vitro* a montré, sur l’exemple du microARN miR-30a, qu’un extrait nucléaire de cellules HeLa ne parvenait pas à maturer efficacement miR-30a, et que l’efficacité en était nettement améliorée par l’ajout d’un extrait cytoplasmique (Lee *et al.*, 2002b) ; cependant, chez la Souris, de longs ARN en tige-boucle peuvent être maturés en siRNA dans le noyau (Shinagawa et Ishii, 2003) ; puisque les voies de maturation des microARN et des siRNA sont similaires (la même enzyme, Dicer, clive les précurseurs bicaténares) (voir Introduction, page 62), le noyau pourrait peut-être contenir tous les facteurs nécessaires à la production des microARN : le seul composant *a priori* cytoplasmique connu pour participer à la biogenèse des microARN est l’enzyme Dicer, or l’expérience de Shinagawa et Ishii montre que, *in vivo*, l’activité de clivage des ARN double brin attribuée à Dicer est également nucléaire. Ce résultat pourrait signifier que, *in vivo*, l’enzyme Dicer peut se localiser dans le noyau ; selon une deuxième interprétation, une enzyme nucléaire (ce pourrait être Drosha qui, comme Dicer, clive les ARN double-brin en libérant des extrémités 3’ sortantes de 2 nt) supplée Dicer.

On le voit, l’implication de miR-127 et miR-136 dans l’empreinte de *Rtl1* par une réaction de RNAi spécifique d’allèle nécessiterait plusieurs aménagements des modèles couramment acceptés, en ce qui concerne la localisation subcellulaire de la réaction de RNAi, et de la maturation des microARN. Cependant, d’autres exceptions à ces modèles sont déjà connues. Si des données génétiques suggèrent que ces microARN déstabilisent les ARN_m issus de la transcription des deux allèles de *Rtl1* (voir Résultats, page 122, et Article n°3, page 124 : en l’absence des microARN, les ARN_m des deux allèles de *Rtl1* sont sur-accumulés, et s’accumulent autant l’un

³Ces résultats sont en contradiction apparente avec ceux de Fire *et al.*, 1998, qui avaient montré que le RNAi était inefficace contre un ARNpré-m. Mais les ARNpré-m ciblés ne sont pas les mêmes, et il est possible qu’un ARNpré-m de courte durée de vie soit insensible au RNAi, alors qu’un ARNpré-m épissé lentement puisse être ciblé (Bosher *et al.*, 1999). L’ARN de *Rtl1*, dépourvu d’intron, pourrait être exporté peu efficacement (chez les Métazoaires, l’export des ARN_m est couplé à leur épissage : voir Reed, 2003, pour revue), donc vulnérable à un RNAi nucléaire.

que l'autre), elles ne sont pas non plus incompatibles avec un RNAi spécifique d'allèle : il est possible qu'une réaction nucléaire, spécifique d'allèle, n'affecte que l'allèle maternel de *Rtl1*, et que simultanément, une réaction cytoplasmique (guidée par des microARN miR-127 et miR-136 exportés du noyau) dégrade indistinctement les ARN_m issus de la transcription des deux allèles de *Rtl1*. Puisque l'ARN_m de l'allèle paternel de *Rtl1* est malgré tout détecté (voir Article n°2), ce RNAi cytoplasmique serait moins efficace que le RNAi nucléaire. La figure 3.3 résume cette situation.

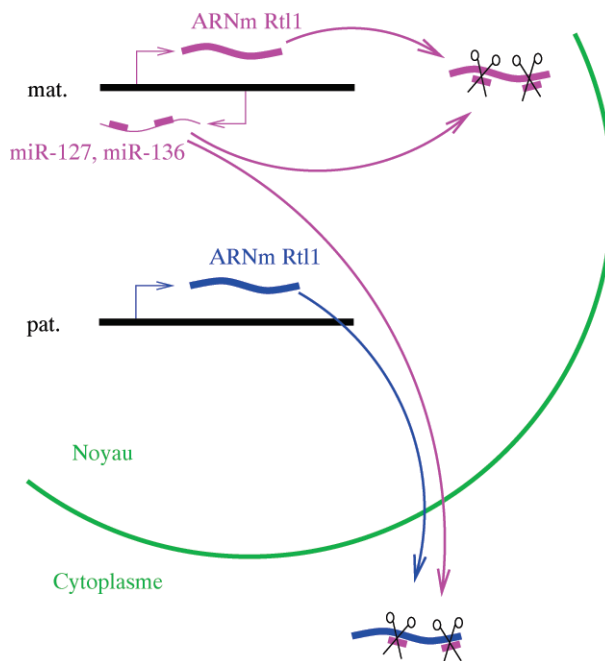


FIG. 3.3 – CONTRIBUTION POST-TRANSCRIPTIONNELLE À L'EMPREINTE ET À LA RÉPRESSION DE *Rtl1* PAR RNAi. « mat. » : chromosome maternel; « pat. » : chromosome paternel. Les microARN miR-127 et miR-136, exprimés uniquement à partir de l'allèle maternel, pourraient guider un clivage de l'ARN_m *Rtl1* à proximité de leur site de transcription, dans le noyau (→ empreinte de *Rtl1*), et également (avec une moindre efficacité) dans le cytoplasme (→ répression biallélique de *Rtl1*); voir texte.

3.2.5 Expression des ARN non-codants du locus 14q32

Tous les ARN non-codants dont les gènes sont portés par le locus 14q32 (Gtl2, les ARN C/D, et les microARN) sont transcrits à partir du même chromosome, le chromosome maternel, dans la même orientation, sous le contrôle d'un même élément régulateur, la DMR dérivée de la lignée germinale, et ils semblent partager le même patron d'expression chez la Souris : chez l'adulte, tous semblent exprimés principalement dans le cerveau (Schmidt *et al.*, 2000; Cavaillé *et al.*, 2002; Seitz *et al.*, 2004), et les ARN C/D et les microARN au moins sont exprimés chez l'embryon (Cavaillé *et al.*, 2002; Seitz *et al.*, 2003; Seitz *et al.*, 2004) (voir articles n°1, 2 et 4). Il est donc possible que tous ces ARN soient co-transcrits. Des produits d'épissage de plusieurs exons encadrant les gènes d'ARN C/D, ou les gènes de microARN (Cavaillé *et al.*, 2002; Seitz *et al.*, 2004), sont en effet détectés : ces exons, donc les petits ARN qu'ils encadrent, sont donc transcrits sous la forme d'un précurseur commun.

Parmi ces ARN non-codants, le seul dont l'extrémité 5' ait été caractérisée est Gtl2; si tous ces ARN non-codants étaient transcrits sous la forme d'un précurseur commun, ce transcrit

primaire serait long d'au moins 197 kb chez la Souris (distance entre l'extrémité 5' de l'ARN Gtl2 et le gène de miR-410/miR-A24, le gène de microARN le plus distal que nous ayons découvert). Dans les banques de données, la plupart des séquences de transcrits Gtl2 ont la même extrémité 3', en amont des gènes d'ARN C/D ; la co-transcription de Gtl2 avec les petits ARN signifierait donc que ce gène possède plusieurs extrémités 3' alternatives.

L'ARN Gtl2 possède plusieurs promoteurs alternatifs, et subit des épissages alternatifs ; la régulation de l'expression de ces différentes isoformes ne semble pas régulée très strictement, puisqu'après le croisement de deux lignées murines différentes, certaines isoformes peuvent être exprimées à partir des deux allèles dans l'embryon ou le placenta (Croteau *et al.*, 2003). Ces souris ne présentent pas de phénotype particulier : les séquences de Gtl2 n'ont donc pas nécessairement de fonction biologique, elles ne constituent peut-être que l'extrémité 5' d'un long transcrit primaire, dont les entités fonctionnelles pourraient être les ARN C/D ou les microARN.

3.2.6 L'ARN, nouveau champ de bataille de la guerre des sexes ?

Les fonctions des ARN C/D et microARN que nous avons décrits restent spéculatives ; quelques observations suggèrent qu'ils pourraient réprimer l'ARN_m de *Rtl1* (voir page 151), et éventuellement l'ARN_m de *Dlk1* (voir page 140). Ces deux ARN_m sont exprimés à partir de l'allèle paternel uniquement, alors que les microARN sont exprimés à partir de l'allèle maternel uniquement : cet antagonisme entre gènes à expression paternelle, et gènes à expression maternelle, est l'une des prédictions de la « théorie du conflit », qui explique l'émergence de l'empreinte génomique parentale par un conflit entre les allèles hérités du père et les allèles hérités de la mère (voir Introduction, page 39). Selon cette théorie, les allèles paternels auront de meilleures chances de se transmettre si l'embryon se porte bien, même au détriment de la mère ou du reste de la portée, alors que les allèles maternels auront de meilleures chances de se transmettre si la gestation n'est pas trop coûteuse à la mère, quitte à ne pas assurer à l'embryon des conditions de croissance optimales. Et effectivement, plusieurs gènes à expression paternelle sont clairement des activateurs de la croissance, et plusieurs gènes à expression maternelle, des répresseurs de la croissance (voir tableau 1.2, page 42) ; en particulier, le produit du gène à expression maternelle *Igf2r* réprime le facteur de croissance *Igf2*, exprimé à partir de son allèle paternel uniquement (Hawkes et Kar, 2004).

S'il était confirmé que les microARN exprimés à partir du locus 14q32 répriment ces ARN_m, ils apporteraient une nouvelle illustration de ce conflit au niveau moléculaire. Et plus précisément, le modèle d'interaction en *trans* entre des gènes du locus, proposé pour expliquer les règles de transmission du phénotype callipyge chez le Mouton (Georges *et al.*, 2003), oppose un gène à expression paternelle (qui favoriserait la croissance musculaire) à un répresseur à expression maternelle (qui, donc, limiterait la croissance musculaire), et les microARN sont les candidats les plus crédibles pour jouer ce rôle de répresseurs à expression maternelle (voir page 140).

Si les microARN à expression maternelle exprimés à partir du locus 14q32 réprimaient la croissance, une pression de sélection pourrait favoriser la multiplication de leurs gènes : l'apparition de nouveaux gènes de microARN, apparentés en séquence (donc : susceptibles de contribuer à la répression du gène-cible) renforcerait la répression de la croissance de l'embryon, au bénéfice probable des allèles maternels ; si une femelle possédait un grand nombre de microARN, ses allèles auraient alors de meilleures chances de se transmettre. Quant au père, le nombre de ses gènes de microARN n'a *a priori* aucune incidence sur la croissance de l'embryon (puisque l'allèle paternel de ces gènes est silencieux) ; en revanche, si l'embryon est femelle, ses allèles (qui sont, pour moitié, ceux de son père) auront de meilleures chances de se transmettre, à la génération suivante, si elle possède de nombreux gènes de microARN. Ainsi, à l'échéance de la deuxième génération, la multiplication des gènes de microARN serait aussi profitable à la transmission des allèles des mâles.

La théorie du conflit prédit donc que, si effectivement les microARN de ce locus répriment la croissance, une pression de sélection tendrait à renforcer leur expression, par exemple, par la multiplication de leurs gènes; cette pression de sélection aurait donc pu contribuer à la multiplication des gènes de microARN dans le locus 14q32.

Il est intéressant de remarquer que la répression traductionnelle par un microARN n'est jamais efficace à 100 %; il faut probablement considérer que de multiples interactions entre un ARN_m et des microARN contribuent chacune partiellement à la répression globale (Bartel et Chen, 2004; Hobert, 2004) : ajouter ou retrancher une de ces interactions n'aurait alors qu'un effet limité sur l'expression du gène-cible. Or, la théorie du conflit propose que les effets des gènes soumis à l'empreinte génomique parentale sur la santé de l'embryon soient subtils : il ne faudrait pas s'attendre à ce qu'un embryon avec une disomie uniparentale paternelle soit nécessairement plus gros et plus sain qu'un embryon sauvage, parce que l'avantage sélectif d'un gène qui favorise la croissance n'est pas le même dans un contexte sauvage (où il est opposé à des répresseurs à expression maternelle), et dans le contexte de la disomie uniparentale (Wilkins et Haig, 2003). Promouvoir la croissance peut être bénéfique dans un contexte où sont exprimés des répresseurs de la croissance, mais détrimental dans un contexte de croissance débridée⁴; les effets des gènes soumis à l'empreinte ne sont donc probablement que de petites impulsions en faveur ou en défaveur de la croissance, autour d'une valeur d'équilibre. Les microARN seraient donc des effecteurs de choix, en régulant finement l'expression de leurs gènes-cibles.

3.3 Perspectives

Les quelques indices dont nous disposons quant à la fonction de ces nombreux gènes non-codants du locus 14q32 humain, et distal 12 murin, n'apportent évidemment aucune preuve formelle, mais nous incitent à orienter les recherches futures dans quelques directions privilégiées.

La délétion des gènes d'ARN C/D, et de microARN, devrait apporter des informations concrètes quant à leurs fonctions. Étant donné que plusieurs sont apparentés en séquence, ils pourraient avoir des rôles redondants, et la méthode la plus efficace consiste probablement à tous les déléter. Notre laboratoire a ainsi entamé un partenariat avec la Clinique de la Souris, à Illkirch, pour insérer des sites de recombinaison *lox* en amont des gènes d'ARN C/D, entre les gènes d'ARN C/D et le cluster R (où se trouvent la plupart des gènes de microARN), et en aval du cluster R. Le croisement de souris dont les gènes d'ARN C/D, ou dont les gènes de microARN, sont flanqués de sites *lox*, avec des souris exprimant la recombinase Cre, permettra de déléter spécifiquement l'une ou l'autre des deux familles de gènes. Il sera également possible de déléter les gènes d'ARN C/D et les gènes de microARN dans une même souris. La délétion du cluster R ne supprimera pas tous les gènes de microARN, puisque quelques-uns (dont *miR-127* et *miR-136*, par exemple) se trouvent en-dehors du cluster R; il faut donc espérer qu'ils ne corrigeront pas l'absence des gènes de microARN du cluster R (desquels ils sont assez différents, par leur séquence nucléotidique).

L'étude de ces souris délétées pour les uns ou les autres des gènes non-codants devrait être très informative. En particulier, il sera intéressant de suivre l'empreinte génomique parentale du locus distal 12 chez les descendants de femelles délétées : si les ARN C/D ou les microARN participent à l'extinction de l'allèle maternel de *Dlk1*, *Rtl1* ou *Dio3*, ces embryons devraient exprimer les deux allèles du (ou : des) gène(s) habituellement réprimé(s) par les petits ARN.

De plus, le croisement des souris portant les insertions de sites *lox*, avec une souche exprimant la recombinase Cre dans le cerveau uniquement, devrait donner des indications sur le rôle de

⁴Ainsi, dans une situation extrême, si les gènes à expression paternelle favorisent trop l'embryon au détriment de la mère, la vulnérabilité de la mère pourra faire échouer la gestation, donc la transmission des allèles paternels, par la mort de la mère; réciproquement, si les gènes à expression maternelle répriment trop la croissance des embryons au point de compromettre leur survie, les allèles maternels auront peu de chances de se transmettre.

ces petits ARN dans le cerveau adulte, en s'affranchissant de la létalité embryonnaire éventuelle de la délétion des gènes de petits ARN. La comparaison de différentes caractéristiques de la traduction dans le cerveau (par l'analyse du profil des gradients de polyribosomes, ou de la localisation subcellulaire de la traduction dans les neurones) pourrait mettre en évidence un rôle des ARN C/D du locus 14q32 dans un aspect particulier du fonctionnement des ribosomes.

La nature précise des précurseurs de ces petits ARN mérite aussi une analyse plus approfondie. La détermination des bornes des séquences transcrites, dans le locus, pourrait révéler les extrémités du (ou : des) ARN précurseur(s) de ces petits ARN. Si tous ces ARN dériveraient d'un unique précurseur commun, le contrôle de l'épissage de ce long ARN affecterait la production des ARN C/D et des microARN, or il est connu que l'ARN Gtl2 subit des épissages alternatifs, contrôlés au cours du développement (Croteau *et al.*, 2003).

D'autre part, la longue séquence (plus de 530 kb chez la Souris) qui sépare le gène de microARN le plus distal, *miR-410/miR-A24*, du gène *Dio3*, est encore mal caractérisée, et pourrait recéler de nouveaux gènes, ou de nouvelles séquences régulatrices. L'analyse de la transcription et de la structure chromatinienne de cette séquence pourrait ouvrir de nouvelles perspectives.

Enfin, ces petits ARN semblent propres aux Mammifères placentaires ; la comparaison de la séquence génomique de ce locus entre Mammifères placentaires et Marsupiaux⁵ pourrait apporter de nouvelles indications quant à la fonction des gènes de ces petits ARN (si, par exemple, il était découvert que ce locus n'est pas soumis à l'empreinte chez les Marsupiaux⁶, la corrélation entre l'expression de ces petits ARN, et l'empreinte de leur locus, serait prolongée), et quant à leur origine évolutive. Ces gènes sont visiblement d'origine récente (ils sont absents du génome du Poulet et du Poisson *Fugu rubripes*), et se sont donc multipliés très rapidement. Si le génome du dernier ancêtre commun aux Mammifères placentaires et aux Marsupiaux représentait une étape particulière de cette évolution (il pouvait contenir un nombre limité de ces gènes, ou bien pas du tout, ou encore, un nombre voisin de celui observé chez les Mammifères placentaires actuels), la confrontation des génomes de Mammifères placentaires et de Marsupiaux pourrait permettre de reconstruire ce génome fossile, et de préciser ainsi la chronologie de l'évolution du locus.

⁵Les génomes de deux Marsupiaux, un Kangourou et un Opossum, sont en cours de séquençage.

⁶L'empreinte génomique parentale est encore mal caractérisée chez les Marsupiaux : on ne connaît que deux gènes soumis à l'empreinte dans ce phylum, *Igf2* (O'Neill *et al.*, 2000) et *Igf2r* (Killian *et al.*, 2000), et il semble que personne n'ait étudié l'orthologue du locus 14q32 humain chez les Marsupiaux.

Chapitre 4

Annexe 1

4.1 Article n°5

Nucleic Acids Research, 2003, Vol. 31, No. 22 6543–6551
DOI: 10.1093/nar/kgk849

Identification of 13 novel human modification guide RNAs

Patrice Vitali, Hélène Royo¹, Hervé Seitz¹, Jean-Pierre Bachellerie¹,
Alexander Hüttenhofer and Jérôme Cavallé^{1,*}

Institute for Molecular Biology, Department of Functional Genomics, University of Innsbruck, Peter-Mayr-Strasse 4b, 6020 Innsbruck, Austria and ¹Laboratoire de Biologie Moléculaire des Eucaryotes, UMR 5099 du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31602 Toulouse Cedex, France

Received August 5, 2003; Revised and Accepted September 23, 2003

ABSTRACT

Members of the two expanding RNA subclasses termed C/D and H/ACA RNAs guide the 2'-O-methylations and pseudouridylations, respectively, of rRNA and spliceosomal RNAs (snRNAs). Here, we report on the identification of 13 novel human intron-encoded small RNAs (U94–U106) belonging to the two subclasses of modification guides. Seven of them are predicted to direct 2'-O-methylations in rRNA or snRNAs, while the remainder represent novel orphan RNA modification guides. From these, U100, which is exclusively detected in Cajal bodies (CBs), is predicted to direct modification of a U6 snRNA uridine, U₉, which to date has not been found to be pseudouridylated. Hence, within CBs, U100 might function in the folding pathway or other aspects of U6 snRNA metabolism rather than acting as a pseudouridylate guide. U106 C/D snoRNA might also possess an RNA chaperone activity only since its two conserved antisense elements match two rRNA sequences devoid of methylated nucleotides and located remarkably close to each other within the 18S rRNA secondary structure. Finally, we have identified a retrogene for U99 snoRNA located within an intron of the *Siat5* gene, supporting the notion that retro-transposition events might have played a substantial role in the mobility and diversification of snoRNA genes during evolution.

INTRODUCTION

Eukaryal, bacterial and archaeal organisms contain an unexpectedly large number of non-coding RNAs (also called non-messenger RNAs or nmRNAs). Although most of them are poorly characterised as yet, they are thought to play important roles at various steps in the control of gene expression (1–3). Experimental RNomics studies in multicellular organisms have recently begun to illustrate the complexity of the nmRNA population in several model organisms (4–9). A

large fraction of nmRNAs belong to well-defined families, e.g. snRNAs, snoRNAs or miRNAs, exhibiting specific structural hallmarks, thus allowing many of them to be identified by computational searches of completely sequenced genomes. The combination of experimental and computational RNomics is now paving the way for a better understanding of the functions of this diverse class of RNA molecules (7).

In eukaryal and archaeal organisms, a substantial proportion of the nmRNAs identified so far belong to the two expanding subclasses, termed C/D and H/ACA RNAs, which guide the two prevalent types of rRNA modifications in both groups of organisms, i.e. 2'-O-methylation and pseudouridylation, respectively. For both subclasses, each guide RNA specifies the nucleotide to be modified through the formation of a canonical duplex spanning the cognate modification site (4,10–13). In eukaryotes, most of these relatively short guide RNAs (60–100 nt for C/D RNAs and 120–140 nt for H/ACA RNAs) accumulate within the nucleolus, hence they are designated as snoRNAs. In addition to targeting rRNAs, C/D and H/ACA snoRNAs also direct modifications of the Pol III transcribed spliceosomal U6 snRNA which is thought to transit through the nucleolus (14). In contrast, the C/D or H/ACA RNAs involved in the modification of the Pol II-transcribed spliceosomal snRNAs, U1, U2, U4 and U5, do not accumulate within the nucleolus but are exclusively found within the Cajal (coiled) bodies (or CBs), a subnuclear compartment in which they appear to interact with their RNA substrates (15). This subset of modification guides have been called small CB-specific RNAs or scaRNAs (16). Cellular RNAs targeted by C/D and H/ACA RNA guides also include tRNAs in Archaea as well as probably a few other RNA species in Eukarya (12,17). Intriguingly, an increasing number of 'orphan snoRNAs', i.e. presumptive modification guides which remain so far without an identified cellular RNA target, have been reported in mammals, some of them being expressed in a tissue-specific manner and subjected to genomic imprinting (6,17–20). The range of cellular functions mediated by both subclasses of modification guides might therefore be larger than recognised so far, particularly in complex organisms such as vertebrates or plants, stressing the need for a thorough identification of the C/D and H/ACA repertoires of such organisms.

*To whom correspondence should be addressed. Tel: +33 5 61 33 59 34; Fax: 33 5 61 33 58 86; Email: cavaille@ibcg.biotoul.fr
Correspondence may also be addressed to Alexander Hüttenhofer. Tel: +43 512 507 3630; Fax: +43 512 507 9880; Email: alexander.huetttenhofer@uibk.ac.at

6544 *Nucleic Acids Research*, 2003, Vol. 31, No. 22

Computational genomic searches of C/D RNAs targeting rRNA or spliceosomal snRNAs have been efficient in many organisms (21–26), mainly based on the presence of a relatively long, continuous antisense element at the target site. However, the power of this approach is considerably limited in the case of C/D specimens targeting unknown cellular RNA species. As for the H/ACA RNAs, due to the bipartite structure of their shorter antisense elements, most of them so far have remained refractory to computational searches, even for those targeting rRNAs or snRNAs, and most of them have been identified through direct experimental approaches (5,6,8).

Here, by a combination of experimental screens of rat libraries and *in silico* searches of the human and mouse genomes, we report on the characterisation of 10 novel C/D and three novel H/ACA small RNAs which are all intron encoded and conserved among human and rodents. Remarkably, six of these new specimens (U97, U98, U99, U100, U101 and U106) do not seem to be involved in the modification of rRNAs or snRNAs. One of them, H/ACA U100, is particularly intriguing as it is the first scaRNA predicted to target a Pol III transcript, snRNA U6, at a nucleotide position not known to be modified, suggesting that it might have a function different from a typical pseudouridylation guide.

MATERIALS AND METHODS

Unless otherwise noted, all techniques for cloning and manipulating nucleic acids were performed according to standard protocols.

Oligonucleotides

The following oligonucleotides were used: U6, 5'-CGT-GTCATCCTTGCGCAGGGGCC-3'; U3, 5'-AAATGTCCC-TGAAAGTATAGTCTT-3'; U94, 5'-TCCGTACCCCTGC-GCCAATCATCA-3'; U97, 5'-TCATATCTCATAATCTTC-GCTCATAGGACG-3'; U98, 5'-AAACAGAACTGCGACC-GTCAAGGAA-3'; U99, 5'-TGTCGCCGCGTTTGAGGAT-AGAACC-3'; U100, 5'-TGTATGGAGCCATCGCACAGA-AAATCTGA-3'; U101, 5'-TCAGACTCTTATGTTTCACT-CATAA-3'; U102, 5'-TCAGAGCCCGGTGAAATGTGTT-TTC-3'; U106, 5'-TCAGAACTAACTGGCAAAATATAA-GACGTCA-3'; RT-18S, 5'-CCTCGTTCATGGGGAATAA-TTGC-3'; and U100-C6dT, 5'-GATAACTAXACAGACC-CTGXC GGCAGGAACCATCTGXTTAAATGTGTGXG-3' (amino-allyl-modified T residues are indicated by an XX).

Search for novel RNA modification guides

Human ribosomal protein genes were systematically searched for C (RUGAUGA) and D (CUGA) boxes. Pairs of correctly spaced (C box 35–140 bp upstream from D box) C and D boxes, with no more than one deviation as compared with the C and D box consensus, and flanked by complementary sequences (at least 4 bp complementarity, with no more than one G–U pair, among the five nucleotides beside each box), were then searched for murine orthologues using a UCSC blat search (<http://genome.ucsc.edu/cgi-bin/hgBlat>). Construction of the C/D snoRNA cDNA library has been described (19). The insert of individual clones was PCR amplified using M13 reverse and forward primers and sequenced using M13 forward primer and the BigDye terminator cycle sequencing

reaction kit (PE Applied Biosystems). Sequences were analysed on an ABI Prism 377 (Perkin Elmer) sequencer using the LASERGENE sequence analysis program package.

Cell fractionation, immunoprecipitation and northern blot analysis

Subcellular fractionation of HeLa cells was performed as described (27). Immunoprecipitation of trimethylated capped RNAs with monoclonal antibody R1131 (kindly provided by Dr R. Luhrmann) was performed according to Cavaillé *et al.* (17). RNAs were fractionated by electrophoresis in 6% acrylamide/7 M urea gels and transferred onto nylon membranes (Qiabran Nylon Plus, Qiagen) using the Biorad semi-dry blotting apparatus (Trans-blot SD, Biorad). After immobilising RNA using the Stratagene cross-linker, nylon membranes were pre-hybridised for 30 min in 1 M sodium phosphate buffer pH 6.2, 7% SDS. Oligonucleotides complementary to the respective snoRNA species were end-labelled with [³²P]ATP and T4 polynucleotide kinase; hybridisation was carried out at 58°C in 1 M sodium phosphate buffer pH 6.2, 7% SDS overnight. Blots were washed twice at room temperature in 2× SSC buffer (20 mM sodium phosphate pH 7.4, 0.3 M NaCl, 2 mM EDTA), 0.1% SDS for 15 min and exposed to Kodak MR-1 film.

In situ hybridisation

The human U100 gene was PCR amplified from HeLa cell genomic DNA and cloned into the ClaI and XhoI sites of the pCMV-globin expression vector [kindly provided by Dr T. Kiss (16)]. This pCMV-U100 vector together with pGFP-coilin (kindly provided by Dr E. Bertrand) was transfected into HeLa cells with Fugene transfection reagent according to the manufacturer's recommendations (Roche). The U100-specific modified oligonucleotide was labelled with FluoroLink Cy3 reactive dye (Amersham), and fluorescent *in situ* hybridisation (FISH) was performed according to the protocol of the laboratory of Dr Singer (<http://singerlab.aecom.yu.edu>). Nuclear DNA was stained by 4',6-diamidino-2-phenylindole (DAPI).

RESULTS AND DISCUSSION

Identification of novel C/D snoRNAs

By screening two C/D snoRNA-specific cDNA libraries obtained from a whole rat brain (19) and from rat testis extracts (P.Vitali and A.Hüttenhofer, unpublished data), we have identified scores of C/D RNAs. While most of them correspond to rat counterparts of previously characterised human (4) or mouse (6) C/D snoRNAs (data not shown), six novel specimens were detected among the sequenced clones. These RNAs, referred to as U94, U95, U103, U104, U105 and U106, exhibit all the C/D hallmarks, and their homologues in mouse and human have been identified by database searches (not shown).

In most cases, the small intron-encoded RNA is hosted by the same gene in human and mouse. In both mammals, two U103 gene copies are located within introns 17 and 21 of the PUM1 gene (28), while U94 and U95 are encoded within introns of the FLJ20758 and GNB2L1 genes, respectively (Table 1). In the case of U95, however, found in GNB2L1

Table 1. Compilation and major features of the RNA modification guides described in this study

Name	Family	RNA target (human)	Antisense element	snoRNA host gene (human)	Accession no.
U94	C/D	Cm62 (U6)	13 bp/D	FLJ20758 (intron 21), conserved hypothetical protein	AY349593
U95 ^a	C/D	Am2792/Cm2801 (28S)	13 bp/D + 10 bp/D'	GNB2L1 (intron 1), guanine nucleotide-binding protein (G protein)	AY349594
U96 ^{a,b}	C/D	Gm75 (5.8S)	12 bp/D	GNB2L1 (intron 2)	AY349595
U96b	C/D	Gm75 (5.8S)	12 bp/D	AMMECR1 (intron 2), unknown function	AY349596
U97	C/D	ND		EIF4G2 (intron 14), translation initiation factor	AY349597
U98a	H/ACA	ND		LOC85028 (intron 2), poorly characterised gene	AY349598
U98b	H/ACA	ND		PPP2R5A (intron 8), protein phosphatase 2, regulatory subunit B	AY349599
U99	H/ACA	ND		MGC2477 (intron 3), conserved hypothetical protein	AY349600
U100	H/ACA	U9? (U6) ^c	(6 + 4)/3' hairpin	FLJ20516 (intron 6), homologous to Tipin	AY349601
U101	C/D	ND		RPS12 (intron 3), ribosomal protein	AY349602
U102	C/D	Gm4010 (28S)	11 bp/D	RPL21 (intron 2), ribosomal protein	AY349603
U103 ^{a,d}	C/D	Gm601 (18S)	10 bp/D'	PUM1 (intron 17), RNA-binding protein (Pumilio family member)	AY349604
U103b ^d	C/D	Gm601 (18S)	10 bp/D'	PUM1 (intron 21)	
U104 ^a	C/D	Cm1320 (28S)	14 bp/D'	Not annotated gene	AY349605
U105	C/D	Um799 (18S)	15 bp/D	SSF1 (intron 3), Peter Pan homologue (contains a Brix domain)	AY349606
U106	C/D	G1536/U1602? (18S) ^c	10 bp/D + 10 bp/D'	BC032480 (intron 1), poorly characterised gene	AY349607

^aU95, U96 and U104 sequences have been independently submitted in the public database as: Z38, Z37 and Z12 snoRNAs, respectively.

^bU96 might be the mammalian counterpart of the *A.thaliana* Z37a and Z37b snoRNA involved in Gm79 formation in *A.thaliana* 5.8S rRNA (24).

^cA question mark means that no Ψ9 or Gm1536 and Um1602 have been detected in human U6 snRNA and 18S rRNA, respectively.

^dU103a and U103b RNAs target the same rRNA position as MBII-251 and thus might be considered as snoRNA isoforms of the latter.

intron 1 in both mammals, an additional gene copy is present in intron 3 of the same gene in mouse only (Fig. 1).

Genes hosting an intron-encoded snoRNA frequently contain one (or more) additional snoRNA(s) within other intron(s). We therefore systematically searched other introns of the three above-mentioned host genes for the potential presence of additional C/D-like sequences. A likely candidate, termed U96a, was identified in intron 2 of the GNB2L1 gene. Subsequently, we noticed that the human genome contains an additional, strongly related copy of the U96 sequence, U96b, located in intron 2 of the AMMECR1 gene. This gene maps to chromosome Xq22 and is potentially involved in the pathogenesis of the AMME (Alport syndrome, mental retardation, midface hypoplasia and ellipocytosis) contiguous gene deletion syndrome (29). Curiously, the murine *Ammecri1* gene does not contain an intron-encoded, U96-like sequence (Fig. 1).

The novel C/D snoRNA candidates were searched for the presence of appropriate complementarities to rRNAs or snRNAs. U95, which contains two antisense elements, is predicted to target two neighbouring 2'-O-methylations, Cm2801 and Am2792, in 28S rRNA. As for U96, U103, U104 and U106 snoRNAs, they are predicted to direct Gm75 (5.8S rRNA), Gm601 (18S rRNA), Cm1320 (28S rRNA) and Um799 (18S rRNA), respectively, through phylogenetically conserved RNA guide duplexes (Table 1).

Remarkably, U106 contains two conserved, 9–10 nt long antisense elements matching two 18S rRNA segments (Fig. 2A) devoid of 2'-O-methylated nucleotides (30) and mapping very close to each other within the rRNA secondary structure (Fig. 2B). We confirmed that the two nucleotides (G1536 and U1602 in 18S rRNA) predicted to be targeted by U106 do not seem to be 2'-O-methylated as judged by the absence of reverse transcription pauses at low dNTP concentration (Fig. 2C). Many archaeal and plant C/D RNAs have been reported to carry two antisense elements interacting with neighbouring rRNA sites which might reflect their important

role as RNA chaperones for rRNA folding in addition to nucleotide modification (22,23,25,26). Thus, our observation suggests that U106 might play a role distinct from directing 2'-O-methylations. Interestingly, this could also be the case for a few C/D snoRNAs, such as MBII-142, MBII-170, MBII-289, MBII-295, MBII-316 and MBII-426, which all can form canonical guide duplexes targeting apparently unmethylated rRNA sites [(6); J.P.Bachellerie, unpublished data].

Finally, U94 contains an 11 nt long antisense element spanning three adjacent methylated nucleotides in the U6 snRNA, Cm60, Cm62 and Cm63, with Cm62 being paired to the fifth position upstream from the D' box (data not shown). Although this RNA duplex contains a bulged nucleotide on the snoRNA strand (between box D' and the targeted nucleotide), such an irregularity does not hamper efficient RNA guide activity (31) and we thus propose that U94 is the *bona fide* RNA guide for Cm62.

Genes hosting both subclasses of intron-encoded snoRNAs belong to the 5' (TOP) family of vertebrate genes, including many ribosomal protein genes (32,33). In an attempt to identify additional intron-encoded C/D snoRNAs, we systematically searched all introns of the complete set of ribosomal protein genes in the human genome for the presence of C/D hallmarks. In this way, we could detect in introns of the RPS12 and RPL21 genes two novel C/D RNAs, called U101 and U102, respectively, which are strongly conserved among mammals. While U102 is predicted to direct a highly conserved 2'-O-methylation in 28S rRNA (Table 1), U101 remains without an identified target in rRNAs or snRNAs.

Finally, we re-examined all the candidate snmRNA genes previously identified by screening of a mouse brain cDNA library that could not have been assigned to a snoRNA subclass (6) essentially because they corresponded to very short, probably truncated cDNA sequences not represented in the genomic sequences in databases at that time. Following the publication of the entire human genome and large parts of the mouse genome, the 5'-terminal sequences of all these

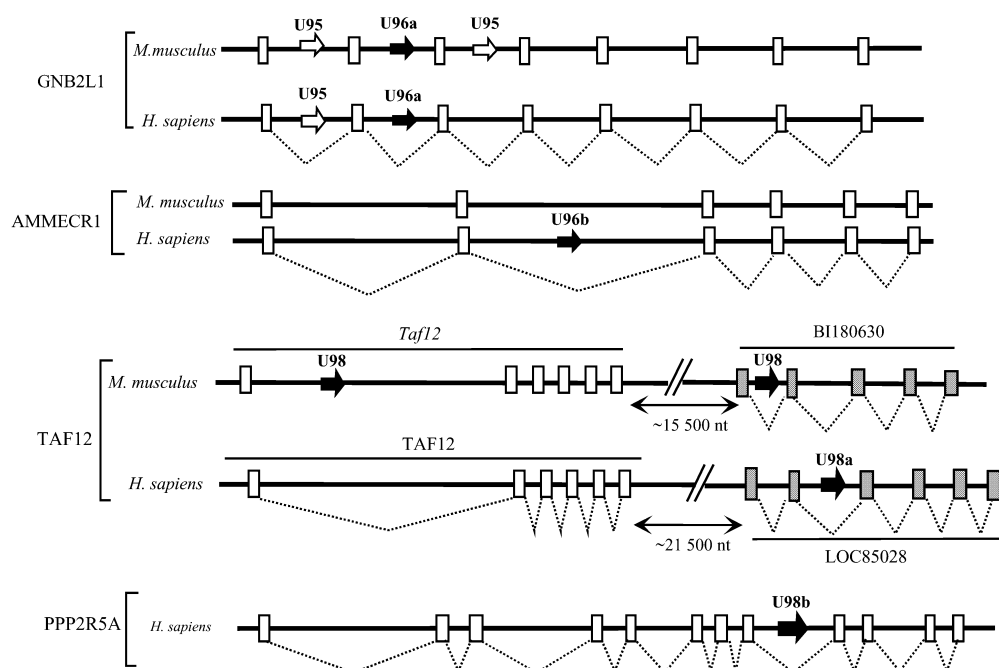
6546 *Nucleic Acids Research*, 2003, Vol. 31, No. 22

Figure 1. Differences in the genomic organisation of the three novel snoRNAs between human and mouse. Each snoRNA sequence (small arrow) is located within an intron of the indicated genes. Exons are represented by boxes and splicing events by dotted lines. Note that AK009175, a spliced EST, and LOC85028, a poorly characterised gene, both located 15–21 kb downstream from TAF12 in mouse and human, are not related to each other. The cartoon is not drawn to scale.

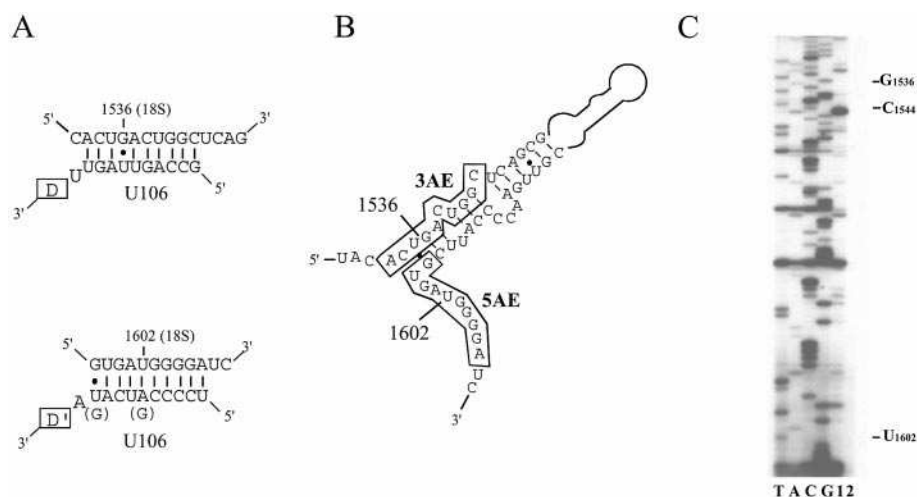


Figure 2. Interaction between C/D snoRNA U106 and 18S rRNA. (A) Predicted base pairing between U106 and 18S rRNA. The sequences shown are for human (nucleotide differences in rodents are indicated in parentheses). Note that G1536 and U1602 do not appear to be 2'-O-methylated (30). (B) Location of the sites of complementarity to snoRNA U106 within the 18S rRNA secondary structure. The 3' and 5' antisense elements of U106 are denoted by 3AE and 5AE, respectively, and the rRNA nucleotides potentially involved in base pairing with the snoRNA are boxed. (C) Mapping of ribose methylated nucleotides in 18S rRNA. Primer extension at low concentrations of dNTP was performed with a 5'-³²P-labelled 18S rRNA-specific oligonucleotide, either at 0.04 mM (lane 1) or at 1 mM (control, lane 2). The significance of a pause at C1544 is unclear since 2'-O-methylation at this position has never been reported so far.

snmRNAs could be examined, leading to the definite assignment of a few of them to one of the two snoRNA subclasses. Thus, in the recently published human genome, we could identify the sequence homologous to mouse MBI-82 snmRNA. The 5'-extended human sequence, located in an intron of the EIF4-G2 gene, clearly exhibits the C/D hallmarks. However, the corresponding RNA, renamed U97, remains without a predicted target in rRNAs or snRNAs. All the novel C/D RNAs mentioned above, whether they were identified in the cDNA libraries or predicted by search of sequence databases, were experimentally verified by northern blotting analysis and found to be ubiquitously expressed (data not shown).

Identification of novel H/ACA snoRNAs

We also re-examined by reference to the updated human and mouse genomic sequences all the snmRNA candidates not yet assigned to snoRNAs in the screen of the mouse brain library mentioned above (6). We found that three of the most strongly expressed specimens in this category, MBII-367, MBII-104 and MBII-201, do in fact correspond to intronic sequences exhibiting the characteristic H/ACA hallmarks, i.e. two large hairpin domains linked by a hinge (containing the H box sequence, ANANNA) and followed by a short tail containing the ACA motif (3 nt away from the RNA 3' end). The expression of these snmRNAs, renamed U98, U99 and U100, which are conserved in mammals, has been experimentally verified by northern blot analysis [(6); data not shown].

In the mouse, two identical U98 genes map very close to each other: one is encoded within the large intron 1 of the strongly conserved *Taf12* gene and the other one within a poorly known gene ~15 kb downstream from *Taf12*. In human, the U98 gene organisation is quite distinct, with the TAF12 gene devoid of any related U98-related sequence while a U98 gene copy is also found immediately downstream from the TAF12 gene. Moreover, another U98 gene copy, U98b, is embedded within intron 8 of the PPP2R5A gene (Fig. 1). Intriguingly, the U99 gene, which maps, in both mouse and human, to intron 3 of a gene encoding a conserved hypothetical protein (MGC2477), is transcribed in the opposite orientation to this gene (Fig. 3A). This might suggest that U99 could be independently transcribed from its own promoter in the intron, rather than being processed from the debranched intron like all known vertebrate C/D or H/ACA snoRNAs. However, lack of immunoprecipitation of U99 with the R1131 antibody (Fig. 3B) that specifically recognises the trimethyl-guanosine cap structure demarking the 5' end of vertebrate snoRNAs transcribed from independent genes is consistent with U99 representing a processed RNA. Moreover, we noticed that several human and mouse spliced expressed sequence tags (ESTs) mapping within intron 3 of the MGC2477 gene were indicative of transcription in the direction opposite to this gene, accompanied by splicing of a U99-containing intron mapping within the boundaries of the MGC2477 intron 3, as shown in Figure 3A. While U99 is likely to be produced by intron processing of the MGC2477 antisense transcript, the biological significance of this antisense genomic organisation, never previously described for a snoRNA modification guide, remains elusive.

We could not identify any significant target uridine in rRNA or snRNA for U98 and U99 H/ACA RNAs. In contrast, U100

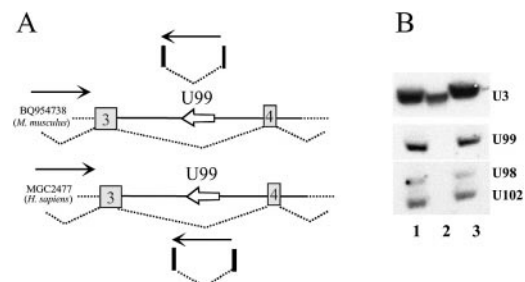


Figure 3. The U99 snoRNA gene is transcribed in the opposite orientation to its host intron. (A) Schematic representation of a part of the mouse (top) and human (bottom) genes hosting U99 snoRNA. Exons 3 and 4 are represented by white boxes, while splicing events are denoted by dotted lines. The U99 snoRNA gene is schematised by a white arrow indicating transcription orientation. Thin black arrows denote the transcription orientation of the indicated genes. Several spliced ESTs (i.e. BU588934, BF219096, BU588653 and BI461867 in human, and a single one, AK011444, in the mouse) overlapping and in the same orientation as the U99 snoRNA gene are also depicted (filled boxes denote exons). Not drawn to scale. (B) U99 does not immunoprecipitate with R1131 antibody. Total RNA from HeLa cells was subjected to immunoprecipitation with R1131 antibody (specific for the trimethyl cap structure), and RNAs recovered from either the pellet or supernatant were assayed for U99 by northern blot hybridisation with a U99-specific oligonucleotide probe. 1, input RNA (1:10); 2, pellet; 3, supernatant (1:10). The 5'-trimethyl-capped snoRNA U3 and the intron-encoded U98 and U102 snoRNAs were used as positive and negative controls, respectively.

can form with snRNA U6 a canonical bipartite duplex predicted to direct pseudouridylation at position 9 within the spliceosomal RNA (Fig. 4A). A fish U100 homolog was identified by a search of the *Fugu rubripes* genomic sequence. Remarkably, the presumptive bipartite guide RNA duplex is conserved between *Homo sapiens* and *F. rubripes* (Fig. 4C). Intriguingly, however, position U₉ in a stem-loop structure of U6 has not been reported to be pseudouridylated (34). We experimentally verified the absence of the predicted pseudouridine by a reverse transcriptase approach following chemical modification by CMC and alkaline treatment of the RNA template (Fig. 4B). Two other previously reported mouse H/ACA snoRNAs, MBI-39 and MBI-164, seem able to direct pseudouridylation on rRNA uridines which are not experimentally found to be modified (6). U100, as well as MBI-39 and MBI-164, might therefore play a role distinct from that of a pseudouridylation guide, possibly acting as RNA chaperones (see below).

U100 H/ACA RNA belongs to the scaRNA family

To date, RNA modification guides have been detected either in the nucleolus or in the nucleoplasm, according to the nature of their RNA substrates (13). Specimens located in the nucleolus include guides for the modification of rRNA as well as Pol III-transcribed snRNA U6 (14), whereas guides for the modification of Pol II-transcribed spliceosomal snRNAs, U1, U2, U4 and U5, are found within the CBs (13,15,16,35). We have investigated the intracellular localisation of the novel orphan C/D and H/ACA RNAs by isolating different subcellular fractions obtained from exponentially growing HeLa cells. All of them, except U100, are highly enriched in the nucleolar

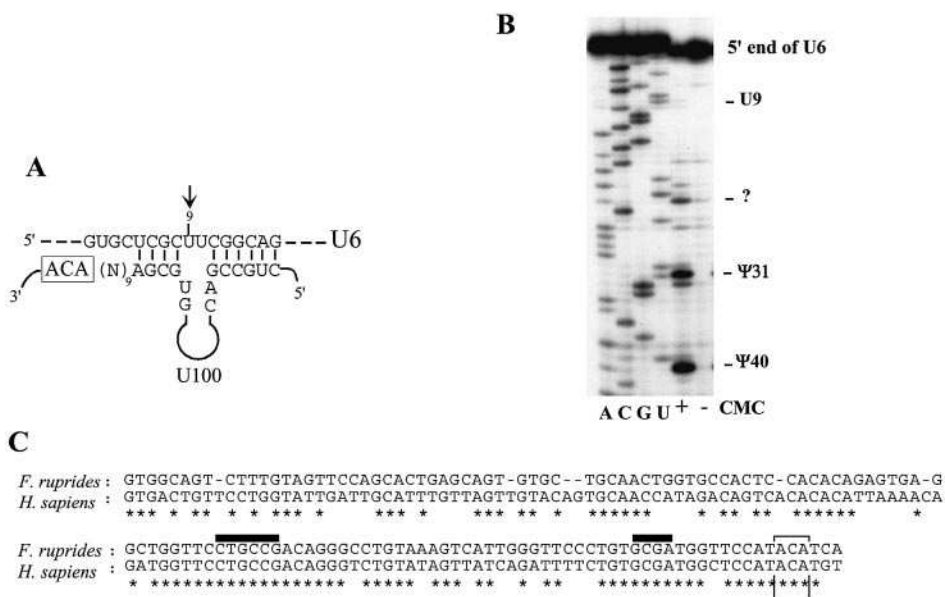


Figure 4. U100, a novel RNA guide targeting U6 snRNA. (A) Predicted base pairing of U100 with U6 snRNA. The U6 nucleotide predicted to be targeted for modification by U100 is indicated by an arrow. Note that only the 3' hairpin domain from U100 is shown. The sequences shown are for human. (B) Mapping of pseudouridines at the 5' end of the spliceosomal U6 snRNA. Total RNA extracted from HeLa cells, treated (+) or not treated (-) with CMC, was subjected to primer extension analysis with a U6-specific ³²P-labelled oligonucleotide. (C) U100 sequence alignment between *H.sapiens* and the fish *F.rubripes*. The fish U100 gene is located in an intron of the Huntington's disease gene homologue (accession no. X82939). The conserved ACA motif and potential bipartite antisense element are denoted (boxed and overlined, respectively). Below the alignment, conserved nucleotides are denoted by asterisks.

fraction (Fig. 5A). To gain further insight about the intranuclear location of U100 RNA, FISH with a U100-specific oligonucleotide probe has been performed in HeLa cells. As endogenous U100 signals were relatively weak (not shown), we cloned the human U100 gene into the pCMV-globin expression construct (16) and co-transfected it into HeLa cells together with a green fluorescent protein (GFP)-coilin expression plasmid which was used as a marker protein for CBs. As shown in Figure 5B, U100 signals did not distribute homogeneously within the nucleoplasm but rather accumulated within distinct domains that co-localise with GFP-coilin signals. We therefore conclude that U100 is a novel member of the scaRNA family, in contrast to previously reported RNA guides for U6 modifications (as well as U94 described in this study) which have been all detected within the nucleolus [(14); Fig. 5A]. Consistent with U100 being a scaRNA, we noticed that it harbors a recently identified CB-specific localisation signal [(27); Fig. 5C]. U6 snRNA has been previously shown to transit through CBs (36). Thus, U100 is the first H/ACA RNA predicted to potentially base-pair with a Pol III transcript outside the nucleolus. Moreover, the absence of pseudouridylation at position 9 in U6 snRNA suggests an unusual role for U100 in the metabolism of U6 in CBs. Thereby, U100 could act as an RNA chaperone presumably within the CBs during U6 snRNP and/or U4/U6 snRNP assembly and trafficking within the nucleoplasm, en route to and from the nucleolus (37).

Detection of an intronic snoRNA retrogene

During our BLAST search of the mouse genome, in addition to the sequences encoding H/ACA RNA U99, we also detected an additional U99-like copy (Fig. 6) preceded by an 81 nt long tract identical to the intronic sequence immediately upstream from the functional U99 gene and immediately followed by a 3' poly(A) stretch. The U99-like copy and upstream 81 nt were flanked by a pair of 13 nt long direct repeats. This probably corresponds to an H/ACA retrogene resulting from the insertion at staggered nicks in genomic DNA of a cDNA produced from an intronic RNA processing intermediate in U99 biogenesis. Intronic snoRNAs can be hosted by different genes in distant species, and retrotransposition mechanisms have been proposed to play a pivotal role in the mobility and diversification of snoRNA genes (12,38,39). In particular, retroposition into an intron may frequently result in the production of a faithfully processed snoRNA retrogene transcript, owing to the presence of internal *cis*-acting processing signals. Remarkably, the full-length U99 retrogene mentioned above is located in the mouse *Siat5* gene (Fig. 6, bottom) within an intron of its 5'-untranslated region as reflected by the detection of many spliced ESTs. We checked by RT-PCR that the corresponding portion of the *Siat5* gene intron was expressed in mouse cells (data not shown). It is noteworthy that within the U99 retrogene sequence, the tracts, which are predicted to correspond to bipartite antisense

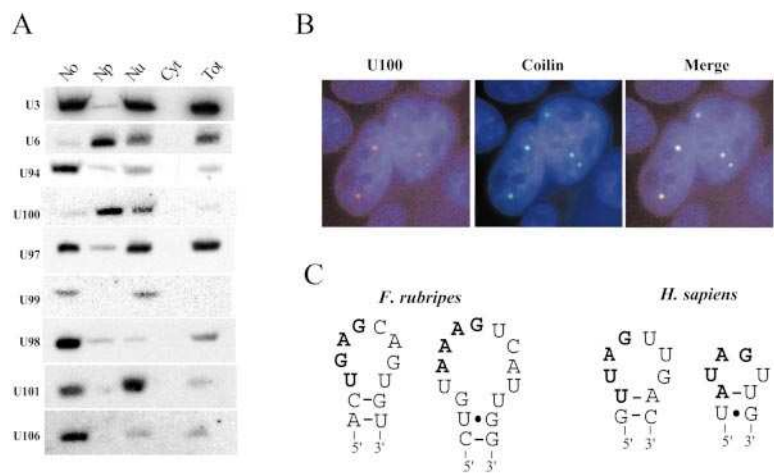


Figure 5. U100 is a novel member of the scaRNA family. (A) Subfractionation of HeLa cells. RNA isolated either from total HeLa cells (Tot) or from cytoplasmic (Cyt), nuclei (Nu), nucleoplasmic (Np) or nucleolar (No) fractions was analysed in a 6% acrylamide/7 M urea gel and the various snoRNAs detected by northern blot analysis using specific oligonucleotide probes. Hybridisations with U3- and U6-specific probes have been used as controls of the cell fractionation procedure. (B) *In situ* hybridisation showing the localisation of transfected human U100. HeLa cells co-transfected with pCMV-hU100 and pGFP-coilin were hybridised with a specific U100 fluorescent oligonucleotide. The Cajal bodies are visualised by co-expressing GFP-coilin fluorescent protein. The merged picture shows that U100 co-localises with coilin. (C) Schematic representation of the terminal loops of the 5' and 3' hairpins of U100. The predicted Cajal body-specific localisation signals (27) are indicated in bold.

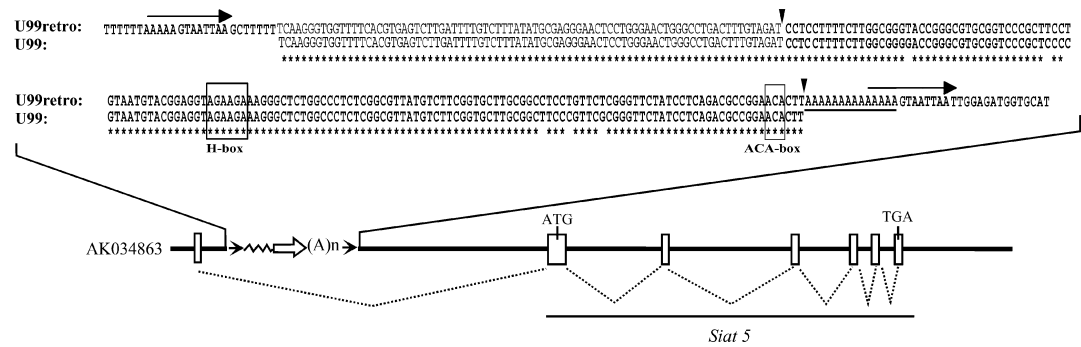


Figure 6. Identification of an intronic U99 retrogene. Sequence alignment of a U99 mouse snoRNA retrogene with its functional counterpart. Boundaries of the mature U99 sequence are denoted by vertical arrowheads. Direct repeats flanking the snoRNA retrogene are overlined (arrow), while the poly(A) stretch is underlined and the H and ACA motifs boxed. Bottom: intronic location of the U99 retrogene within the mouse *Siat5* gene, with indication of a spliced EST (AK034863) connecting the U99 retrogene-containing intron to the rest of the *Siat5* gene. Exons are represented by white boxes, splicing events by dotted lines, and the U99 retrogene by an open arrow (the flanking direct repeats are depicted by arrowheads). Nucleotides conserved between gene and retrogene are indicated by asterisks.

element(s) of the H/ACA RNA, are perfectly conserved by reference to the *bona fide* U99 gene. More generally, none of the few sequence differences exhibited by the U99 retrogene are expected to be detrimental to H/ACA snoRNA structure and function. The U99 retrogene might well encode a functional snoRNA, a possibility without precedent to our knowledge. Overall, sequences of the U99 retrogene and *bona fide* U99 gene diverge by only 5%, pointing to a relatively recent origin of the snoRNA retrogene in the mouse lineage. In line with this notion, the homologous intron of the human *SIAT5* gene is devoid of any U99-like sequence.

Conclusions

Among the 13 novel specimens of the C/D or H/ACA subclasses characterised in this study, six are especially intriguing. Four (C/D snoRNAs U97 and U101 and H/ACA snoRNAs U98 and U99) belong to the expanding group of orphan snoRNAs and remain without identified targets in rRNAs or snRNAs. They could either target cellular RNA species distinct from rRNA or snRNAs, or perform an entirely different function. In the case of the other two specimens, the H/ACA scaRNA U100 and the C/D snoRNA U106, our

6550 *Nucleic Acids Research*, 2003, Vol. 31, No. 22

present evidence points to the second possibility, suggesting that they might be involved in an RNA chaperone function in the metabolism of U6 snRNA and 18S rRNA, respectively.

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Daniel Vitali. We thank T. Kiss for sharing unpublished data about the Cajal body-specific localisation signal, J. Brosius for critical reading of the manuscript, E. Bertrand for advice on RNA FISH, and M.-L. Bortolin and T. Rozhdestvensky for technical support. This work was supported by grants from the 'Toulouse Génopole', the Programme Interdisciplinaire du Centre National de la Recherche Scientifique 'Dynamique et Réactivité des Assemblages Biologiques' and 'La Ligue contre le Cancer/Comité de Haute-Garonne' (to J.C.), and by the German Human Genome Project through the BMBF (01KW9966) and an IZKF grant (Teilprojekt IKF3 G6, Münster) (to A.H.).

REFERENCES

- Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.*, **2**, 919–929.
- Mattick,J.S. and Gagen,M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, **18**, 1611–1630.
- Wassaman,K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell*, **109**, 141–144.
- Kiss-Laszlo,Z., Henry,Y., Bachelier,J.P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
- Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.
- Huttenhofer,A., Kieffmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachelier,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
- Huttenhofer,A., Brosius,J. and Bachelier,J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.
- Marker,C., Zemann,A., Terhorst,T., Kieffmann,M., Kastenmayer,J.P., Green,P., Bachelier,J.P., Brosius,J. and Huttenhofer,A. (2002) Experimental RNomics. Identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
- Yuan,G., Klambt,C., Bachelier,J.P., Brosius,J. and Huttenhofer,A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
- Cavaille,J., Nicoloso,M. and Bachelier,J.P. (1996) Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, **383**, 732–735.
- Ganot,P., Bortolin,M.L. and Kiss,T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
- Bachelier,J.P., Cavaille,J. and Huttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
- Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
- Ganot,P., Jady,B.E., Bortolin,M.L., Darzacq,X. and Kiss,T. (1999) Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol. Cell. Biol.*, **19**, 6906–6917.
- Jady,B.E., Darzacq,X., Tucker,K.E., Matera,A.G., Bertrand,E. and Kiss,T. (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J.*, **22**, 1878–1888.
- Darzacq,X., Jady,B.E., Verheggen,C., Kiss,A.M., Bertrand,E. and Kiss,T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746–2756.
- Cavaille,J., Buiting,K., Kieffmann,M., Lalande,M., Brannan,C.I., Horsthemke,B., Bachelier,J.P., Brosius,J. and Huttenhofer,A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
- Jady,B.E. and Kiss,T. (2000) Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2'-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. *Nucleic Acids Res.*, **28**, 1348–1354.
- Cavaille,J., Vitali,P., Basyuk,E., Huttenhofer,A. and Bachelier,J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.
- Cavaille,J., Seitz,H., Paulsen,M., Ferguson-Smith,A.C. and Bachelier,J.P. (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.*, **11**, 1527–1538.
- Lowe,T.M. and Eddy,S.R. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
- Brown,J.W., Clark,G.P., Leader,D.J., Simpson,C.G. and Lowe,T. (2001) Multiple snoRNA gene clusters from *Arabidopsis*. *RNA*, **7**, 1817–1832.
- Barnache,F., Gaspin,C., Guyot,R. and Echeverria,M. (2001) Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. *J. Mol. Biol.*, **311**, 57–73.
- Qu,L.H., Meng,Q., Zhou,H., Chen,Y.Q., Liang-Hu,Q., Qing,M., Hui,Z. and Yue-Qin,C. (2001) Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*. *Nucleic Acids Res.*, **29**, 1623–1630.
- Omer,A.D., Lowe,T.M., Russell,A.G., Ebhardt,H., Eddy,S.R. and Dennis,P.P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517–522.
- Gaspin,C., Cavaille,J., Erauso,G. and Bachelier,J.P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.*, **297**, 895–906.
- Richard,P., Darzacq,X., Bertrand,E., Jady,B.E., Verheggen,C. and Kiss,T. (2003) A common sequence motif determines the Cajal body-specific localization of H/ACA scaRNAs. *EMBO J.*, **22**, 4283–4293.
- Spassov,D.S. and Jurecic,R. (2002) Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins. *Gene*, **299**, 195–204.
- Vitelli,F., Meloni,L., Fineschi,S., Favara,F., Tiziana Storlazzi,C., Rocchi,M. and Renieri,A. (2000) Identification and characterization of mouse orthologs of the AMMECR1 and FACIL4 genes deleted in AMME syndrome: orthology of Xq22.3 and MmuXF1-F3. *Cytogenet. Cell Genet.*, **88**, 259–263.
- Maden,B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **39**, 241–303.
- Cavaille,J. and Bachelier,J.P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.*, **26**, 1576–1587.
- Smith,C.M. and Steitz,J.A. (1998) Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol.*, **18**, 6897–6909.
- Pelczar,P. and Filipowicz,W. (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.*, **18**, 4509–4518.
- Massenet,S., Mougin,A. and Branlant,C. (1998) *Posttranscriptional Modifications in the U snRNAs*. ASM Press, Washington, DC.
- Kiss,A.M., Jady,B.E., Darzacq,X., Verheggen,C., Bertrand,E. and Kiss,T. (2002) A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res.*, **30**, 4643–4649.

Nucleic Acids Research, 2003, Vol. 31, No. 22 **6551**

36. Carmo-Fonseca,M., Pepperkok,R., Carvalho,M.T. and Lamond,A.I. (1992) Transcription-dependent colocalization of the U1, U2, U4/U6 and U5 snRNPs in coiled bodies. *J. Cell Biol.*, **117**, 1–14.
37. Stanek,D., Rader,S.D., Klingauf,M. and Neugebauer,K.M. (2003) Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.*, **160**, 505–516.
38. Bachelierie,J.P., Nicoloso,M., Qu,L.H., Michot,B., Caizergues-Ferrer,M., Cavaillie,J. and Renalier,M.H. (1995) Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem. Cell Biol.*, **73**, 835–843.
39. Maxwell,E.S. and Fournier,M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897–934.

Chapitre 5

Annexe 2

5.1 Problématique

Après la découverte d'ARN C/D dont les gènes sont soumis à l'empreinte génomique parentale (Cavaillé *et al.*, 2000 ; Runte *et al.*, 2001 ; Cavaillé *et al.*, 2001 ; Cavaillé *et al.*, 2002), la question se pose de savoir s'ils participent eux aussi à la biogenèse des ARN ribosomiques. La présente étude donne les résultats des recherches de complémentarités entre plusieurs de ces ARN C/D et le précurseur de 45S de l'ARN ribosomique humain, et les compare aux complémentarités attendues entre le précurseur de 45S et une séquence aléatoire de même composition en bases que l'élément antisens de l'ARN C/D considéré.

Les complémentarités recherchées sont parfaites (*i.e.* : sans mésappariement), sans appariement G-U, et se trouvent immédiatement en amont de la boîte D (ou D'), ou 1 nucléotide en amont de la boîte (voir figure 5.2). Ces caractéristiques sont celles de la plupart des complémentarités fonctionnelles (connues pour guider une méthylation sur le précurseur de l'ARN ribosomique).

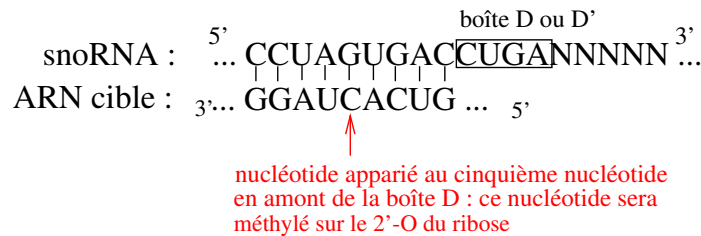


FIG. 5.1 – GUIDAGE DE LA MÉTHYLATION DU 2'-O DU RIBOSE PAR LES ARN C/D.

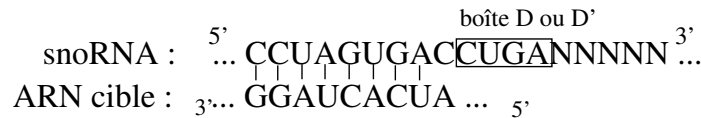


FIG. 5.2 – COMPLÉMENTARITÉ NE FAISANT PAS INTERVENIR LE NUCLÉOTIDE IMMÉDIATEMENT EN AMONT DE LA BOÎTE.

Le modèle mathématique utilisé est une simplification : il néglige les biais de répartition des dinucléotides dans les séquences génomiques humaines, et il utilise quelques approximations, pour alléger les calculs (voir section 5.2).

5.1.1 Données présentées

Cette annexe rassemble les statistiques sur les complémentarités d'ARN C/D humains dont les gènes sont soumis à l'empreinte génomique parentale, sur le précurseur de 45S des ARN ribosomiques humains. Les analyses de tous les ARN C/D dont les gènes sont soumis à l'empreinte ne sont pas présentées ici, seules celles de quelques exemples représentatifs, choisis pour la similarité de leur élément antisens à celui du consensus de leur famille (voir figures 5.3, 5.4 et 5.5 ; les alignements ont été réalisés avec le programme Multalin¹ : Corpet, 1988). Six snoRNA dont les gènes ne sont pas soumis à l'empreinte, et dont les complémentarités à l'ARNpré-r sont connues (Kiss-László *et al.*, 1996) ont été inclus dans cette analyse.

¹Interface web du programme Multalin : <http://prodes.toulouse.inra.fr/multalin/multalin.html>

	1	10	20
	-----+-----		
14q32(II)1	GTGAATACAGG	TCTGGAA	GT
14q32(II)5	ATGAATACGTGTCTAGAA	CT	AACT
14q32(II)25	ATGAATACGTGTCTAGAA	CT	AACT
14q32(II)28	ATGAATACGTGTCTAGAA	CT	AACT
14q32(II)9	ATGAATACGTGTCTGGAA	CT	AACT
14q32(II)21	ATGAATACGTGTCTGGAA	CT	AACT
14q32(II)23	ATGAATACGTGTCTGGAA	CT	AACT
14q32(II)26	ATGAATACGTGTCTGGAA	CT	AACT
14q32(II)10	ATGAATACATGTCTGAA	CT	AACT
14q32(II)27	ATGAATACATGTCTGGAA	CT	AACT
14q32(II)29	GTGAATACATGTCTGGAA	CT	AACT
14q32(II)22	ATGAATACGTGTTTGGAA	CT	AACT
14q32(II)20	GTGAATACGTCTCTGGAA	CT	AACT
14q32(II)24	ATGAGTACGTATCTGGAA	CT	AACT
14q32(II)6	GTGAATATATGTCTGGAA	CT	AACT
14q32(II)15	ATGAATATGTGTCTGGAA	CT	AACT
14q32(II)30	GTGAATGCCATCTGGAA	CT	AACT
14q32(II)8	GTGAATACATGTTTGA	AACT	
14q32(II)18	ATGAATACATGTCTGA	AACT	
14q32(II)7	ATAAATACATGCC	TGAGAACT	
14q32(II)3	AATACTACGTGTCTGA	AACT	
14q32(II)12	AATAATACGTGTCTGGAA	CT	AACT
14q32(II)14	AATAATACGTGTCTAGAA	CT	AACT
14q32(II)31	AATACTAAGTGCCTGGAA	CT	AACT
14q32(II)4	AATAATGTGTGTCTAGAA	CT	AACT
14q32(II)17	AATGATATGTGTCTGA	AACT	
14q32(II)11	AATATCATGTGTCTGGAA	CT	AACT
14q32(II)13	AACACCATGTTTCTAGAA	CT	AACT
14q32(II)16	AACACCACGTGTCTGGAA	CT	AACT
14q32(II)19	GATGATACGTGTGTGGAA	CT	AACT
14q32(II)2	AACAGTTGTATCTGGAA	AT	
Consensus	atgAat.acgTgTcTggAAcT		

FIG. 5.3 – ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D 14q32(II). Les nucléotides représentés en rouge sont ceux qui sont conservés dans plus de 90 % des séquences alignées, pondérées par le programme Multalin ; en bleu, ceux qui sont conservés dans plus de 50 % des séquences, mais strictement moins que 90 % ; en noir, ceux qui sont conservés dans strictement moins que 50 % des séquences.

	1	10	2022
	-----+-----+		
HBII-52-1	TCATGCTCAATAGGATTACG		
HBII-52-2	TCATGCTCAATAGGATTACG		
HBII-52-3	TCATGCTCAATAGGATTACG		
HBII-52-4	TCATGCTCAATAGGATTACG		
HBII-52-5	TCATGCTCAATAGGATTACG		
HBII-52-6	TCATGCTCAATAGGATTACG		
HBII-52-7	TCATGCTCAATAGGATTACG		
HBII-52-9	TCATGCTCAATAGGATTACG		
HBII-52-10	TCATGCTCAATAGGATTACG		
HBII-52-11	TCATGCTCAATAGGATTACG		
HBII-52-38	TCATGCTCAATAGGATTACG		
HBII-52-12	TCATGCTCAATAGGATTACG		
HBII-52-13	TCATGCTCAATAGGATTACG		
HBII-52-35	TCATGCTCAATAGGATTACG		
HBII-52-34	TCATGCTCAATAGGATTACG		
HBII-52-33	TCATGCTCAATAGGATTACG		
HBII-52-32	TCATGCTCAATAGGATTACG		
HBII-52-14	TCATGCTCAATAGGATTACG		
HBII-52-15	TCATGCTCAATAGGATTACG		
HBII-52-29	TCATGCTCAATAGGATTACG		
HBII-52-16	TCATGCTCAATAGGATTACG		
HBII-52-17	TCATGCTCAATAGGATTACG		
HBII-52-26	TCATGCTCAATAGGATTACG		
HBII-52-18	TCATGCTCAATAGGATTACG		
HBII-52-19	TCATGCTCAATAGGATTACG		
HBII-52-23	TCATGCTCAATAGGATTACG		
HBII-52-22	TCATGCTCAATAGGATTACG		
HBII-52-21	TCATGCTCAATAGGATTACG		
HBII-52-20	TCATGCTCAATAGGATTACG		
HBII-52-25	TCATGCTCAATAGGATTACA		
HBII-52-30	TCATGCTCAATAGGATTACA		
HBII-52-39	TCATGCTCAATAGGATTACA		
HBII-52-36	TAATGCTCAATAGGATTACG		
HBII-52-8	TCATGCTCAATAGGATTATG		
HBII-52-37	TCATGCTCAATAGGATTATG		
HBII-52-40	TCATGTTCAATAGGATTACG		
HBII-52-41	CCATGCTCAATAGGATTACA		
HBII-52-46	TCATGCTCAATAGGATTACG		
HBII-52-31	TCATGCTTATAGGATTACG		
HBII-52-24	TCATTCTCAAAAGGATTATG		
HBII-52-27	TCATTCTCAAAAGGATTATG		
HBII-52-28	TCATTCTCAAAAGGATTATG		
HBII-52-43	TCATGCTTATAGTATTATG		
HBII-52-45	TTAAGTTCAAGTTGGATTACG		
HBII-52-44	TACCATGCTCAAT--GATTACG		
HBII-52-42	TGCTCAATATAGGATTACG		
Consensus	..tcaTgctcAataggATTAcG		

FIG. 5.4 – ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D HBII-52. Convention de couleurs : voir figure 5.3.

	1	10	21
	-----+-----		
HBII-85-1	TACCGTCGTTCTCAGCGGAA		
HBII-85-2	TACCGTCGTTCTCATCGGAA		
HBII-85-3	TACCGTCGTTCTCATCGGAA		
HBII-85-5	TACCGTCGTTCTCATCGGAA		
HBII-85-7	TACCGTCGTTCTCATCGGAA		
HBII-85-4	TACCGTCGTTCTCATCAGAA		
HBII-85-8	TACCGTCATTCTCATCGGAA		
HBII-85-9	TACCGTCATTCTCATCGGAA		
HBII-85-6	TAACGTCATTCTCATCGGAA		
HBII-85-10	GTACCATCATCCTCATCTAA		
HBII-85-13	ATACCATCATCCTCATTGAA		
HBII-85-12	ATACTGTCATCCTCTTCAA		
HBII-85-14	ATACCGTCGTCCTCGTCAA		
HBII-85-17	ATACCGTCATCCTCGTCAA		
HBII-85-15	ATACCGTCATCCTCGTCGAA		
HBII-85-16	ATACCGTCATCCTCGTCGAA		
HBII-85-19	ATACCGTCATCCTCGTCGAA		
HBII-85-20	ATACCGTCATCCTCGTCGAA		
HBII-85-22	ATACCGTCATCCTCGTCGAA		
HBII-85-21	ATACCGTCATCCTCGTCGAA		
HBII-85-23	ATACTGTCATCCTCGTCGAA		
HBII-85-18	ATACCGTCATTCTCGTCGAA		
HBII-85-24	ATACCGTCATCTTCGTTGAA		
HBII-85-11	GTACTATCATCTTAGTTGAA		
Consensus	aTACcgTCaTccTcgTcgaA.		

FIG. 5.5 – ALIGNEMENT DES ÉLÉMENTS ANTISENS DES COPIES DE L'ARN C/D HBII-85. Convention de couleurs : voir figure 5.3.

5.1.2 Séquences utilisées

Les séquences utilisées dans cette analyse sont les séquences de GenBank de numéros d'accèsion :

- ARNpré-r de 45S : U13369, nt 1 à 13350
- U14 : D88010, snoRNA U14-S13-3
- U20 : Z34290
- U21 : NR_000006
- U24 : Z48764
- U45b : X96645
- U47 : X96647
- HBII-13 : NR_001294
- HBII-52, copie n°1 : AF250841
- HBII-85, copie n°2 : NG_002690, nt 289522 à 289618
- HBII-85, copie n°15 : NG_002690, nt 301858 à 301951
- HBII-436 : AY055806
- HBII-437 : AY055807
- HBII-438 : AY055808.

Ainsi que les séquences suivantes :

- 14q32(0)
- 14q32(I), copie n°3
- 14q32(II), copie n°21,

décrites dans Cavaillé *et al.*, 2002.

5.2 Principe des calculs

5.2.1 Probabilité de trouver au moins une complémentarité parfaite dans la séquence-cible

On appellera « séquence-cible » la séquence dans laquelle on recherche des complémentarités aux éléments antisens des ARN C/D (ici : la séquence du précurseur de 45S des ARN ribosomiques humains).

On s'intéresse aux séquences de n nucléotides de long (ce sont des n -uplets d'éléments pris parmi : $\{A; C; G; T\}$). Soit A_i l'événement : « le n -uplet commençant au nucléotide i de la séquence-cible est parfaitement complémentaire à la séquence de n nucléotides en amont de la boîte D de l'ARN C/D ». Soient $snoA$, $snoC$, $snoG$ et $snoT$, respectivement, le nombre de A, de C, de G et de T dans la séquence de n nucléotides en amont de la boîte D de l'ARN C/D (on a donc : $snoA + snoC + snoG + snoT = n$), et p_A , p_C , p_G et p_T , respectivement, les proportions de la séquence-cible en nucléotides A, C, G et T.

A_i est l'intersection des événements : « le nucléotide i est complémentaire au premier nucléotide en amont de la boîte D », « le nucléotide $i + 1$ est complémentaire au deuxième nucléotide en amont de la boîte D », ... , « le nucléotide $i + n - 1$ est complémentaire au n ème nucléotide en amont de la boîte D ».

Si l'on considère ces événements indépendants (c'est à dire qu'on néglige les biais dans la répartition des dinucléotides dans la séquence-cible), la probabilité de l'événement A_i sera égale au produit de leurs probabilités, soit :

$p(A_i) = (p_A)^{snoT} * (p_C)^{snoG} * (p_G)^{snoC} * (p_T)^{snoA}$ qui est indépendant de i (notons N la taille de la séquence-cible, en nucléotides) :

$$\forall i \in \llbracket 1; N \rrbracket \quad p(A_i) = p(A_1) = (p_A)^{snoT} * (p_C)^{snoG} * (p_G)^{snoC} * (p_T)^{snoA}$$

Soit A l'événement : « au moins un des événements A_i est réalisé ».

$$\overline{A} = \bigcap_{i=1}^N \overline{A_i}$$

Considérons les A_i mutuellement indépendants (en réalité, pour la plupart des ARN C/D, tous les A_i ne sont pas indépendants deux à deux : par exemple, si le premier nucléotide de l'élément antisens de l'ARN C/D est différent du deuxième nucléotide de l'élément antisens, alors si A_j est réalisé, A_{j+1} est impossible ; les A_i , pour des valeurs de i proches les unes des autres, ne sont donc pas indépendants, et il faudrait corriger le calcul de $p(A)$ en en tenant compte ; nous allons négliger cette correction, en considérant que, pour un A_i donné, seuls quelques A_j ne sont pas indépendants de A_i , sur les $\approx N$ A_j possibles) ; dans ce cas :

$$p(\overline{A}) = \prod_{i=1}^N p(\overline{A_i})$$

$$p(\overline{A}) = \prod_{i=1}^N p(\overline{A_1})$$

$$p(\overline{A}) = p(\overline{A_1})^N$$

$$p(\overline{A}) = (1 - p(A_1))^N$$

$$p(A) = 1 - (1 - p(A_1))^N$$

$$p(A) = 1 - (1 - (p_A)^{snoT} * (p_C)^{snoG} * (p_G)^{snoC} * (p_T)^{snoA})^N$$

Par exemple, pour l'ARN C/D HBII-52 (copie n°1), dont l'élément antisens a pour séquence :
5' TCATGCTCAATAGGATTACG 3' :

$$n = 5 : p(A_1) = (p_A)^2 * (p_C)^1 * (p_G)^1 * (p_T)^1$$

Étant donnée la composition de la séquence-cible (nucléotides 1 à 13350 de la séquence de numéro d'accèsion : U13369 : 1612 « A », 4746 « C », 4917 « G », et 2075 « T ») :

$$p(A_1) \approx 3,0.10^{-4}$$

$$p(A) \approx 0,98$$

$$n = 6 : p(A_1) = (p_A)^2 * (p_C)^1 * (p_G)^1 * (p_T)^2$$

$$p(A_1) \approx 4,6.10^{-5}$$

$$p(A) \approx 0,46$$

$$n = 7 : p(A_1) = (p_A)^2 * (p_C)^2 * (p_G)^1 * (p_T)^2$$

$$p(A_1) \approx 1,6.10^{-5}$$

$$p(A) \approx 0,19$$

5.2.2 Nombre le plus probable de complémentarités

Appelons Y la variable aléatoire indiquant le nombre d'événements A_i réalisés ; l'événement ($Y = y$) sera donc l'union de tous les événements pour lesquels y événements élémentaires A_i seront réalisés, et $(N - y)$ événements élémentaires A_i ne seront pas réalisés :

$$(Y = y) = [(A_1 \text{ vrai}) \cap (A_2 \text{ vrai}) \cap \dots \cap (A_y \text{ vrai}) \cap (A_{y+1} \text{ faux}) \cap \dots \cap (A_N \text{ faux})] \\ \cup [(A_1 \text{ vrai}) \cap (A_2 \text{ vrai}) \cap \dots \cap (A_{y-1} \text{ vrai}) \cap (A_y \text{ faux}) \cap (A_{y+1} \text{ vrai}) \cap (A_{y+2} \text{ faux}) \cap \dots \cap (A_N \text{ faux})] \\ \cup \dots$$

($Y = y$) est donc l'union de C_N^y événements incompatibles, que nous allons noter : B_j (avec par exemple : $B_1 = [(A_1 \text{ vrai}) \cap (A_2 \text{ vrai}) \cap \dots \cap (A_y \text{ vrai}) \cap (A_{y+1} \text{ faux}) \cap \dots \cap (A_N \text{ faux})]$). On a alors :

$$p(Y = y) = \sum_{j=1}^{C_N^y} p(B_j)$$

Si on néglige les incompatibilités entre événements élémentaires A_i pour des valeurs de i voisines (par exemple, pour la plupart des ARN C/D, A_1 et A_2 sont incompatibles : cf paragraphe 5.2.1), $p(B_j)$ est indépendant de j :

$\forall j \in \llbracket 1; C_N^y \rrbracket$ $p(B_j) = p(B_1) = \prod_{i=1}^y p(A_i) \cdot \prod_{i=y+1}^N p(\overline{A_i})$ (on considère les A_i mutuellement indépendants, c'est à dire qu'on néglige les biais de répartition des dinucléotides dans la séquence-cible)

Ainsi :

$$p(Y = y) = \sum_{j=1}^{C_N^y} p(B_j)$$

$$p(Y = y) = C_N^y p(B_1)$$

$$p(Y = y) = C_N^y \prod_{i=1}^y p(A_i) \cdot \prod_{i=y+1}^N p(\overline{A_i})$$

Or $\forall i \in \llbracket 1; N \rrbracket$ $p(A_i) = p(A_1) = (p_A)^{snoT} * (p_C)^{snoG} * (p_G)^{snoC} * (p_T)^{snoA}$ (cf paragraphe 5.2.1), donc :

$$p(Y = y) = C_N^y \cdot p(A_1)^y \cdot (1 - p(A_1))^{N-y}$$

La variable aléatoire Y suit donc une loi binômiale, de paramètres N et $p(A_1)$.

La valeur de y qui maximise $p(Y = y)$ (c'est à dire : le nombre de complémentarités le plus probable dans la séquence-cible) est recherchée par comparaison de $p(Y = y)$ à $p(Y = y + 1)$:

$$\frac{p(Y=y+1)}{p(Y=y)} = \frac{C_N^{y+1} \cdot p(A_1)^{y+1} \cdot (1-p(A_1))^{N-y-1}}{C_N^y \cdot p(A_1)^y \cdot (1-p(A_1))^{N-y}}$$

$$\frac{p(Y=y+1)}{p(Y=y)} = \frac{N!}{(y+1)!(N-y-1)!} \cdot \frac{y!(N-y)!}{N!} \cdot \frac{p(A_1)}{1-p(A_1)}$$

$$\frac{p(Y=y+1)}{p(Y=y)} = \frac{N-y}{y+1} \cdot \frac{p(A_1)}{1-p(A_1)}$$

Si on appelle $r(y)$ cette valeur $\frac{p(Y=y+1)}{p(Y=y)}$, la valeur de y qui maximise $p(Y = y)$ sera l'entier immédiatement successif à la valeur y_0 telle que : $r(y_0) = 1$. Or :

$r(y) = 1 \iff p(A_1) \cdot (N - y) = (y + 1) \cdot (1 - p(A_1))$ (les deux facteurs, $(y + 1)$ et $[1 - p(A_1)]$, ne peuvent pas s'annuler)

$$r(y) = 1 \iff y \cdot [(1 - p(A_1)) + p(A_1)] = N \cdot p(A_1) - (1 - p(A_1))$$

$$r(y) = 1 \iff y = N \cdot p(A_1) - 1 + p(A_1)$$

$$r(y) = 1 \iff y = (N + 1) \cdot p(A_1) - 1$$

Donc si on note y_{max} le nombre de complémentarités le plus probable dans la séquence-cible : $y_{max} = E[(N + 1) \cdot p(A_1) - 1] + 1$.

Remarque : pour N grand (séquences analysées longues), cette valeur de y , le nombre de complémentarités le plus probable, est proche de l'espérance de la variable aléatoire Y (l'espérance d'une variable aléatoire qui suit une loi binômiale de paramètres N et $p(A_1)$ vaut : $N \cdot p(A_1)$).

La variance de Y , quant à elle, vaut : $N \cdot (1 - p(A_1)) \cdot p(A_1)$, donc son écart-type vaut : $\sigma = \sqrt{N \cdot (1 - p(A_1)) \cdot p(A_1)}$.

5.3 Résultats de l'analyse

Dans les pages suivantes sont présentés les résultats des recherches de complémentarités entre les ARN C/D dont les gènes sont soumis à l'empreinte génomique parentale (ainsi que six contrôles, les snoRNA ubiquitairement exprimés U14, U20, U21, U24, U45b et U47) sur le précurseur de 45S des ARN ribosomiques ; sur ces graphiques sont superposés les résultats attendus, calculés tel que présenté dans la section 5.2.

5.3.1 Probabilité de trouver au moins une complémentarité parfaite dans la séquence-cible

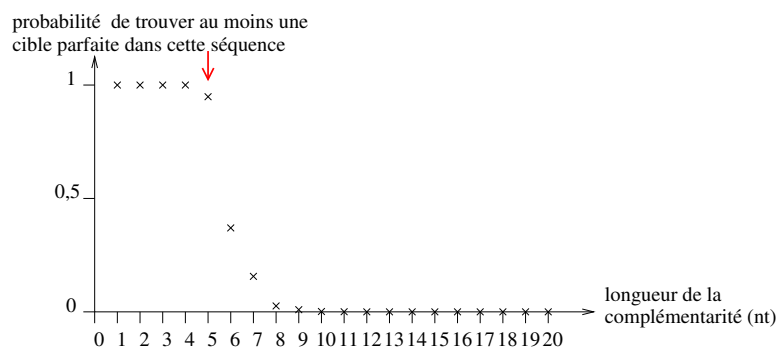
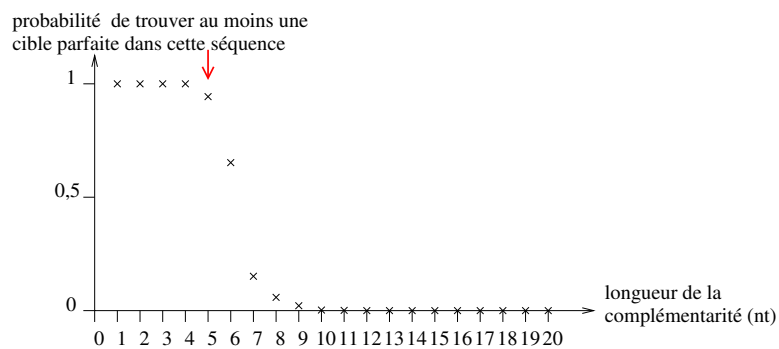
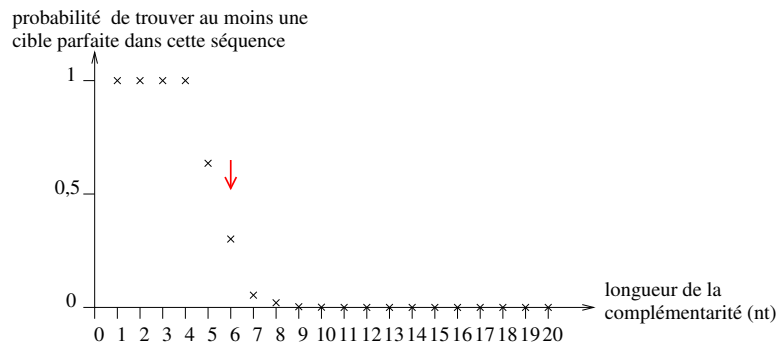
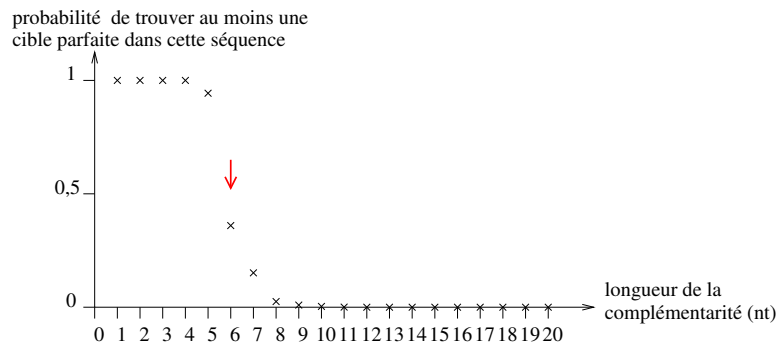
Les graphiques suivants présentent, en ordonnée, la probabilité de trouver au moins une complémentarité parfaite longue de n nucléotides à l'ARNpré-r, en fonction de cette longueur n (en abscisse). L'abscisse de la flèche rouge indique la taille de la plus grande complémentarité effectivement retrouvée.

Pour chaque ARN C/D, quatre graphiques sont donnés : deux pour la recherche de complémentarités à la séquence en amont de la boîte D (en tenant compte du nucléotide immédiatement en amont de la boîte — voir un exemple figure 5.1 — et sans en tenir compte — voir un exemple figure 5.2), et les deux autres, pour la recherche de complémentarités à la séquence en amont de la boîte D' (en tenant compte du nucléotide immédiatement en amont de la boîte, et sans en tenir compte).

Seuls sont représentés les résultats des analyses pour les ARN C/D 14q32(0) et 14q32(I) n°3, qui montrent qu'aucune complémentarité détectée n'est plus longue que les complémentarités statistiquement attendues avec une probabilité raisonnable ; à titre de comparaison, les résultats de la même analyse, sur les séquences des snoRNA U14 et U20 (connus pour participer à la maturation de l'ARN ribosomique), sont également présentés.

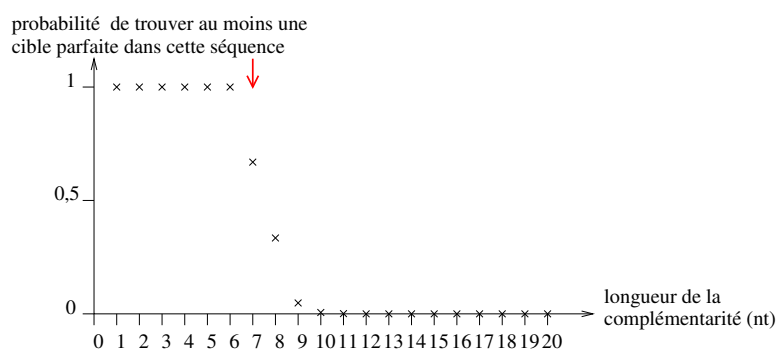
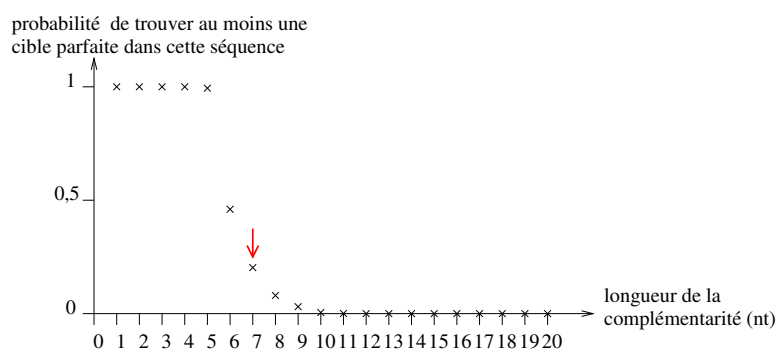
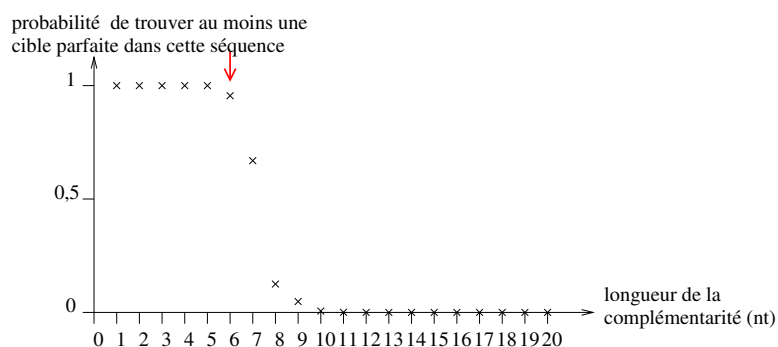
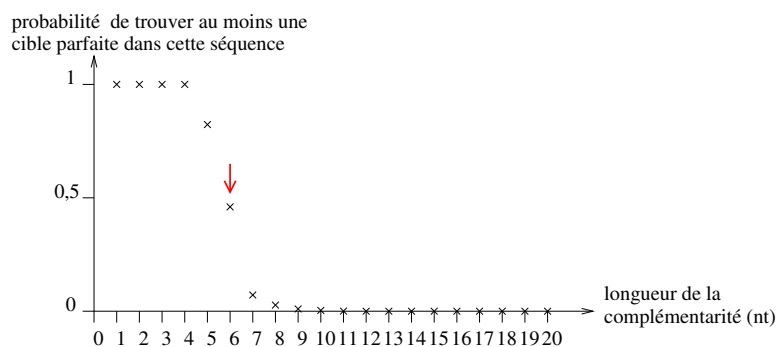
Les résultats des analyses des ARN C/D 14q32(II) n°21, HBII-13, HBII-52 n°1, HBII-85 n°2, HBII-85 n°15, HBII-436, HBII-437 et HBII-438 sont très similaires à ceux des analyses de 14q32(0) et 14q32(I) n°3, et ne sont pas montrés. Parmi les snoRNA U14, U20, U21, U24, U45b et U47, connus pour participer à la maturation de l'ARN ribosomique, seul U47 ne présente pas de complémentarité parfaite qui ne soit pas attendue (la plus longue, un nucléotide en amont de la boîte D, est longue de 7 nt) : la reconnaissance de sa cible par U47 implique une paire G-U (Kiss-László *et al.*, 1996).

ARN C/D 14q32(0)



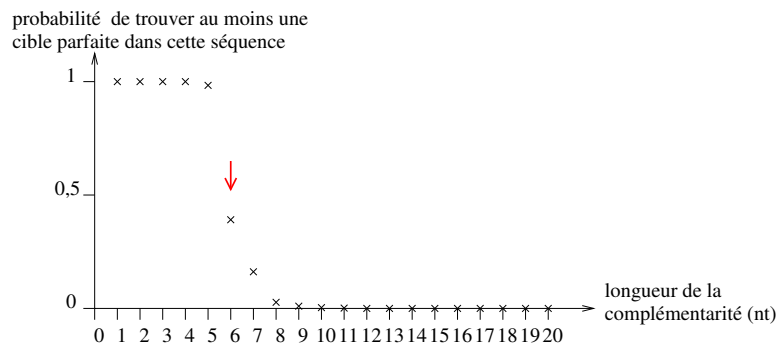
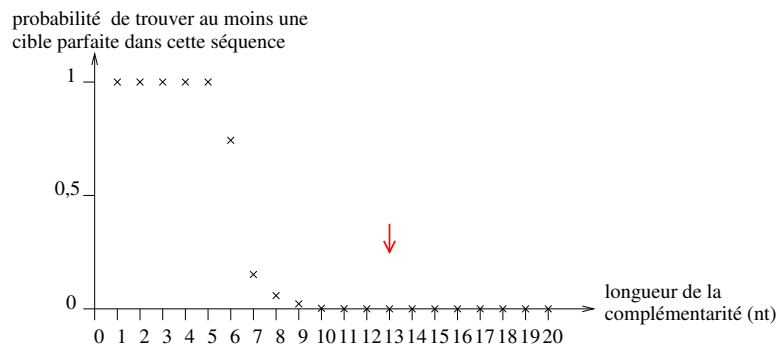
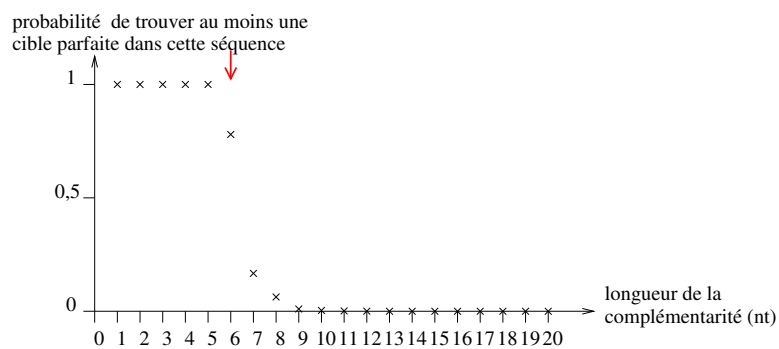
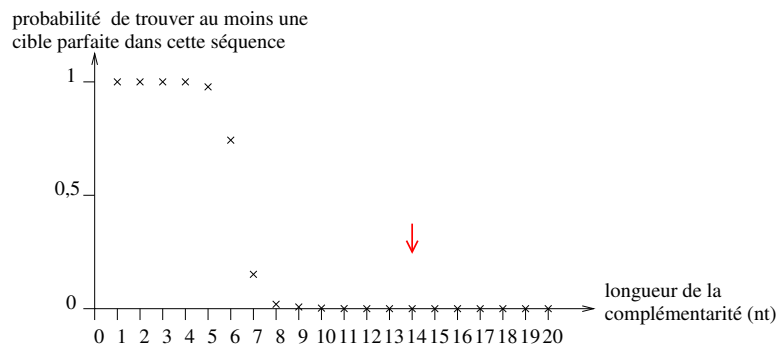
Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. L'abscisse de la flèche rouge indique la taille de la plus grande complémentarité effectivement retrouvée.

ARN C/D 14q32(I) n°3



Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. L'abscisse de la flèche rouge indique la taille de la plus grande complémentarité effectivement retrouvée.

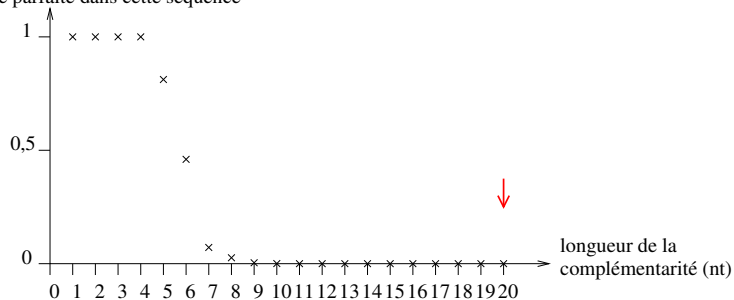
SnoRNA U14



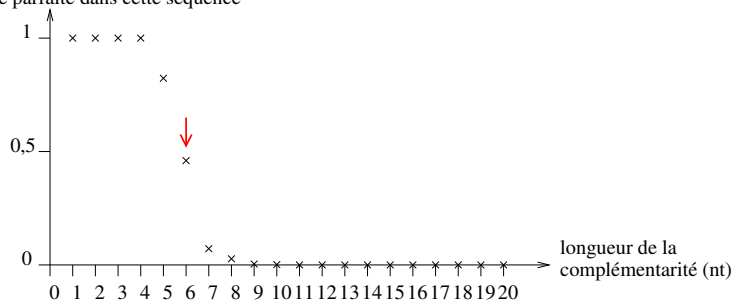
Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. L'abscisse de la flèche rouge indique la taille de la plus grande complémentarité effectivement retrouvée.

SnoRNA U20

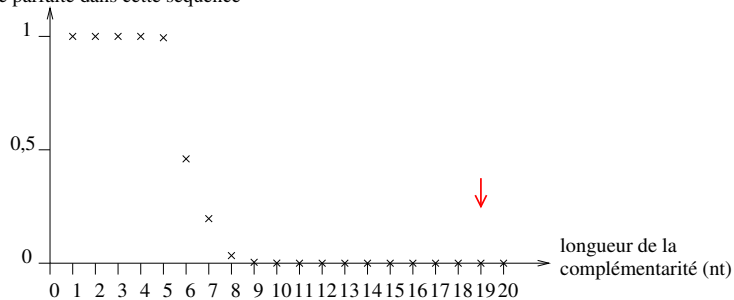
probabilité de trouver au moins une
cible parfaite dans cette séquence



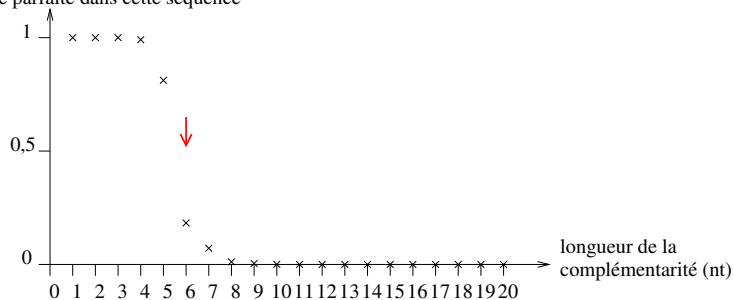
probabilité de trouver au moins une
cible parfaite dans cette séquence



probabilité de trouver au moins une
cible parfaite dans cette séquence



probabilité de trouver au moins une
cible parfaite dans cette séquence



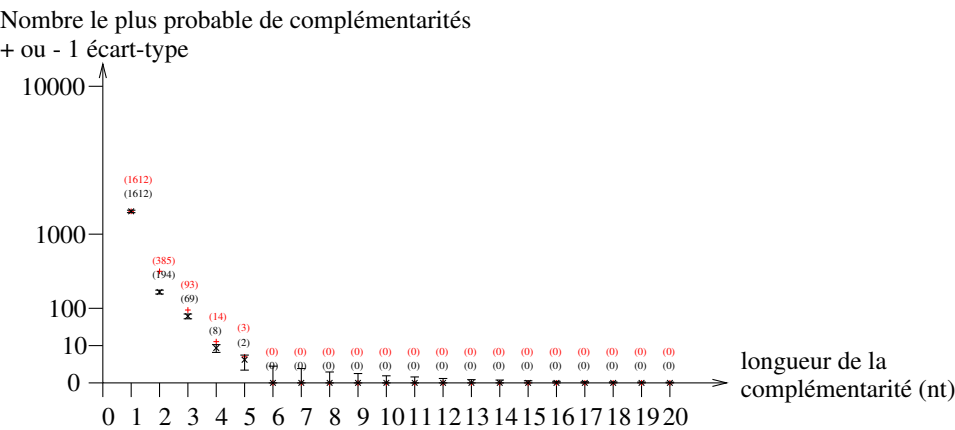
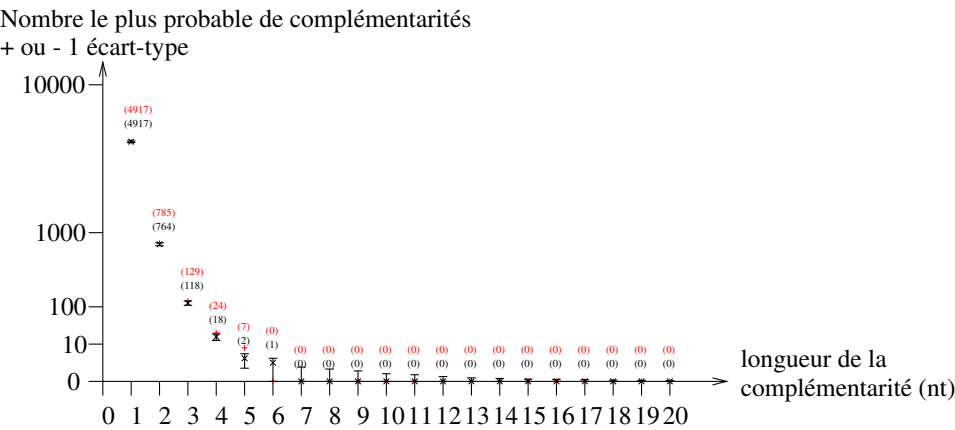
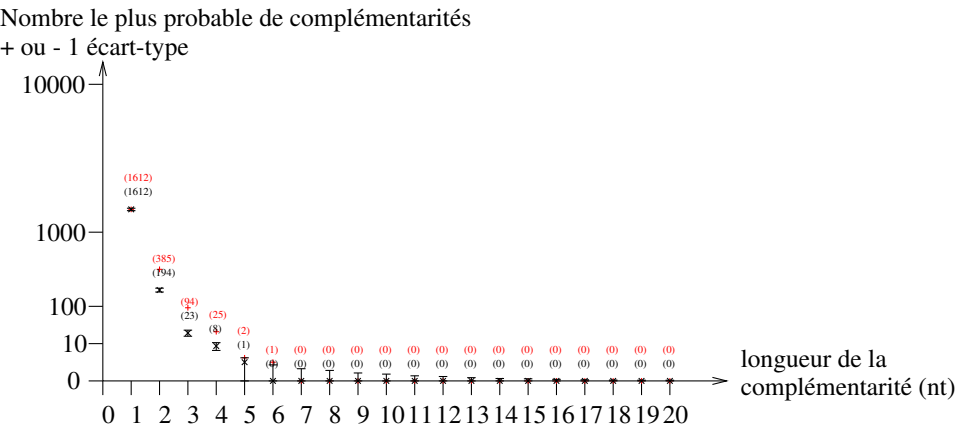
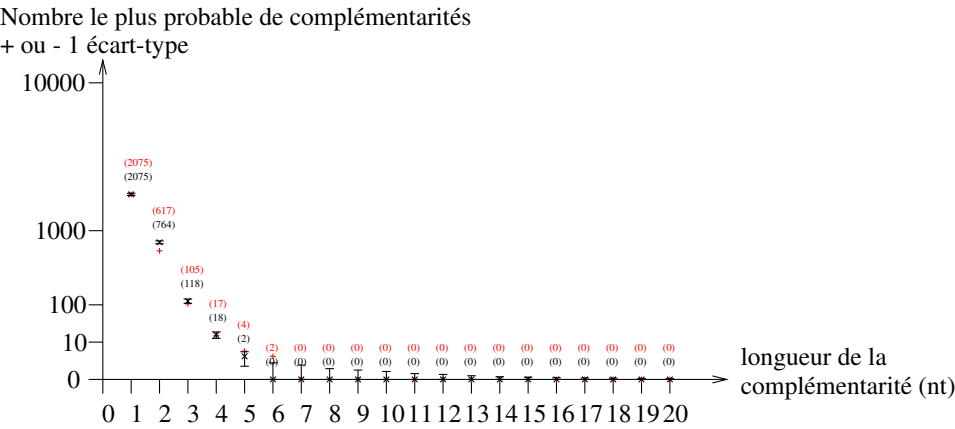
Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. L'abscisse de la flèche rouge indique la taille de la plus grande complémentarité effectivement retrouvée.

5.3.2 Nombre le plus probable de complémentarités

Les graphiques suivants présentent, en ordonnée, le nombre le plus probable de complémentarités parfaites longues de n nucléotides à l'ARNpré-r (croix noires), en fonction de cette longueur n (en abscisse). Les croix rouges représentent les complémentarités parfaites effectivement retrouvées pour l'ARN C/D considéré. Les valeurs « nombre le plus probable de complémentarités - écart-type de la distribution théorique » et « nombre le plus probable de complémentarités + écart-type de la distribution théorique » sont représentées par les deux extrémités des barres noires autour de chaque croix noire. Les nombres entre parenthèses, en noir et en rouge, indiquent respectivement les nombres les plus probables, et les nombres effectivement retrouvés, de complémentarités parfaites.

Pour chaque ARN C/D, quatre graphiques sont donnés : deux pour la recherche de complémentarités à la séquence en amont de la boîte D (en tenant compte du nucléotide immédiatement en amont de la boîte — voir un exemple figure 5.1 — et sans en tenir compte — voir un exemple figure 5.2), et les deux autres, pour la recherche de complémentarités à la séquence en amont de la boîte D' (en tenant compte du nucléotide immédiatement en amont de la boîte, et sans en tenir compte).

ARN C/D 14q32(0)

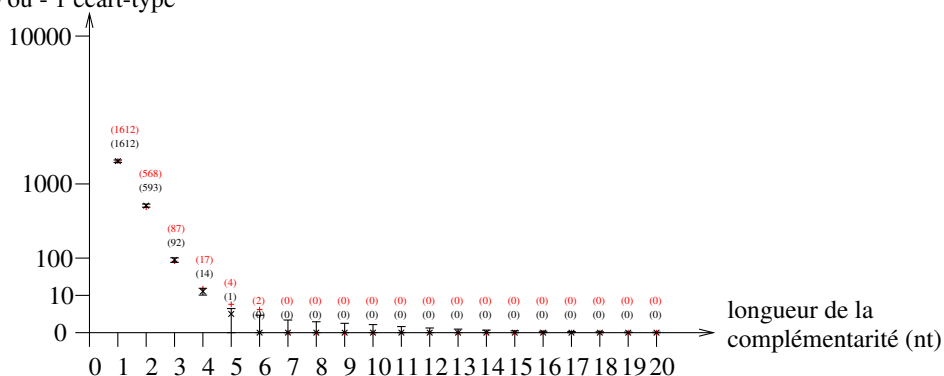


Les croix noires indiquent les nombres les plus probables de complémentarités entre un ARN C/D aléatoire et le précurseur de 45S, en fonction de la longueur de l'appariement, + et - écart-type (voir p. 186 pour plus de détails); les croix rouges indiquent les nombres de complémentarités effectivement trouvées.

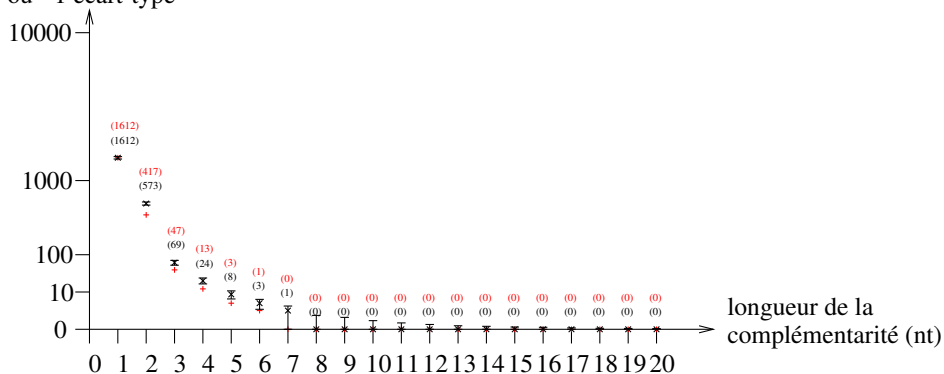
Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. Voir p. 186 pour les détails.

ARN C/D 14q32(I) n°3

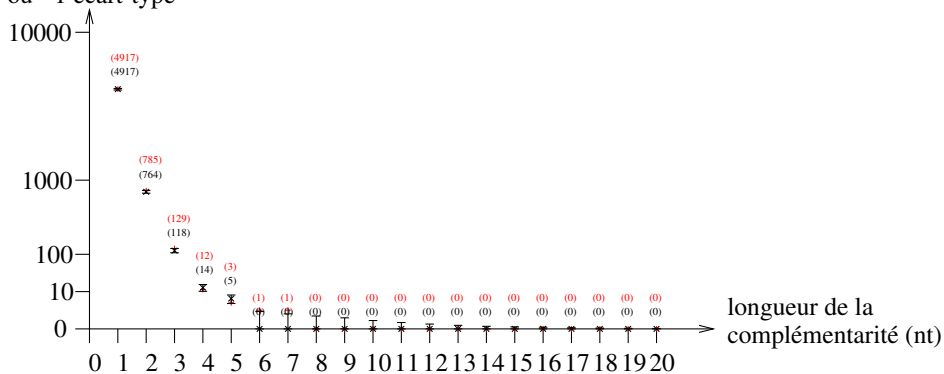
Nombre le plus probable de complémentarités
+ ou - 1 écart-type

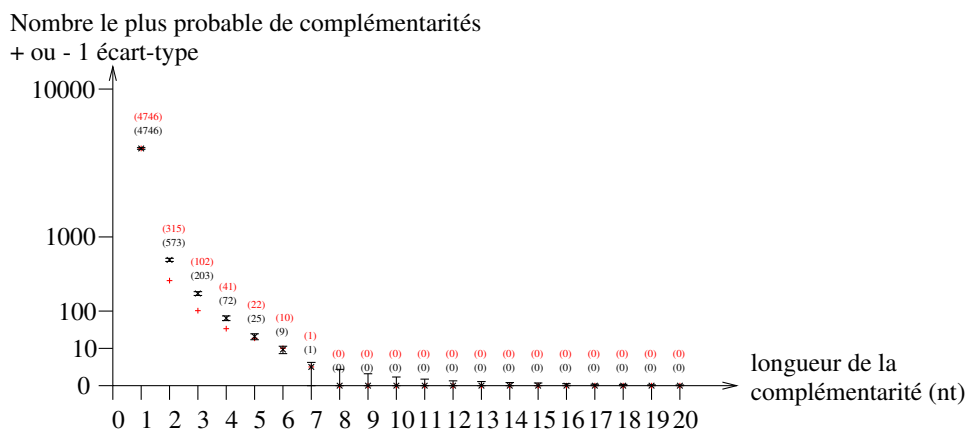


Nombre le plus probable de complémentarités
+ ou - 1 écart-type



Nombre le plus probable de complémentarités
+ ou - 1 écart-type

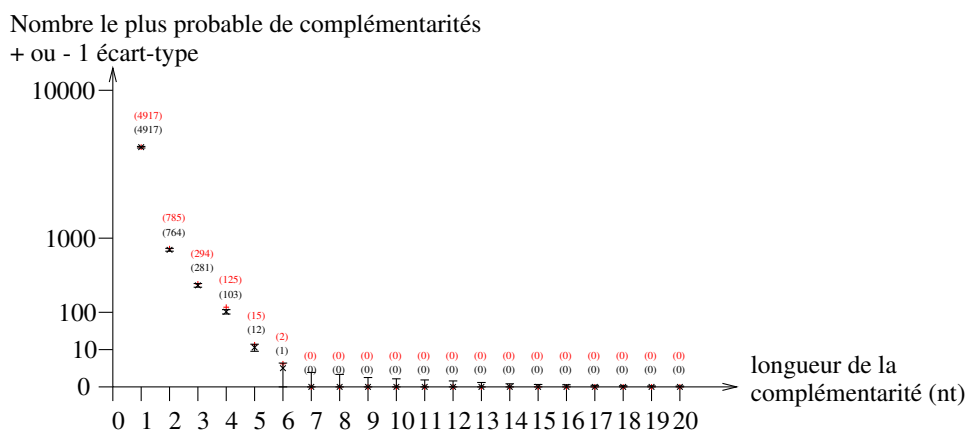
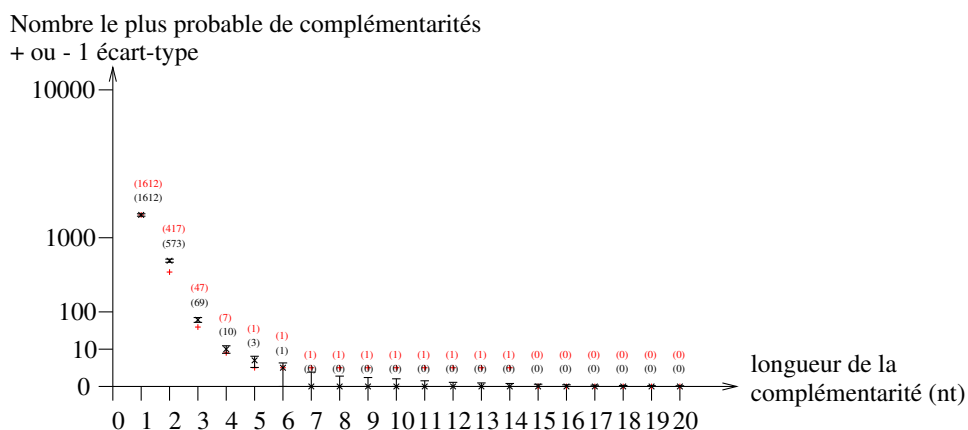




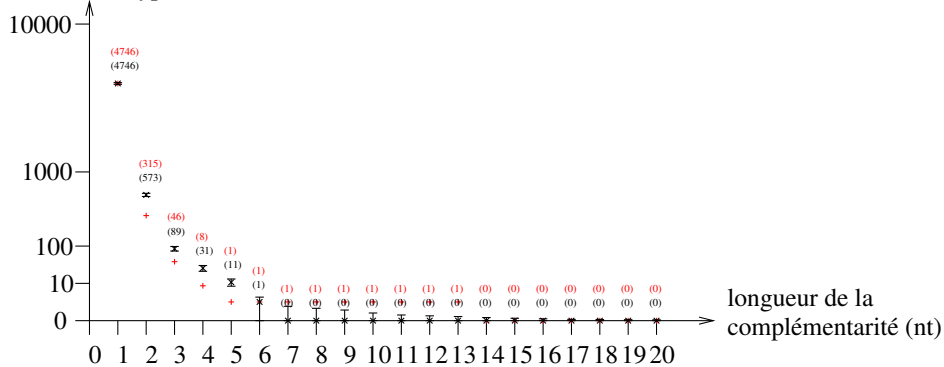
Les croix noires indiquent les nombres les plus probables de complémentarités entre un ARN C/D aléatoire et le précurseur de 45S, en fonction de la longueur de l'appariement, + et - écart-type (voir p. 186 pour plus de détails); les croix rouges indiquent les nombres de complémentarités effectivement trouvées.

Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. Voir p. 186 pour les détails.

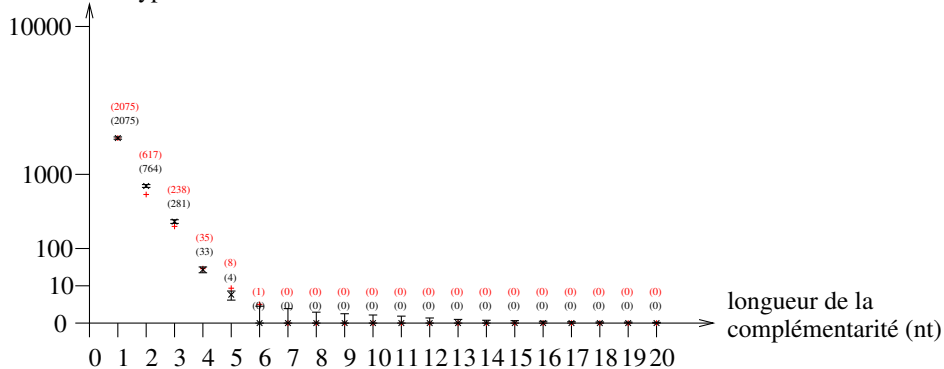
SnoRNA U14



Nombre le plus probable de complémentarités
+ ou - 1 écart-type



Nombre le plus probable de complémentarités
+ ou - 1 écart-type

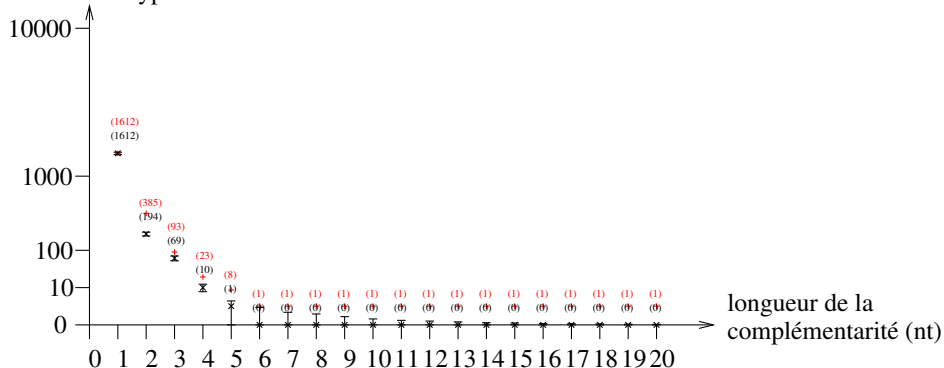


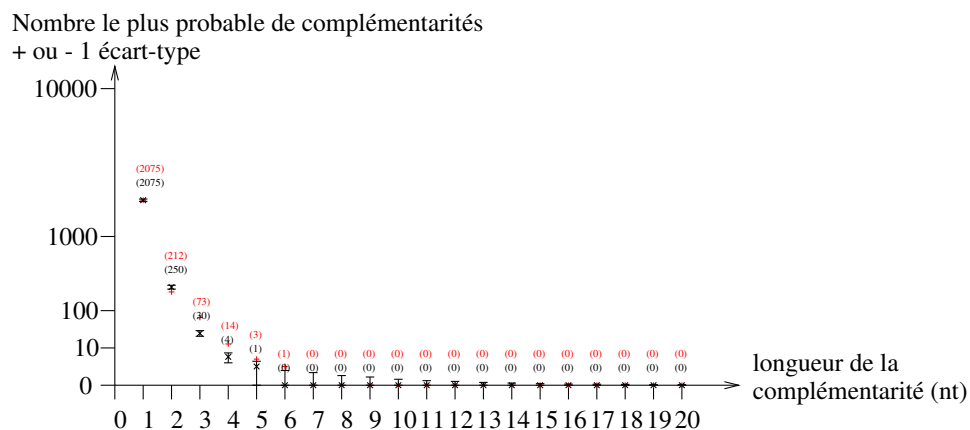
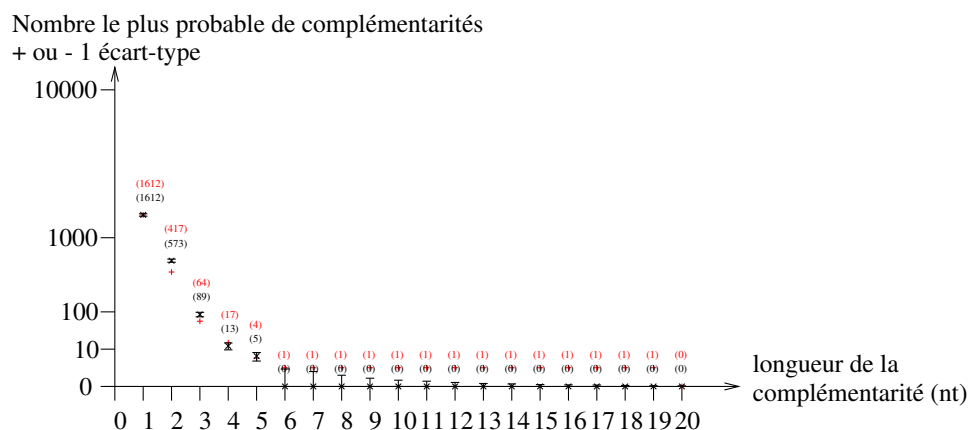
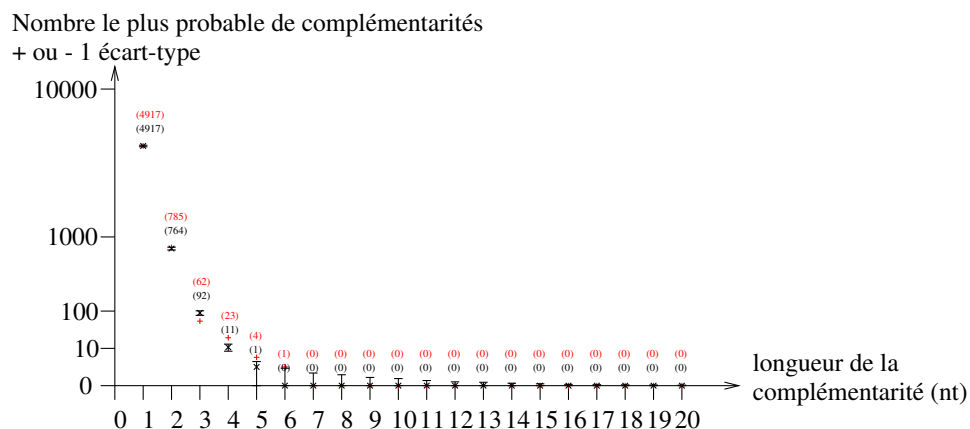
Les croix noires indiquent les nombres les plus probables de complémentarités entre un ARN C/D aléatoire et le précurseur de 45S, en fonction de la longueur de l'appariement, + et - écart-type (voir p. 186 pour plus de détails); les croix rouges indiquent les nombres de complémentarités effectivement trouvées.

Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D ; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D' ; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D ; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. Voir p. 186 pour les détails.

SnoRNA U20

Nombre le plus probable de complémentarités
+ ou - 1 écart-type





Les croix noires indiquent les nombres les plus probables de complémentarités entre un ARN C/D aléatoire et le précurseur de 45S, en fonction de la longueur de l'appariement, + et - écart-type (voir p. 186 pour plus de détails); les croix rouges indiquent les nombres de complémentarités effectivement trouvées.

Légende des quatre courbes : la première : avec le nucléotide immédiatement en amont de la boîte, boîte D; la deuxième : avec le nucléotide immédiatement en amont de la boîte, boîte D'; la troisième : sans le nucléotide immédiatement en amont de la boîte, boîte D; la quatrième : sans le nucléotide immédiatement en amont de la boîte, boîte D'. Voir p. 186 pour les détails.

5.3.3 Complémentarités aux snRNA

J'ai réalisé la même recherche de complémentarités contre les snRNA U1 (numéro d'accèsion : HUMSNU1C1 de GenBank), U2 (numéro d'accèsion : HSU2SNRNA), U4 (numéro

d'accension : HSU4SNRNA), U4atac (numéro d'accension : HSU62822), U5a (numéro d'accension : HSU5ARN), U5b (numéro d'accension : HSRNAU5B), U5b1 (numéro d'accension : HUMUR5B1), U5ds (numéro d'accension : HUMU5DS), U5e (numéro d'accension : HUMU5E), U5f (numéro d'accension : HUMU5F), U6 (numéro d'accension : HSU6SNRNA) et U6atac (numéro d'accension : HSU62823) humains. Sur ces séquences, nettement plus courtes que le précurseur de 45S des ARN_r, les approximations du modèle mathématique présenté dans la section 5.2 ne seraient plus négligeables. J'ai donc uniquement cherché les complémentarités entre les ARN C/D et ces séquences, sans les comparer à des valeurs prédites : aucune complémentarité de 9 nt ou plus n'est détectée entre les séquences situées immédiatement en amont des boîtes D ou D', ou un nucléotide en amont, et les snRNA.

5.3.4 Complémentarités hors de l'élément antisens

Certains snoRNA à boîtes C et D ne guident pas de méthylation, mais participent à des clivages des ARN_{pré-r}, et en dirigent éventuellement le repliement, en s'appariant avec leur ARN-cible par des séquences qui ne sont pas situées immédiatement en amont (ou un nucléotide en amont) de leur boîte D ou D' (voir pages 49 et 51). Les ARN C/D issus des deux loci soumis à l'empreinte génomique parentale pourraient ainsi jouer un rôle de chaperones pour les ARN_r, et altérer la conformation de domaines particuliers.

J'ai donc recherché les complémentarités entre les ARN C/D humains des deux loci soumis à l'empreinte génomique parentale (le locus 15q11-13 et le locus 14q32) et le précurseur de 45S de l'ARN_r humain (numéro d'accension : U13369 de GenBank, nucléotides 1 à 13350), d'au moins 9 nt de long (sans appariement G-U)²; à la différence des complémentarités recensées en 5.3.1 et 5.3.2, les complémentarités recherchées ici ne se limitent pas aux séquences immédiatement en amont (ou un nucléotide en amont) des boîtes D et D' d'ARN C/D représentatifs, elles sont recherchées sur toute la séquence de l'ARN C/D, pour tous les ARN C/D dont les gènes sont situés dans l'un de ces deux loci.

Le tableau 5.1, page 193, rassemble les complémentarités détectées (parmi les nombreuses complémentarités de 9 nt, seules sont représentées celles impliquant un ARN C/D déjà impliqué dans une complémentarité plus longue).

De façon remarquable, onze de ces douze complémentarités³ (et en particulier, les neuf complémentarités d'au moins 10 nt de long) impliquent des séquences des ARN_r matures, alors que celles-ci ne représentent que 53 % de la longueur du précurseur de 45S. Cependant, si ces complémentarités étaient requises pour des repliements de l'ARN_r, on devrait s'attendre à ce qu'un même ARN C/D reconnaisse au moins deux séquences d'une même molécule d'ARN_r, or un seul ARN C/D (HBII-438) remplit cette condition, et les séquences qu'il reconnaît, sur l'ARN de 18S, sont très éloignées sur la structure spatiale de la petite sous-unité du ribosome : l'une est un segment monocaténaire qui relie les hélices H18 et H19, et l'autre est la boucle terminale de l'hélice H42 (Wimberly *et al.*, 2000). Il est donc peu probable que HBII-438 aide au repliement de l'ARN_r de 18S, sauf à imaginer d'importants réarrangements dans la structure de l'ARN_r de 18S après dissociation de l'ARN C/D.

La même recherche de complémentarités entre des séquences quelconques des ARN C/D et les snRNA (voir partie 5.3.3) révèle des complémentarités de 11 nt entre plusieurs isoformes de HBII-52 et les snRNA U2, U5ds et U5e, mais jamais plus d'une par isoforme de HBII-52; comme un rôle de chaperone semble peu crédible sur ces ARN probablement déjà repliés (voir page 52), je n'ai pas poursuivi leur analyse.

²Pour les guides de méthylation, une complémentarité de 8 nt entre un snoRNA modèle et sa cible ne guide pas de 2'-O-méthylation détectable (Cavaillé et Bachellerie, 1998); donc pour les guides de méthylation au moins, les complémentarités plus courtes que 9 nt ne semblent pas fonctionnelles.

³Quatre complémentarités sont identiques : elles impliquent des isoformes de HBII-52 sur une séquence qui

ARN C/D	cible putative	duplex (haut : ARN C/D ; bas : cible)
HBII-438	18S nt 557 à 568	5' GGACUCAUCCA 3' 3' CCUGAGUAAGGU 5'
14q32(II) n°14	18S nt 1560 à 1570	5' CUGCCGGCGUA 3' 3' GACGGCCGCAU 5'
14q32(II) n°14	28S nt 2068 à 2077	5' ACUGCCGGCG 3' 3' UGACGGCCGC 5'
14q32(II) n°31	28S nt 1454 à 1445	5' GCCGGUGGUG 3' 3' CGGCCACCAC 5'
14q32(0)	28S nt 1856 à 1865	5' CCAGUUCUGC 3' 3' GGUCAAGACG 5'
HBII-52 n°14	28S nt 4049 à 4058	5' GGGUCAGUGA 3' 3' CCCAGUCACU 5'
HBII-52 n°20	28S nt 4049 à 4058	5' GGGUCAGUGA 3' 3' CCCAGUCACU5 5'
HBII-52 n°21	28S nt 4049 à 4058	5' GGGUCAGUGA 3' 3' CCCAGUCACU 5'
HBII-52 n°30	28S nt 4049 à 4058	5' GGGUCAGUGA 3' 3' CCCAGUCACU 5'
HBII-52 n°35	28S nt 2818 à 2827	5' UACAUUGUUC 3' 3' AUGUAACAAG 5'
HBII-52 n°38	18S nt 1531 à 1540	5' CAGUCAGUGU 3' 3' GUCAGUCACA 5'
HBII-438	28S nt 4 à 13	5' AUCUGAGGUC 3' 3' UAGACUCCAG 5'
14q32(II) n°14	28S nt 4 à 12	5' UCUGAGGUC 3' 3' AGACUCCAG 5'
HBII-438	18S nt 1618 à 1626	5' GAAUAAUUG 3' 3' CUUAUUAAC 5'
HBII-438	5'ETS nt 830 à 838	5' CUGAGGGAC 3' 3' GACUCCCUG 5'

TAB. 5.1 – COMPLÉMENTARITÉS ENTRE LES ARN C/D DES LOCI HUMAINS 15q11-13 ET 14q32 ET LE PRÉCURSEUR D'ARN_r HUMAIN DE 45S. Elles sont classées par longueur décroissante ; les complémentarités de 9 nt sont très nombreuses : seules sont représentées (sous la double barre horizontale) celles impliquant un ARN C/D déjà impliqué dans une complémentarité plus longue. Dans la première colonne, le nom de l'ARN C/D est complété, pour les ARN C/D à gènes multiples, par le numéro de la copie du gène. Longueurs des séquences : 5'ETS : 3656 nt ; 18S : 1871 nt ; ITS1 : 1095 nt ; 5,8S : 157 nt ; ITS2 : 1155 nt ; 28S : 5035 nt ; 3'ETS : 381 nt.

5.4 Un rôle peu probable dans la biogenèse des ARN_r et snRNA

En conclusion de cette analyse, il apparaît que les ARN C/D transcrits à partir des loci 15q11-13 et 14q32 ne présentent pas de complémentarité parfaite d'au moins 9 nt aux ARN_r ni aux snRNA en amont de leurs boîtes D, ou D', ou un nucléotide en amont (la longueur minimale requise pour le guidage de la méthylation dans un système modèle est de 9 nt : Cavaillé et Bachellerie, 1998). Il semble donc peu probable qu'ils guident une méthylation de l'ARN_r. Ils ne présentent pas non plus de longue complémentarité aux ARN_r ni aux snRNA par des segments quelconques, ce qui semble exclure un rôle de chaperone.

Cependant, de rares exemples de complémentarités fonctionnelles imparfaites sont connus (c'est notamment le cas de la complémentarité du snoRNA U47, dont l'appariement à sa cible implique une paire G-U). Il reste donc possible que ces ARN C/D reconnaissent un précurseur d'ARN_r ou de snRNA par ce type de complémentarité.

leur est commune ; je considère donc qu'elles ne totalisent qu'une complémentarité.

Chapitre 6

Annexe 3

6.1 Article n°6 : revue

Biol. Chem., Vol. 385, pp. 905–911, October 2004 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2004.118

Minireview

Imprinted small RNA genes

Hervé Seitz¹, Hélène Royo¹, Shau-Ping Lin²,
Neil Youngson², Anne C. Ferguson-Smith² and
Jérôme Cavaille^{1,*}

¹Laboratoire de Biologie Moléculaire des Eucaryotes du CNRS, Institut d'Exploration Fonctionnelle des Génomes, 118 route de Narbonne, F-31062 Toulouse Cedex 4, France

²Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

*Corresponding author
e-mail: cavaille@ibcg.biotoul.fr

Abstract

Genomic imprinting is an epigenetic phenomenon that results in differential expression of both alleles, depending on their parent of origin. We have recently identified many imprinted small non-coding RNA genes belonging to the C/D RNA and microRNA gene families, both of which are usually known to play key roles in post-transcriptional metabolism of specific genes (e.g. C/D RNAs guide ribosome methylation of target RNAs while microRNAs elicit either translational repression or RNA interference). Although the functional and evolutionary significance of this association between C/D RNA genes, microRNA genes and genomic imprinting is still highly elusive, these observations provide a framework for further analysis of the potential role of small non-coding RNAs in epigenetic control.

Keywords: genomic imprinting; microRNAs; non-coding RNAs; snoRNAs.

Introduction: Scientific background and state of the art

In mammals, some genes behave differently according to the parental origin of their alleles: either the maternal or the paternal allele is expressed. This phenomenon, called 'genomic imprinting', has been also reported in angiosperms (flowering plants). Several hypotheses have been proposed to explain the emergence of genomic imprinting, according to one of them it would protect females from the development of parthenogenotes (Wilkins and Haig, 2003). A recent work by Kono et al. (2004) showed that genomic imprinting is the sole barrier to parthenogenesis in mammals, as normalization of the expression of imprinted genes (by transgenic manipulation) is sufficient to allow the development of viable parthenogenetic mice. Interestingly, correction of the imprinted expression of the respective genes in those mice leads to a global normalization of genome expres-

sion, as compared to un-manipulated parthenogenetic embryos. This suggests that imprinted genes control the rate of expression of many other genes during embryonic development. Imprinted genes are often clustered into large chromosomal domains, with maternally-expressed genes alongside paternally-expressed ones. As both parental alleles are virtually identical in sequence, their differential expression is not due to a genetic difference but rather to epigenetic regulation. Accordingly, allele-specific patterns for classical epigenetic marks (e.g. DNA methylation and histone tail modifications) are commonly found associated with imprinted domains and regulate imprinted gene activity and repression in *cis* within these domains. The mechanism of differential expression may also involve non-coding RNAs. Indeed, non-coding genes have been reported within imprinted loci, with many of them (but not all) producing antisense RNA to reciprocally imprinted neighboring protein-coding genes (reviewed in Sleutels and Barlow, 2002). Although their functions remain poorly understood, one of them – *Air* at the *Igf2r* locus – seems to be implicated in the bi-directional silencing of several imprinted genes on mouse chromosome 17 (Sleutels et al., 2002).

This review is mainly focused on the recently described imprinted C/D RNA and microRNA (miRNA) genes that map at two distinct human imprinted chromosomal domains: 15q11q13 (mouse 7C) and 14q32 (mouse distal 12). We will briefly summarize how they have been identified through either *in silico* or experimental approaches and then address their potential functions both in terms of target RNAs and epigenetic regulatory pathways. Due to space limits, many original papers of outstanding interest could not be cited and thus, we strongly recommend to readers the following reviews on genomic imprinting: Reik and Walter (2001), Verona et al. (2003); miRNAs/RNA interference (Lai, 2003; Bartel, 2004) and C/D RNAs (Kiss, 2001; Bachellerie et al., 2002).

C/D RNAs direct 2'-O-methylation

Small RNAs of the C/D type (or C/D RNAs) represent one of the largest families of eukaryotic non-coding RNAs, with more than 100 different RNA species described so far. Most of them are involved in the generation of 2'-O-methylation of ribosomal RNAs (rRNAs) and/or spliceosomal RNAs (UsnRNAs). They act through the formation of a specific RNA duplex (10 to 21 nt long) at each modification site in which the methylated nucleotide is systematically paired to the fifth position upstream from the D or D' box (5'-CUGA-3'), a conserved sequence motif found in all the C/D RNAs. Hence, these small RNAs have been also termed RNA methylation guides. Each C/D RNA interacts with a specific set of 4 common core proteins (Nop56, Nop58, Fibrillarin and NHPX) to form a ribonucleoparticle (RNP) which can accumulate either in

906 H. Seitz et al.

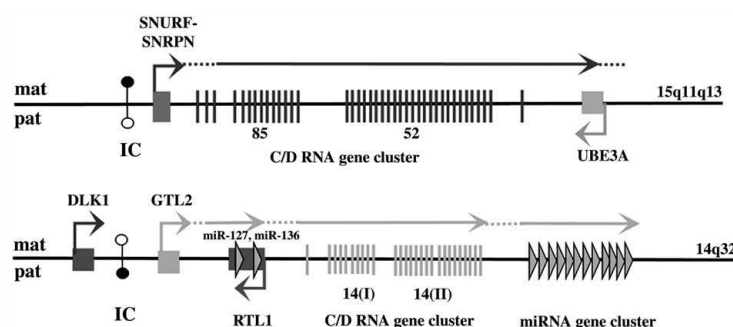


Figure 1 Schematic representation of the imprinted human 15q11q13 (mouse 7C) and 14q32 domains (distal 12). The relative position of several imprinted genes is indicated by squares, vertical bars (C/D RNA genes) or triangles (microRNA genes). Maternally expressed and paternally expressed genes are indicated in light grey and dark grey, respectively. It is noteworthy that imprinted expression of human C/D RNA and miRNA genes at 14q32 has not yet been formally demonstrated. Differentially methylated regions (DMR) acting as imprinting centers are represented by circles (filled: hypermethylated, open: hypomethylated). Poorly characterized small RNA host genes are symbolized by horizontal arrows. Mat: maternal chromosome. Pat: paternal chromosome. IC: imprinting center. Only a portion of each imprinted domain is shown and the picture is not drawn to scale.

the nucleolus (called C/D snoRNAs) or in the Cajal bodies (termed C/D scaRNAs). In the nucleolus, C/D snoRNAs modify both rRNAs and the RNA polymerase III-transcribed U6 spliceosomal RNA, while in the Cajal bodies scaRNAs are believed to direct modifications on the RNA polymerase II-transcribed spliceosomal small RNAs (U1, U2, U4 and U5). One should also keep in mind that many orphan C/D RNAs (devoid of any significant complementarities to pre-rRNA and pre-srRNA) have also been reported. Whether they are able to guide 2'-O-methylation of other cellular RNAs or whether they perform completely different cellular functions remains to be determined. Regarding their biosynthesis, all vertebrate C/D RNA genes described to date are located in introns and appear to result from the exonucleolytic degradation of the host-intron after RNA splicing.

MicroRNAs are key players in RNA-mediated gene silencing phenomena

MicroRNAs are small (ca. 21 to 23 nt) single-stranded RNAs which are processed from larger RNA precursors through two sequential RNA processing steps mediated by at least two RNase-III endonucleases. First, Drosha in the nucleus converts a larger primary miRNA intermediate (pri-miRNA) to a precursor miRNA species (pre-miRNA; an irregular, approx. 70 nt long hairpin structure) from which the mature miRNA is released through the activity of Dicer in the cytoplasm. miRNAs have been implicated in two kinds of post-transcriptional regulation: translational repression and RNA interference (or RNAi). In the translational repression process, miRNAs and their associated proteins (the miRNP complex) hybridize to specific mRNA targets that exhibit imperfect complementarities to the miRNA and inhibit mRNA translation without notably affecting stability. The molecular mechanisms underlying miRNA-directed inhibition of protein production is still poorly known and one should keep in mind that other mechanisms – not directly related to translation inhibition – cannot be formally ruled out (e.g. a rapid cotranslational degradation of the nascent polypeptide). In the RNAi process, miRNAs and their associated proteins (the RISC complex) hybridize to mRNAs

that have perfect or near-perfect complementarity to the miRNA and trigger RNA cleavages. While a few miRNA targets have been isolated or predicted in plants, *C. elegans* and *D. melanogaster*, vertebrate miRNA targets have rarely been formally demonstrated, and on the basis of cotransfection studies and the analysis of Hox mRNA degradation products, Yekta et al. proposed that miR-196a elicits *in vivo* RNAi on the homeobox gene family mRNAs (Yekta et al., 2004). Small RNAs and/or their associated proteins are also involved in chromatin modifications in plants, in the fission yeast *S. pombe*, in the ciliate *Tetrahymena*, and in *Drosophila* (for recent reviews see Grewal and Rice, 2004; Matzke et al., 2004). Whether mammalian miRNAs can also perform such RNA-mediated silencing mechanisms is still an open question.

Paternally expressed C/D RNA genes on human 15q11-13 (murine 7C)

Pioneering RNomics approaches performed by A. Hüttenhofer and J. Brosius led to the identification of many novel C/D RNAs (Hüttenhofer et al., 2001). Surprisingly a few of them – MBII-13, MBII-52 and MBII-85 – were found to be mainly (or exclusively) expressed in the mouse brain, in contrast to all previously known C/D RNAs that are ubiquitously expressed (Cavaillat et al., 2000). In humans, the orthologous genes (called HBII-13, HBII-52 and HBII-85) are clustered at the 15q11q13 locus between the *SNURF-SNRPN* and *UBE3A* genes (Figure 1). This chromosomal domain harbors several imprinted genes (Figure 1) and is implicated in two distinct human clinical disorders, the Prader-Willi syndrome (PWS) and Angelman syndrome (AS). HBII-85 (also called HMR) has been also independently isolated by two other groups (de los Santos et al., 2000; Meguro et al., 2001) and subsequent sequence analysis revealed additional C/D RNA genes (HBII-436, HBII-437, HBII-438A and HBII-438B) within the same genomic interval (Runte et al., 2001). All the C/D RNA genes that have been tested are not expressed in a PWS patient exhibiting a large paternal deletion of the whole imprinted domain nor in a

PWS mouse model, indicating that they are expressed from the paternal allele only (Cavaille et al., 2000; de los Santos et al., 2000).

Almost all of those C/D RNA genes at 15q11q13 are embedded within introns flanked by non-coding exons. Such a basic unit in which the C/D RNA gene is more conserved than the surrounding intronic sequences can be repeated in tandem many times (e.g. 47 and 27 copies of HBII-52 and HBII-85, respectively) (Cavaille et al., 2000; Runte et al., 2001; see Figure 1). Moreover, the C/D RNA gene is the most conserved part of the repeat unit between human and rodents, suggesting biological importance. Very little is known about their biogenesis, but most of them, if not all, are predicted to be processed from a single very long primary transcript (>460 kb) believed to start at the *SNURF-SNRPN* promoter and overlapping the maternally expressed *UBE3A* gene in the antisense orientation (Runte et al., 2001; Figure 1).

These imprinted C/D RNAs do not seem to be involved in rRNA or snRNA biogenesis as they lack long complementarities to their precursors. However, their nucleolar localization and their interaction with the classical C/D snoRNA-associated proteins suggest a function similar to previously known C/D snoRNAs (unpublished data).

H/MBII-52 and the metabolism of serotonin 5-HT_{2c} mRNA

H/MBII-52 contains an 18 nt long complementarity to the brain-specific serotonin receptor 5-HT_{2c} mRNA (Cavaille et al., 2000). This base pairing is not only conserved but also it involves a portion of 5-HT_{2c} mRNA subjected to both alternative RNA splicing and adenine-to-inosine RNA editing at 5 distinct positions (A-E sites). During translation, inosine behaves like guanine and thus RNA editing alters the genetic information encoded within the 5-HT_{2c} mRNA. Furthermore, the serotonin receptor 2C derived from the fully edited mRNA is less active than the genomically encoded receptor (Burns et al., 1997). Surprisingly, the nucleotide believed to be modified by MBII-52 corresponds to one of these edited sites: the C-site. Because it has been reported *in vitro* for model RNA substrates that ribose methylation at the target adenosine hampers dramatically its deamination to inosine (Yi-Brunozzi et al., 1999), it is tempting to speculate that 2'-O-methylation targeted by MBII-52 might negatively regulate RNA editing at the C-site. Accordingly, *in vivo* experiments demonstrated that a 2'-O-methylation targeted by a C/D RNA can efficiently inhibit A-to-I deamination (unpublished data). The C/D RNA antisense element lies at only 13 nt away from a cryptic 5' donor splice site and therefore a role of H/MBII-52 to control alternative splicing of 5-HT_{2c} pre-mRNA (e.g. through steric hindrance) cannot be formally excluded. If H/MBII-52 fine-tunes RNA editing and/or RNA splicing, it could constitute the first example of a novel family of post-transcriptional regulators for protein-coding genes.

H/MBII-85 and the Prader-Willi syndrome

PWS is a complex neurodevelopmental disorder characterized by neonatal hypotonia, failure to thrive during infancy and hyperphagia in early childhood leading to severe obesity. All the known genetic abnormalities pre-

dict that the lack of expression of one (or several) paternally expressed gene(s) encoded at 15q11q13 results in PWS (Nicholls and Knepper, 2001). In the mouse, a paternally inherited deletion from the *Snrpn* to *Ube3a* genes causes hypotonia and growth retardation with 80% lethality before weaning (Tsai et al., 1999) while a deletion from *lpw* to *Ube3a* does not (Johnson et al., 1995). Taken together, this strongly suggests that the candidate PWS gene(s) lies between the *Snrpn* and *lpw* genes. Studies from human PWS patients with rare balanced translocations also defined a minimal critical region mapping at the same relative position within 15q11q13 domain (Wirth et al., 2001; Gallagher et al., 2002). The only conserved genes mapping within this approx. 121 kb long minimal interval are the HBII-85 repeated RNA genes (and also HBII-438A), supporting a biological role for these C/D RNAs in the etiology of PWS (Cavaille et al., 2000; de los Santos et al., 2000; Meguro et al., 2001; Wirth et al., 2001; Gallagher et al., 2002). We assume that these C/D small RNAs function as RNA guides to modify other cellular RNAs, presumably in a tissue-specifically and/or developmentally regulated manner. However, identifying the cognate cellular RNA targets and understanding the biochemical pathways in which they might be involved, is an extremely challenging task. While *in silico* searches can help identify H/MBII-85 RNA targets they clearly will not be sufficient since 10–11 bp long complementarity is known to efficiently direct 2'-O-methylation (Cavaille and Bachellerie, 1998) and thus the number of false positives is very large (unpublished data).

Maternally expressed C/D RNA and miRNA genes on human 14q32/murine distal 12

The screening of a specific C/D RNA library made from a rat brain led to the identification of RBII-36, another brain-specific C/D RNA (Cavaille et al., 2001). Surprisingly, RBII-36 also exhibits a repetitive gene organization highly reminiscent of that previously described at the PWS locus. Indeed, RBII-36 is intron-encoded within *Bsr*, a non-coding RNA gene mapping at 6q32 consisting of roughly 100 tandem repeats of similar 0.9 kb long sequences (Komine et al., 1999; Cavaille et al., 2001). A computer-assisted approach designed to search for repeated sequences at the human and mouse regions of syntenic homology (imprinted 14q32 and distal 12 domain in human and mouse, respectively) allowed the identification of a large number of tandemly repeated C/D RNA genes located downstream from *GTL2*, a large and spliced maternally expressed non-coding RNA gene (Cavaille et al., 2002). In humans, most of them are arranged into two tandem arrays of 31 and 9 gene copies of related sequences [14q(I) and 14q(II), respectively]. Again, all these C/D RNA genes are intron-located within poorly characterized non-coding RNA genes (called MEG8 and *Rian*, in human and mouse, respectively) which are expressed mainly within the brain. In the mouse, these C/D RNA genes are subject to genomic imprinting with an expression restricted to the maternally inherited chromosome. Like the PWS C/D RNA genes, the function of maternally expressed C/D RNAs is still

unknown as they lack significant region of complementarity to rRNA or snRNA. In contrast to C/D RNA genes at the PWS locus, C/D RNA genes at 14q32/distal 12 are not highly conserved between mammalian species; for example, RBII-36 is a rat-specific small RNA not found in mouse, human or sheep (Cavaillat et al., 2001). Moreover, even within each cluster of related C/D RNA genes, substantial nucleotide mutations within the antisense elements (e.g. sequences upstream from the D/D'-boxes) are commonly observed indicating that these RNAs are rapidly evolving molecules. The lack of selective pressure for conservation of the antisense element sequences might suggest that several C/D RNAs are functionally redundant or may interact with multiple different RNA targets. Alternatively, they might also exert a molecular function distinct from the classical methylation guide activities.

C/D RNA genes are not the only small non-coding RNAs encoded at the imprinted *Dlk1-Gtl2* domain as many potential miRNA genes (ca. 46) have been also recently identified through an *in silico* search combined with systematic experimental detection (Seitz et al., 2003, 2004). Most of them reside in a large cluster of closely related sequences mapping approx. 30 kb downstream from the C/D RNA gene cluster. Again, they are only expressed from the maternally inherited chromosome with strongest expression in the adult brain. Imprinted expression of all these small RNA genes is controlled by a proximal intergenic differentially methylated region (IG-DMR) located between the *Dlk1* and *Gtl2* genes (Takada et al., 2002). Indeed, we have demonstrated that a targeted deletion of the unmethylated maternal IG-DMR results in repression of all maternally expressed non-coding RNA genes (e.g. *Gtl2*, C/D RNAs and miRNAs) with concomitant bi-allelic expression of all the paternally expressed protein coding genes (*Dlk1*, *Rtl1* and *Dio3*). The same deletion transmitted paternally has no obvious effect (Lin et al., 2003). Because all the C/D RNA and miRNA genes are expressed from the same DNA strand, in the same orientation, and because they seem to exhibit nearly the same expression pattern, we cannot rule out the possibility that they are all processed from a single, polycistronic transcription unit, which may also include the *Gtl2* gene as well.

A potential link between RNAi, genomic imprinting and retrotransposon silencing?

Remarkably, two miRNAs – miR-127 and miR-136 – are processed from a large maternally expressed antisense RNA to *Rtl1* (Seitz et al., 2003). *Rtl1* (also called *PEG11* in sheep and human) is a paternally expressed gene coding for a conserved large open reading frame exhibiting strong homology to the retrotransposon Gal and Pol domains (Charlier et al., 2001b; Lynch and Tristem, 2003; N. Youngson and A.C. Ferguson-Smith, unpublished). Because these two miRNAs are perfectly complementary to *Rtl1* mRNA, they are predicted to trigger RNA cleavages within *Rtl1* mRNA. Consistent with this assumption, when the two miRNAs are no longer expressed due to the maternal IG-DMR deletion, the level of expression of the *Rtl1* gene is higher than expected upon reactivation of the maternal allele (four- to five-fold for *Rtl1* versus

two-fold as observed for *Dlk1* and *Dio3* genes), suggesting that those miRNAs normally negatively regulate *Rtl1* expression, perhaps through RNA interference (Lin et al., 2003). The sense/antisense gene organization together with the reciprocal imprinting of the miRNAs relative to their putative RNA target gene is also very puzzling, suggesting also an involvement of the small RNAs in *Rtl1* genomic imprinting. This could be achieved either by a preferential nuclear RNA degradation from the *Rtl1* maternal allele only. However, in the IG-DMR deletion, the overexpression of *Rtl1* is roughly equivalent on both parental chromosomes suggesting that potential negative regulation by miRNAs does not distinguish the two alleles. Alternatively or additionally, miRNAs might also trigger chromatin epigenetic changes specifically at the maternal *Rtl1* gene, consistent with the involvement of the RNAi machinery in heterochromatin formation in *S. pombe* (Grewal and Rice, 2004). Current work is in progress to distinguish between these hypotheses.

Imprinted miRNAs and the Callipyge phenomenon

In contrast to miR-127 and miR-136, none of the other imprinted miRNA genes exhibit perfect complementarity to a cellular mRNA, suggesting that they probably inhibit mRNA translation rather than triggering mRNA degradation. Given the importance of both imprinted genes and miRNAs in development, it is likely that these imprinted miRNAs play a pivotal role during mouse ontogeny. Embryos with paternal uniparental disomy for chromosome 12 (patUPD12; embryos with two chromosomes 12 originating from their father) die before the end of gestation and have multiple developmental abnormalities (Georgiades et al., 2000), strongly suggesting that expression of one (or several) maternally expressed non-coding RNA genes is required for correct embryonic development. The miRNA genes are expressed throughout the developing mouse embryo as well as in the placenta and thus the lack of their expression in patUPD12 could account for at least some of these defects.

Insights to the putative functions of miRNAs can also be derived from the analysis of the 'Callipyge phenotype' in the sheep. This phenotype (a muscular hypertrophy of the hindquarters) is linked to a single point mutation (A to G) between the *DLK1* and *GTL2* genes at the orthologous imprinted ovine locus (Freking et al., 2002; Smit et al., 2003). Transmission rules for this phenotype (called polar overdominance) are very intriguing, as only heterozygous animals which acquired the CLPG mutation from their sire express the phenotype. Heterozygous animals, with a maternally inherited mutation and also, surprisingly, homozygous individuals, are normal. This suggests a *trans* effect in which the abnormal maternal chromosome can somehow abrogate the effect of the mutation on the paternal chromosome. At the molecular level, the CLPG mutation does not affect the imprinting status but rather increases gene expression in *cis*: *DLK1* and *RTL1* gene expression is increased in $+^{mat}/CLPG^{pat}$ individuals while *GTL2*, *ANTI-RTL1* gene and *MEG8* (the C/D RNA host gene) are upregulated in a $CLPG^{mat}/+^{pat}$ genetic background (Charlier et al., 2001a). Consistent with this, the hypothesis that has been suggested is that (i) overexpression of a (some) paternally expressed

gene(s) might lead to the Callipyge phenotype and (ii) overexpression of a (some) maternally expressed *trans*-acting repressor gene(s) in CLPG^{pat}/CLPG^{mat} would counteract the effect of the paternal overexpression, thus leading to a wild-type phenotype in CLPG^{pat}/CLPG^{mat} individuals (Georges et al., 2003). Given the known roles of microRNAs in RNA mediated-gene silencing mechanisms, they represent obvious candidates for playing such a role and their potential involvement is currently being tested.

In conclusion, human 15q11q13 and 14q32 domains, and their rodent counterparts, share common structural and functional features: they are characterized by the presence of many imprinted small RNA genes which (i) are mainly expressed in a tissue-specific manner with strongest expression in the adult brain, (ii) are processed from large transcription units spanning many tandemly repeated C/D snoRNA genes (at 14q32 and 15q11q13 loci) and microRNA genes (only at 14q32), and (iii) are devoid of full complementarity to any cellular RNA (except in the case of H/MBII-52, miR-136 and miR-127). In addition, these poorly characterized transcription units also give rise not only to the mature small RNAs, but also to a complex spectrum of unspliced and spliced larger non-coding RNAs, the functions of which, if any, are still unknown.

Genomic imprinting and small non-coding RNA genes: an enigmatic association

The repeated organization of the imprinted C/D RNA genes at these two loci appear to be unique as *in silico* searches in the whole human genome did not reveal any other such concentration of putative C/D RNA genes (unpublished data). Is this observed association between genomic imprinting and repeated small RNA gene clusters fortuitous or does it have a biological or evolutionary significance? Curiously, we noted that most of the imprinted C/D RNA genes at both loci have nearly the same 5'-3' terminal stem structure, suggesting that some of these imprinted C/D RNAs might have evolved from a common C/D RNA ancestor gene, presumably arising by retrotransposition (Cavaille et al., 2002). Accordingly, many modification guide RNA retrogenes, including imprinted C/D RNA genes, have been detected in mammalian genomes (Bachelier et al., 2002; and unpublished data). Thus, one could imagine that imprinted loci – perhaps due to a specific chromatin structure in the germ line – might represent 'hot spots' for C/D RNA gene integration allowing subsequent gene amplification by still unknown mechanisms. The identification of several intronless genes at the mouse PWS locus indeed suggests that this genomic region might be prone to retrotransposition (Chai et al., 2001). Thus, in such a scenario the monoallelic expression of the small RNA genes can be considered simply as a consequence of the widespread imprinting at these two loci.

Small RNA gene clusters might have occurred before the emergence of genomic imprinting and might have even played a role in its establishment and/or maintenance. In the light of the recent discovery of the involve-

ment of non-coding RNAs in epigenetic regulations (Sleutels et al., 2002; Grewal and Rice, 2004), it is tempting to speculate that expression of these non-coding RNA genes (the small RNAs or their host genes) might contribute to some aspects of genomic imprinting control. Consistent with this proposal, the expression of the distal part of the C/D RNA gene cluster at the 7C domain in the mouse is believed to silence *in cis* expression of the paternal *Ube3A* gene (Rougeulle et al., 1998; Chamberlain and Brannan, 2001; Runte et al., 2001). Moreover, at the mouse distal 12 domain, the lack of expression of all the maternally expressed non-coding RNA genes is correlated with the activation of the normally repressed *Dlk1*, *Rtl1* and *Dio3* maternal alleles (Lin et al., 2003). Alternatively, we can not exclude that active transcription of these large pieces of non-coding DNA sequences – and not RNA itself – could somehow also influence expression of the adjacent imprinted protein-coding genes (e.g. the passage of RNA polymerase through regulatory *cis*-acting elements might interfere with gene expression). Indeed, a role of intergenic transcription in gene regulation has been already documented (Gribnau et al., 2000; Martens et al., 2004).

The huge number of repetitive small RNA genes should be also taken into consideration. Among the numerous hypotheses which have been proposed to explain the evolution of imprinting, one is based on the assumption that genomic imprinting might be seen as an adaptation of the host defence mechanisms that DNA methylation plays against (retro-) transposon mobilization (Barlow, 1993; Bird, 1997; Yoder et al., 1997; Martienssen, 1998; Whitelaw and Martin, 2001). Interestingly, C/D RNAs can behave to some extent like mobile elements as their host genes can differ between distant species. The same might also be true for miRNA genes as several of them are highly related in their sequence and can be found at distinct chromosomal loci. In addition, at least in plant systems, it has been suggested that some miRNAs might derive from transposable elements (Mette et al., 2002). Therefore, one can speculate that amplification of repeated C/D RNA genes and/or miRNA genes may have been perceived by the host cell as 'foreign DNA', thus imposing fortuitously epigenetic regulation of the neighboring protein-coding genes. Finally, the finding that miR-127 and miR-136 have the potential to silence an imprinted retrotransposon-like gene might also nicely illustrate another potential link between RNAi machinery, genomic imprinting and retrotransposon silencing (Seitz et al., 2003).

All those small RNA genes were relatively poorly conserved during evolution. They seem to be present only in eutherian mammals as we and other failed to detect them in chicken or even in marsupials (de los Santos et al., 2000; Cavaille et al., 2001; Meguro et al., 2001; and unpublished data). Given that genomic imprinting might be restricted to marsupials and eutherian mammals (Kilian et al., 2000), comparative sequence analysis at non-eutherian mammalian loci together with characterization of their imprinted status should be extremely informative to provide further insights into the potential significance of this enigmatic association between repeated small RNA genes and genomic imprinting.

910 H. Seitz et al.

In conclusion, although an important role in post-transcriptional regulation (e.g. RNA modification, translation repression and RNA interference) is expected for the imprinted small RNAs, we hypothesize that either the small RNAs and/or their repeated host genes could also contribute to some aspects of genomic imprinting, either at the establishment steps during development or at the maintenance steps in somatic differentiated cells. Further *in vivo* analysis of these small RNAs (e.g. knock-out mice models) will contribute to our understanding of the evolution and the gene regulation network operating at imprinted loci.

Acknowledgments

We apologize to colleagues whose work has not been mentioned due to space limitations. We also wish to thank our friends and colleagues Jean-Pierre Bachelier, Marie-Line Bortolin, Alexander Hüttenhofer, Jürgen Brosius, Patrice Vitali, Carole Charlier and Michel Georges who participated to the work presented in this article. This work is mainly supported by grants from an A.C.I. 'Biologie Cellulaire Moléculaire et Structurale' and by laboratory funds from the Centre National de la Recherche Scientifique and Université Paul-Sabatier, Toulouse (to J.C.).

References

- Bachelier, J. P., Cavaille, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* 84, 775–790.
- Barlow, D. P. (1993). Methylation and imprinting: from host defense to gene regulation? *Science* 260, 309–310.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Bird, A. (1997). Does DNA methylation control transposition of selfish elements in the germline? *Trends Genet.* 13, 469–472.
- Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., and Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 387, 303–308.
- Cavaille, J., and Bachelier, J. P. (1998). SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.* 26, 1576–1587.
- Cavaille, J., Buiting, K., Kieffmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., Bachelier, J. P., Brosius, J., and Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA* 97, 14311–14316.
- Cavaille, J., Vitali, P., Basyuk, E., Hüttenhofer, A., and Bachelier, J. P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.* 276, 26374–26383.
- Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A. C., and Bachelier, J. P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* 11, 1527–1538.
- Chai, J. H., Locke, D. P., Ohta, T., Grealia, J. M., and Nicholls, R. D. (2001). Retrotransposed genes such as *Frat3* in the mouse chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm. Genome* 12, 813–821.
- Chamberlain, S. J., and Brannan, C. I. (2001). The Prader-Willi syndrome imprinting center activates the paternally expressed murine *Ube3a* antisense transcript but represses paternal *Ube3a*. *Genomics* 73, 316–322.
- Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., and Georges, M. (2001a). The callipyge mutation enhances the expression of coregulated imprinted genes in *cis* without affecting their imprinting status. *Nat. Genet.* 27, 367–369.
- Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., Shay, T., Weissenbach, J., Cockett, N., Gyapay, G., and Georges, M. (2001b). Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (*clpg*) locus and identification of six imprinted transcripts: *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*. *Genome Res.* 11, 850–862.
- de los Santos, T., Schweizer, J., Rees, C. A., and Francke, U. (2000). Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from *PWCR1*, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *Am. J. Hum. Genet.* 67, 1067–1082.
- Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L., and Smith, T. P. (2002). Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* 12, 1496–1506.
- Gallagher, R. C., Pils, B., Albalwi, M., and Francke, U. (2002). Evidence for the role of *PWCR1/HBII-85* C/D box small nucleolar RNAs in Prader-Willi syndrome. *Am. J. Hum. Genet.* 71, 669–678.
- Georges, M., Charlier, C., and Cockett, N. (2003). The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet.* 19, 248–252.
- Georgiades, P., Watkins, M., Surani, M. A., and Ferguson-Smith, A. C. (2000). Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* 127, 4719–4728.
- Grewal, S. I., and Plice, J. C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* 16, 230–238.
- Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human β -globin locus. *Mol. Cell* 5, 377–386.
- Hüttenhofer, A., Kieffmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachelier, J. P., and Brosius, J. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.* 20, 2943–2953.
- Johnson, D. K., Stubbs, L. J., Culliat, C. T., Montgomery, C. S., Russell, L. B., and Rinchik, E. M. (1995). Molecular analysis of 36 mutations at the mouse pink-eyed dilution (*p*) locus. *Genetics* 141, 1563–1571.
- Killian, J. K., Byrd, J. C., Jirtle, J. V., Munday, B. L., Stoskopf, M. K., MacDonald, R. G., and Jirtle, R. L. (2000). *M6P/IGF2R* imprinting evolution in mammals. *Mol. Cell* 5, 707–716.
- Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 20, 3617–3622.
- Komine, Y., Tanaka, N. K., Yano, R., Takai, S., Yuasa, S., Shiroyoshi, T., Tsuchiya, K., and Yamamori, T. (1999). A novel type of non-coding RNA expressed in the rat brain. *Brain Res. Mol. Brain Res.* 66, 1–13.
- Lai, E. C. (2003). microRNAs: runts of the genome assert themselves. *Curr. Biol.* 13, R925–936.
- Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A. C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nat. Genet.* 35, 97–102.

- Lynch, C., and Tristem, M. (2003). A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. *Curr. Biol.* 13, 1518–1523.
- Martens, J. A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429, 571–574.
- Martienssen, R. (1998). Transposons, DNA methylation and gene control. *Trends Genet.* 14, 263–264.
- Matzke, M., Aufsatz, W., Kanno, T., Daxinger, L., Papp, I., Mette, M. F., and Matzke, A. J. (2004). Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim. Biophys. Acta* 1677, 129–141.
- Meguro, M., Mitsuya, K., Nomura, N., Kohda, M., Kashiwagi, A., Nishigaki, R., Yoshioka, H., Nakao, M., Oishi, M., and Oshimura, M. (2001). Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes. *Hum. Mol. Genet.* 10, 383–394.
- Mette, M. F., van der Winden, J., Matzke, M., and Matzke, A. J. (2002). Short RNAs can identify new candidate transposable element families in *Arabidopsis*. *Plant Physiol.* 130, 6–9.
- Nicholls, R. D., and Knepper, J. L. (2001). Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.* 2, 153–175.
- Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32.
- Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L., and Lalande, M. (1998). An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat. Genet.* 19, 15–16.
- Runte, M., Hüttenhofer, A., Gross, S., Kiefmann, M., Horthemke, B., and Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* 10, 2687–2700.
- Seitz, H., Youngson, N., Lin, S. P., Dalbert, S., Paulsen, M., Bachelier, J. P., Ferguson-Smith, A. C., and Cavaille, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* 34, 261–262.
- Seitz, H., Royo, H., Bortolin, M. L., Lin, S. P., Ferguson-Smith, A. C., and Cavaille, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.* 14, 1741–1748.
- Sleutels, F., and Barlow, D. P. (2002). The origins of genomic imprinting in mammals. *Adv. Genet.* 46, 119–163.
- Sleutels, F., Zwart, R., and Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813.
- Smit, M., Segers, K., Carrascosa, L. G., Shay, T., Baraldi, F., Gyapay, G., Snowden, G., Georges, M., Cockett, N., and Charlier, C. (2003). Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* 163, 453–456.
- Takada, S., Paulsen, M., Tevendale, M., Tsai, C. E., Kelsey, G., Cattanaach, B. M., and Ferguson-Smith, A. C. (2002). Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum. Mol. Genet.* 11, 77–86.
- Tsai, T. F., Jiang, Y. H., Bressler, J., Armstrong, D., and Beaudet, A. L. (1999). Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader-Willi syndrome. *Hum. Mol. Genet.* 8, 1357–1364.
- Verona, R. I., Mann, M. R., and Bartolomei, M. S. (2003). Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu. Rev. Cell. Dev. Biol.* 19, 237–259.
- Whitelaw, E., and Martin, D. I. (2001). Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* 27, 361–365.
- Wilkins, J. F., and Haig, D. (2003). What good is genomic imprinting: the function of parent-specific gene expression. *Nat. Rev. Genet.* 4, 359–368.
- Wirth, J., Back, E., Hüttenhofer, A., Nothwang, H. G., Lich, C., Gross, S., Menzel, C., Schinzel, A., Kioschis, P., Tommerup, N. et al. (2001). A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF-SNRPN transcription unit on chromosome 15. *Hum. Mol. Genet.* 10, 201–210.
- Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304, 594–596.
- Yi-Brunozzi, H. Y., Easterwood, L. M., Kamilar, G. M., and Beal, P. A. (1999). Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2. *Nucleic Acids Res.* 27, 2912–2917.
- Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340.

Bibliographie

- Accardo, M. C., Giordano, E., Riccardo, S., Digilio, F. A., Iazzetti, G., Calogero, R. A., et Furia, M. (2004). A computational search for box C/D snoRNA genes in the *D. melanogaster* genome. *Bioinformatics*, Sous presse.
- Adhya, S. et Gottesman, M. (1982). Promoter occlusion : transcription through a promoter may inhibit its activity. *Cell*, 29(3) :939–944.
- Agris, P. F. (1996). The importance of being modified : roles of modified nucleosides and Mg^{2+} in RNA structure and function. *Prog Nucleic Acid Res Mol Biol*, 53 :79–129.
- Ainscough, J. F., John, R. M., Barton, S. C., et Surani, M. A. (2000). A skeletal muscle-specific mouse *Igf2* repressor lies 40 kb downstream of the gene. *Development*, 127(18) :3923–3930.
- Akhtar, A. et Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell*, 5(2) :367–375.
- Al-Kaff, N. S., Covey, S. N., Kreike, M. M., Page, A. M., Pinder, R., et Dale, P. J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science*, 279(5359) :2113–2115.
- Albrecht, U., Sutcliffe, J. S., Cattanach, B. M., Beechey, C. V., Armstrong, D., Eichele, G., et Beaudet, A. L. (1997). Imprinted expression of the murine Angelman syndrome gene, *Ube3a*, in hippocampal and Purkinje neurons. *Nat Genet*, 17(1) :75–78.
- Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*. *Cell*, 57(1) :49–57.
- Ambros, V. et Horvitz, H. R. (1987). The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev*, 1(4) :398–414.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., et Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol*, 13(10) :807–818.
- Amrein, H. et Axel, R. (1997). Genes expressed in neurons of adult male *Drosophila*. *Cell*, 88(4) :459–469.
- Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H., et Vance, V. B. (1998). A viral suppressor of gene silencing in plants. *Proc Natl Acad Sci USA*, 95(22) :13079–13084.
- Aravin, A. A., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., et Gvozdev, V. A. (2004). Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol Cell Biol*, 24(15) :6742–6750.
- Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., et Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev Cell*, 5(2) :337–350.
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., et Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol*, 11(13) :1017–1027.

- Arney, K. L. (2003). *H19* and *Igf2* — enhancing the confusion? *Trends Genet*, 19(1) :17–23.
- Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J., et Matzke, M. (2002). RNA-directed DNA methylation in *Arabidopsis*. *Proc Natl Acad Sci USA*, 99(Suppl 4) :16499–16506.
- Avner, P., Prissette, M., Arnaud, D., Courtier, B., Cecchi, C., et Heard, E. (1998). Molecular correlates of the murine *Xce* locus. *Genet Res*, 72(3) :217–224.
- Ayesh, S., Matouk, I., Schneider, T., Ohana, P., Laster, M., Al-Sharef, W., De-Groot, N., et Hochberg, A. (2002). Possible physiological role of H19 RNA. *Mol Carcinog*, 35(2) :63–74.
- Aza-Blanc, P., Cooper, C. L., Wagner, K., Batalov, S., Deveraux, Q. L., et Cooke, M. P. (2003). Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol Cell*, 12(3) :627–637.
- Bachellerie, J.-P. et Cavaillé, J. (1998). *Modification and Editing of RNA*, chapitre Small nucleolar RNAs guide the ribose methylations of eukaryotic rRNAs, pages 255 – 272. ASM Press.
- Bachellerie, J.-P., Cavaillé, J., et Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, 84(8) :775–790.
- Bachellerie, J.-P., Cavaillé, J., et Qu, L.-H. (2000). *The Ribosome : structure, function, antibiotics, and cellular interactions*, chapitre 17, pages 191 – 203. ASM Press. « Nucleotide modifications of eucaryotic rRNAs : the world of small nucleolar RNA guides revisited ».
- Bachellerie, J.-P., Nicoloso, M., Qu, L. H., Michot, B., Caizergues-Ferrer, M., Cavaillé, J., et Renalier, M.-H. (1995). Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem Cell Biol*, 73(11-12) :835–843.
- Balakireva, M. D., Shevelyov, Y. Y., Nurminsky, D. I., Livak, K. J., et Gvozdev, V. A. (1992). Structural organization and diversification of Y-linked sequences comprising *Su(Ste)* genes in *Drosophila melanogaster*. *Nucleic Acids Res*, 20(14) :3731–3736.
- Banerjee, A. K. (1980). 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiol Rev*, 44(2) :175–205.
- Barlow, D. P. (1993). Methylation and imprinting : from host defense to gene regulation? *Science*, 260(5106) :309–310.
- Barlow, D. P. (1997). Competition — a common motif for the imprinting mechanism? *EMBO J*, 16(23) :6899–6905.
- Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K., et Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. *Nature*, 349(6304) :84–87.
- Barr, M. L. (1959). Sex chromatin and phenotype in man. *Science*, 130 :679–685.
- Bartel, B. et Bartel, D. P. (2003). MicroRNAs : at the root of plant development? *Plant Physiol*, 132(2) :709–717.
- Bartel, D. P. (2004). MicroRNAs : genomics, biogenesis, mechanism, and function. *Cell*, 116(2) :281–297.
- Bartel, D. P. et Chen, C. Z. (2004). Micromanagers of gene expression : the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet*, 5(5) :396–400.
- Bartolomei, M. S., Webber, A. L., Brunkow, M. E., et Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse *H19* gene. *Genes Dev*, 7(9) :1663–1673.

- Bartolomei, M. S., Zemel, S., et Tilghman, S. M. (1991). Parental imprinting of the mouse *H19* gene. *Nature*, 351(6322) :153–155.
- Bashaw, G. J. et Baker, B. S. (1995). The *msl-2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. *Development*, 121(10) :3245–3258.
- Baulcombe, D. C. (1996a). Mechanisms of pathogen-derived resistance to viruses in transgenic plants. *Plant Cell*, 8(10) :1833–1844.
- Baulcombe, D. C. (1996b). RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol Biol*, 32(1-2) :79–88.
- Beaudet, A. L. et Jiang, Y. H. (2002). A rheostat model for a rapid and reversible form of imprinting-dependent evolution. *Am J Hum Genet*, 70(6) :1389–1397.
- Bell, A. C. et Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*, 405(6785) :482–485.
- Beltrame, M. et Tollervey, D. (1995). Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J*, 14(17) :4350–4356.
- Berghmans, S., Segers, K., Shay, T., Georges, M., Cockett, N., et Charlier, C. (2001). Breakpoint mapping positions the callipyge gene within a 450-kilobase chromosome segment containing the *DLK1* and *GTL2* genes. *Mamm Genome*, 12(2) :183–185.
- Bernard, P., Maure, J.-F., Partridge, J. F., Genier, S., Javerzat, J.-P., et Allshire, R. C. (2001). Requirement of heterochromatin for cohesion at centromeres. *Science*, 294(5551) :2539–2542.
- Bernstein, E., Caudy, A. A., Hammond, S. M., et Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818) :363–366.
- Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., et Hannon, G. J. (2003). Dicer is essential for mouse development. *Nat Genet*, 35(3) :215–217.
- Bidwell, C. A., Kramer, L. N., Perkins, A. C., Hadfield, T. S., Moody, D. E., et Cockett, N. E. (2004). Expression of *PEG11* and *PEG11AS* transcripts in normal and callipyge sheep. *BMC Biol*, 2(1) :17.
- Bielinska, B., Blaydes, S. M., Buiting, K., Yang, T., Krajewska-Walasek, M., Horsthemke, B., et Brannan, C. I. (2000). *De novo* deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. *Nat Genet*, 25(1) :74–78.
- Billy, E., Brondani, V., Zhang, H., Müller, U., et Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc Natl Acad Sci USA*, 98(25) :14428–14433.
- Boccaccio, I., Glatt-Deeley, H., Watrin, F., Roeckel, N., Lalande, M., et Muscatelli, F. (1999). The human *MAGEL2* gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region. *Hum Mol Genet*, 8(13) :2497–2505.
- Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C., et Allis, C. D. (2002). Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet*, 30(1) :73–76.
- Borovjagin, A. V. et Gerbi, S. A. (2004). *Xenopus* U3 snoRNA docks on pre-rRNA through a novel base-pairing interaction. *RNA*, 10(6) :942–953.
- Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H. F., Avner, P., et Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature*, 351(6324) :325–329.

- Bosher, J. M., Dufourcq, P., Sookhareea, S., et Labouesse, M. (1999). RNA interference can target pre-mRNA : consequences for gene expression in a *Caenorhabditis elegans* operon. *Genetics*, 153(3) :1245–1256.
- Bourc'his, D. et Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004) :96–99.
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., et Bestor, T. H. (2001). Dnmt3l and the establishment of maternal genomic imprints. *Science*, 294(5551) :2536–2539.
- Bowman, A. B., Levorse, J. M., Ingram, R. S., et Tilghman, S. M. (2003). Functional characterization of a testis-specific DNA binding activity at the *H19/Igf2* imprinting control region. *Mol Cell Biol*, 23(22) :8345–8351.
- Brandeis, M., Kafri, T., Ariel, M., Chaillet, J. R., McCarrey, J., Razin, A., et Cedar, H. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J*, 12(9) :3669–3677.
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., et Cohen, S. M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*, 113(1) :25–36.
- Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W., et Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J*, 17(22) :6739–6746.
- Brockdorff, N. (2002). X-chromosome inactivation : closing in on proteins that bind *Xist* RNA. *Trends Genet*, 18(7) :352–358.
- Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., et Rastan, S. (1991). Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. *Nature*, 351(6324) :329–331.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., et Rastan, S. (1992). The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, 71(3) :515–526.
- Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., et Willard, H. F. (1991a). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349(6304) :38–44.
- Brown, C. J. et Grealley, J. M. (2003). A stain upon the silence : genes escaping X inactivation. *Trends Genet*, 19(8) :432–438.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., et Willard, H. F. (1992). The human *XIST* gene : analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, 71(3) :527–542.
- Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., Ledbetter, D. H., Levy, E., Craig, I. W., et Willard, H. F. (1991b). Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature*, 349(6304) :82–84.
- Brown, C. J. et Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature*, 368(6467) :154–156.
- Brown, J. W., Echeverria, M., et Qu, L. H. (2003). Plant snoRNAs : functional evolution and new modes of gene expression. *Trends Plant Sci*, 8(1) :42–49.
- Bucher, E., Sijen, T., De Haan, P., Goldbach, R., et Prins, M. (2003). Negative-strand tospoviruses and tenuiviruses carry a gene for a suppressor of gene silencing at analogous genomic positions. *J Virol*, 77(2) :1329–1336.

- Buiting, K., Barnicoat, A., Lich, C., Pembrey, M., Malcolm, S., et Horsthemke, B. (2001). Disruption of the bipartite imprinting center in a family with Angelman syndrome. *Am J Hum Genet*, 68(5) :1290–1294.
- Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., et Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, 387(6630) :303–308.
- Caffarelli, E., Arese, M., Santoro, B., Fragapane, P., et Bozzoni, I. (1994). In vitro study of processing of the intron-encoded U16 small nucleolar RNA in *Xenopus laevis*. *Mol Cell Biol*, 14(5) :2966–2974.
- Caffarelli, E., Fatica, A., Prislei, S., De Gregorio, E., Fragapane, P., et Bozzoni, I. (1996). Processing of the intron-encoded U16 and U18 snoRNAs : the conserved C and D boxes control both the processing reaction and the stability of the mature snorna. *EMBO J*, 15(5) :1121–31.
- Cahill, N. M., Friend, K., Speckmann, W., Li, Z. H., Terns, R. M., Terns, M. P., et Steitz, J. A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J*, 21(14) :3816–3828.
- Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G., et Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet*, 21(1) :103–107.
- Canton, H., Emeson, R. B., Barker, E. L., Backstrom, J. R., Lu, J. T., Chang, M. S., et Sanders-Bush, E. (1996). Identification, molecular cloning, and distribution of a short variant of the 5-hydroxytryptamine_{2C} receptor produced by alternative splicing. *Mol Pharmacol*, 50(4) :799–807.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., et Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298(5595) :1039–1043.
- Cao, X., Aufsatz, W., Zilberman, D., Mette, M. F., Huang, M. S., Matzke, M., et Jacobsen, S. E. (2003). Role of the *DRM* and *CMT3* methyltransferases in RNA-directed DNA methylation. *Curr Biol*, 13(24) :2212–2217.
- Carlson, L. L., Page, A. W., et Bestor, T. H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos : implications for genomic imprinting. *Genes Dev*, 6(12B) :2536–2541.
- Carmell, M. A. et Hannon, G. J. (2004). RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol*, 11(3) :214–218.
- Carmell, M. A., Xuan, Z., Zhang, M. Q., et Hannon, G. J. (2002). The Argonaute family : tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, 16(21) :2733–2742.
- Cattanach, B. M. et Beechey, C. V. (1990). Autosomal and X-chromosome imprinting. *Dev Suppl*, pages 63–72.
- Cattanach, B. M. et Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, 315(6019) :496–498.
- Caudy, A. A., Ketting, R. F., Hammond, S. M., Denli, A. M., Bathoorn, A. M., Tops, B. B., Silva, J. M., Myers, M. M., Hannon, G. J., et Plasterk, R. H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. *Nature*, 425(6956) :411–414.
- Caudy, A. A., Myers, M., Hannon, G. J., et Hammond, S. M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev*, 16(19) :2491–2496.

- Cavaillé, J. et Bachellerie, J.-P. (1998). SnoRNA-guided ribose methylation of rRNA : structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res*, 26(7) :1576–1587.
- Cavaillé, J., Buiting, K., Kieffmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., Bachellerie, J.-P., Brosius, J., et Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci USA*, 97(26) :14311–14316.
- Cavaillé, J., Nicoloso, M., et Bachellerie, J.-P. (1996). Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, 383(6602) :732–5.
- Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A. C., et Bachellerie, J.-P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet*, 11(13) :1527–1538.
- Cavaillé, J., Vitali, P., Basyuk, E., Hüttenhofer, A., et Bachellerie, J.-P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a non-coding RNA gene in rats. *J Biol Chem*, 276(28) :26374–26383.
- Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K., et Gingeras, T. R. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, 116(4) :499–509.
- Cerrato, F., Dean, W., Davies, K., Kagotani, K., Mitsuya, K., Okumura, K., Riccio, A., et Reik, W. (2003). Paternal imprints can be established on the maternal *Igf2-H19* locus without altering replication timing of DNA. *Hum Mol Genet*, 12(23) :3123–3132.
- Cervelli, M., Cecconi, F., Giorgi, M., Annesi, F., Oliverio, M., et Mariottini, P. (2002). Comparative structure analysis of Vertebrate U17 small nucleolar RNA (snoRNA). *J Mol Evol*, 54(2) :166–179.
- Chai, J. H., Locke, D. P., Ohta, T., Grealis, J. M., et Nicholls, R. D. (2001). Retrotransposed genes such as *Frat3* in the mouse chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm Genome*, 12(11) :813–821.
- Chalfie, M., Horvitz, H. R., et Sulston, J. E. (1981). Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell*, 24(1) :59–69.
- Chalker, D. L. et Yao, M. C. (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol Cell Biol*, 16(7) :3658–3667.
- Chamberlain, S. J. et Brannan, C. I. (2001). The Prader-Willi syndrome imprinting center activates the paternally expressed murine *Ube3a* antisense transcript but represses paternal *Ube3a*. *Genomics*, 73(3) :316–322.
- Chan, S. W., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., et Jacobsen, S. E. (2004). RNA silencing genes control de novo DNA methylation. *Science*, 303(5662) :1336.
- Charalambous, M., Menheniott, T. R., Bennett, W. R., Kelly, S. M., Dell, G., Dandolo, L., et Ward, A. (2004). An enhancer element at the *Igf2/H19* locus drives gene expression in both imprinted and non-imprinted tissues. *Dev Biol*, 271(2) :488–497.
- Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., et Georges, M. (2001a). The callipyge mutation enhances the expression of coregulated imprinted genes in *cis* without affecting their imprinting status. *Nat Genet*, 27(4) :367–369.

- Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., Shay, T., Weisenbach, J., Cockett, N., Gyapay, G., et Georges, M. (2001b). Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (*clpg*) locus and identification of six imprinted transcripts : *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*. *Genome Res*, 11(5) :850–862.
- Chaudhuri, S. et Messing, J. (1994). Allele-specific parental imprinting of *dazl*, a posttranscriptional regulator of zein accumulation. *Proc Natl Acad Sci USA*, 91(11) :4867–4871.
- Chédin, F., Lieber, M. R., et Hsieh, C. L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proc Natl Acad Sci USA*, 99(26) :16916–16921.
- Chen, C. Z., Li, L., Lodish, H. F., et Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science*, 303(5654) :83–86.
- Chicas, A., Cogoni, C., et Macino, G. (2004). RNAi-dependent and RNAi-independent mechanisms contribute to the silencing of RIPed sequences in *Neurospora crassa*. *Nucleic Acids Res*, 32(14) :4237–4243.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E., et Fischer, R. L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell*, 110(1) :33–42.
- Clemson, C. M., McNeil, J. A., Willard, H. F., et Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase : evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol*, 132(3) :259–275.
- Clerc, P. et Avner, P. (1998). Role of the region 3' to *Xist* exon 6 in the counting process of X-chromosome inactivation. *Nat Genet*, 19(3) :249–253.
- Clouet d'Orval, B., Bortolin, M.-L., Gaspin, C., et Bachellerie, J.-P. (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp}. *Nucleic Acids Res*, 29(22) :4518–4529.
- Cockett, N. E., Jackson, S. P., Shay, T. L., Farnir, F., Berghmans, S., Snowden, G. D., Nielsen, D. M., et Georges, M. (1996). Polar overdominance at the ovine callipyge locus. *Science*, 273(5272) :236–238.
- Cockett, N. E., Jackson, S. P., Shay, T. L., Nielsen, D., Moore, S. S., Steele, M. R., Barendse, W., Green, R. D., et Georges, M. (1994). Chromosomal localization of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc Natl Acad Sci USA*, 91(8) :3019–3023.
- Cockett, N. E., Shay, T. L., et Smit, M. (2001). Analysis of the sheep genome. *Physiol Genomics*, 7(2) :69–78.
- Cogoni, C. et Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. *Curr Opin Genet Dev*, 10(6) :638–643.
- Constância, M., Dean, W., Lopes, S., Moore, T., Kelsey, G., et Reik, W. (2000). Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat Genet*, 26(2) :203–206.
- Cook, H. A., Koppetsch, B. S., Wu, J., et Theurkauf, W. E. (2004). The *Drosophila* SDE3 homolog *armitage* is required for *oskar* mRNA silencing and embryonic axis specification. *Cell*, 116(6) :817–829.
- Cooper, D. W., VandeBerg, J. L., Sharman, G. B., et Poole, W. E. (1971). Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. *Nat New Biol*, 230(13) :155–157.

- Cooper, P. R., Smilnich, N. J., Day, C. D., Nowak, N. J., Reid, L. H., Pearsall, R. S., Reece, M., Prawitt, D., Landers, J., Housman, D. E., Winterpacht, A., Zabel, B. U., Pelletier, J., Weissman, B. E., Shows, T. B., et Higgins, M. J. (1998). Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics*, 49(1) :38–51.
- Corbin, V. et Maniatis, T. (1989). Role of transcriptional interference in the *Drosophila melanogaster Adh* promoter switch. *Nature*, 337(6204) :279–282.
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 16(22) :10881–10890.
- Costanzi, C., Stein, P., Worrada, D. M., Schultz, R. M., et Pehrson, J. R. (2000). Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development*, 127(11) :2283–2289.
- Courtier, B., Heard, E., et Avner, P. (1995). *Xce* haplotypes show modified methylation in a region of the active X chromosome lying 3' to *Xist*. *Proc Natl Acad Sci USA*, 92(8) :3531–3535.
- Croteau, S., Charron, M.-C., Latham, K. E., et Naumova, A. K. (2003). Alternative splicing and imprinting control of the *Meg3/Gtl2-Dlk1* locus in mouse embryos. *Mamm Genome*, 14(4) :231–241.
- Côté, C. A., Greer, C. L., et Peculis, B. A. (2002). Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA*, 8(6) :786–797.
- Cullen, B. R., Lomedico, P. T., et Ju, G. (1984). Transcriptional interference in avian retroviruses — implications for the promoter insertion model of leukaemogenesis. *Nature*, 307(5948) :241–245.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., et Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, 111(2) :185–196.
- Dao, D., Frank, D., Qian, N., O'Keefe, D., Vosatka, R. J., Walsh, C. P., et Tycko, B. (1998). IMPT1, an imprinted gene similar to polyspecific transporter and multi-drug resistance genes. *Hum Mol Genet*, 7(4) :597–608.
- Darzacq, X., Jádý, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., et Kiss, T. (2002). Cajal body-specific small nuclear RNAs : a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J*, 21(11) :2746–2756.
- Davis, D. R. (1998). *Modification and Editing of RNA*, chapitre Biophysical and conformational properties of modified nucleosides in RNA (nuclear magnetic resonance studies), pages 85 – 102. ASM Press.
- Davis, E., Jensen, C. H., Schroder, H. D., Farnir, F., Shay-Hadfield, T., Kliem, A., Cockett, N., Georges, M., et Charlier, C. (2004). Ectopic expression of DLK1 protein in skeletal muscle of padumnal heterozygotes causes the callipyge phenotype. *Current Biology*, Sous presse.
- Davis, T. L., Yang, G. J., McCarrey, J. R., et Bartolomei, M. S. (2000). The *H19* methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet*, 9(19) :2885–2894.
- de la Casa-Esperón, E. et Sapienza, C. (2003). Natural selection and the evolution of genome imprinting. *Annu Rev Genet*, 37 :349–370.
- de la Puente, A., Hall, J., Wu, Y. Z., Leone, G., Peters, J., Yoon, B. J., Soloway, P., et Plass, C. (2002). Structural characterization of *Rasgrf1* and a novel linked imprinted locus. *Gene*, 291(1-2) :287–297.

- de los Santos, T., Schweizer, J., Rees, C. A., et Francke, U. (2000). Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from *PWCR1*, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *Am J Hum Genet*, 67(5) :1067–1082.
- Debrand, E., Chureau, C., Arnaud, D., Avner, P., et Heard, E. (1999). Functional analysis of the *DXPas34* locus, a 3' regulator of *Xist* expression. *Mol Cell Biol*, 19(12) :8513–8525.
- DeChiara, T. M., Robertson, E. J., et Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, 64(4) :849–859.
- Dehio, C. et Schell, J. (1994). Identification of plant genetic loci involved in a posttranscriptional mechanism for meiotically reversible transgene silencing. *Proc Natl Acad Sci USA*, 91(12) :5538–5542.
- Delaval, K. et Feil, R. (2004). Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev*, 14(2) :188–195.
- Demakova, O. V., Kotlikova, I. V., Gordadze, P. R., Alekseyenko, A. A., Kuroda, M. I., et Zhimulev, I. F. (2003). The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster*. *Chromosoma*, 112(3) :103–115.
- Di Croce, L., Raker, V. A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Lo Coco, F., Kouzarides, T., Nervi, C., Minucci, S., et Pelicci, P. G. (2002). Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science*, 295(5557) :1079–1082.
- Dittrich, B., Buiting, K., Korn, B., Rickard, S., Buxton, J., Saitoh, S., Nicholls, R. D., Poustka, A., Winterpacht, A., Zabel, B., et Horsthemke, B. (1996). Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. *Nat Genet*, 14(2) :163–170.
- Djikeng, A., Shi, H., Tschudi, C., Shen, S., et Ullu, E. (2003). An siRNA ribonucleoprotein is found associated with polyribosomes in *Trypanosoma brucei*. *RNA*, 9(7) :802–808.
- Doench, J. G., Petersen, C. P., et Sharp, P. A. (2003). siRNAs can function as miRNAs. *Genes Dev*, 17(4) :438–442.
- Doench, J. G. et Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. *Genes Dev*, 18(5) :504–511.
- Dorsett, Y. et Tuschl, T. (2004). siRNAs : applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov*, 3(4) :318–329.
- Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., et Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA*, 9(2) :180–186.
- Drewell, R. A., Brenton, J. D., Ainscough, J. F., Barton, S. C., Hilton, K. J., Arney, K. L., Dandolo, L., et Surani, M. A. (2000). Deletion of a silencer element disrupts *H19* imprinting independently of a DNA methylation epigenetic switch. *Development*, 127(16) :3419–3428.
- Duharcourt, S., Butler, A., et Meyer, E. (1995). Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev*, 9(16) :2065–2077.
- Duthie, S. M., Nesterova, T. B., Formstone, E. J., Keohane, A. M., Turner, B. M., Zakian, S. M., et Brockdorff, N. (1999). *Xist* RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum Mol Genet*, 8(2) :195–204.
- Eicher, E. M. (1970). X-autosome translocations in the mouse : total inactivation versus partial inactivation of the X chromosome. *Adv Genet*, 15 :175–259.
- Elbashir, S. M., Lendeckel, W., et Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev*, 15(2) :188–200.

- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., et Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J*, 20(23) :6877–6888.
- Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F., et Bowman, J. L. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr Biol*, 13(20) :1768–1774.
- Engemann, S., Strödicke, M., Paulsen, M., Franck, O., Reinhardt, R., Lane, N., Reik, W., et Walter, J. (2000). Sequence and functional comparison in the Beckwith-Wiedemann region : implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet*, 9(18) :2691–2706.
- Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., et Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol*, 5(1) :R1.
- Fahrner, J. A., Eguchi, S., Herman, J. G., et Baylin, S. B. (2002). Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Res*, 62(24) :7213–7218.
- Fatica, A., Morlando, M., et Bozzoni, I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J*, 19(22) :6218–6229.
- Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M., et Bartolomei, M. S. (2004). Transgenic RNAi reveals essential function for CTCF in *H19* gene imprinting. *Science*, 303(5655) :238–240.
- Feil, R., Boyano, M. D., Allen, N. D., et Kelsey, G. (1997). Parental chromosome-specific chromatin conformation in the imprinted *U2af1-rs1* gene in the mouse. *J Biol Chem*, 272(33) :20893–20900.
- Feil, R., Khosla, S., Cappai, P., et Loi, P. (1998). Genomic imprinting in ruminants : allele-specific gene expression in parthenogenetic sheep. *Mamm Genome*, 9(10) :831–834.
- Feil, R., Walter, J., Allen, N. D., et Reik, W. (1994). Developmental control of allelic methylation in the imprinted mouse *Igf2* and *H19* genes. *Development*, 120(10) :2933–2943.
- Ferguson-Smith, A. C., Lin, S.-P., et Youngson, N. (2004). Regulation of gene activity and repression : a consideration of unifying themes. *Curr Top Dev Biol*, 60 :197–213.
- Ferguson-Smith, A. C., Sasaki, H., Cattanaach, B. M., et Surani, M. A. (1993). Parental-origin-specific epigenetic modification of the mouse *H19* gene. *Nature*, 362(6422) :751–755.
- Finnegan, E. J., Margis, R., et Waterhouse, P. M. (2003). Posttranscriptional gene silencing is not compromised in the *Arabidopsis* *CARPEL FACTORY (DICER-LIKE1)* mutant, a homolog of Dicer-1 from *Drosophila*. *Curr Biol*, 13(3) :236–240.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., et Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669) :806–811.
- Fitzpatrick, G. V., Soloway, P. D., et Higgins, M. J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet*, 32(3) :426–431.
- Flavell, R. B. (1994). Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci USA*, 91(9) :3490–3496.
- Floyd, S. K. et Bowman, J. L. (2004). Gene regulation : ancient microRNA target sequences in plants. *Nature*, 428(6982) :485–486.
- Forney, J. D., Yantiri, F., et Mikami, K. (1996). Developmentally controlled rearrangement of surface protein genes in *Paramecium tetraurelia*. *J Euk Microbiol*, 43 :462–467.

- Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A. M., Esteller, M., et Feil, R. (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J*, 21(23) :6560–6570.
- Franke, A. et Baker, B. S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol Cell*, 4(1) :117–122.
- Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L., et Smith, T. P. (2002). Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res*, 12(10) :1496–1506.
- Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., et Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol*, 6(8) :784–791.
- Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., et Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet*, 24(1) :88–91.
- Fuks, F., Hurd, P. J., Deplus, R., et Kouzarides, T. (2003a). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res*, 31(9) :2305–2312.
- Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., et Kouzarides, T. (2003b). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem*, 278(6) :4035–4040.
- Gallagher, R. C., Pils, B., Albalwi, M., et Francke, U. (2002). Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader-Willi syndrome. *Am J Hum Genet*, 71(3) :669–678.
- Ganot, P., Jádý, B. E., Bortolin, M.-L., Darzacq, X., et Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol*, 19(10) :6906–6917.
- Garnier, O., Serrano, V., Duharcourt, S., et Meyer, E. (2004). RNA-Mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol Cell Biol*, 24(17) :7370–7379.
- Gaspin, C., Cavaillé, J., Erauso, G., et Bachellerie, J.-P. (2000). Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs : lessons from the *Pyrococcus* genomes. *J Mol Biol*, 297(4) :895–906.
- Georges, M., Charlier, C., et Cockett, N. (2003). The callipyge locus : evidence for the *trans* interaction of reciprocally imprinted genes. *Trends Genet*, 19(5) :248–252.
- Georgiades, P., Watkins, M., Surani, M. A., et Ferguson-Smith, A. C. (2000). Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development*, 127(21) :4719–4728.
- Giddings, S. J., King, C. D., Harman, K. W., Flood, J. F., et Carnaghi, L. R. (1994). Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. *Nat Genet*, 6(3) :310–313.
- Gilbert, W. (1992). *The Code of Codes*, chapitre Visions of the Grail, pages 83–97. Harvard University Press.
- Gogolevskaya, I. K., Makarova, J. A., Gause, L. N., Kulichkova, V. A., Konstantinova, I. M., et Kramerov, D. A. (2002). U87 RNA, a novel C/D box small nucleolar RNA from mammalian cells. *Gene*, 292(1-2) :199–204.

- Goodwin, J., Chapman, K., Swaney, S., Parks, T. D., Wernsman, E. A., et Dougherty, W. G. (1996). Genetic and biochemical dissection of transgenic RNA-mediated virus resistance. *Plant Cell*, 8(1) :95–105.
- Gould, T. D. et Pfeifer, K. (1998). Imprinting of mouse *Kvlqt1* is developmentally regulated. *Hum Mol Genet*, 7(3) :483–487.
- Grad, Y., Aach, J., Hayes, G. D., Reinhart, B. J., Church, G. M., Ruvkun, G., et Kim, J. (2003). Computational and experimental identification of *C. elegans* microRNAs. *Mol Cell*, 11(5) :1253–1263.
- Grandjean, V., O'Neill, L., Sado, T., Turner, B., et Ferguson-Smith, A. (2001). Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted *Igf2-H19* domain. *FEBS Lett*, 488(3) :165–169.
- Gray, T. A., Saitoh, S., et Nicholls, R. D. (1999). An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc Natl Acad Sci USA*, 96(10) :5616–5621.
- Greger, I. H., Aranda, A., et Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, 97(15) :8415–8420.
- Gregory, R. I., Randall, T. E., Johnson, C. A., Khosla, S., Hatada, I., O'Neill, L. P., Turner, B. M., et Feil, R. (2001). DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes *Snrpn* and *U2af1-rs1*. *Mol Cell Biol*, 21(16) :5426–5436.
- Gribnau, J., Hochedlinger, K., Hata, K., Li, E., et Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev*, 17(6) :759–773.
- Griffiths-Jones, S. (2004). The microRNA registry. *Nucleic Acids Res*, 32(supplément 1) :D109–D111.
- Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., et Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell*, 106(1) :23–34.
- Gu, W., Wei, X., Pannuti, A., et Lucchesi, J. C. (2000). Targeting the chromatin-remodeling MSL complex of *Drosophila* to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. *EMBO J*, 19(19) :5202–5211.
- Guillemot, F., Caspary, T., Tilghman, S. M., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Anderson, D. J., Joyner, A. L., Rossant, J., et Nagy, A. (1995). Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nat Genet*, 9(3) :235–242.
- Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., et Joyner, A. L. (1994). Essential role of *Mash-2* in extraembryonic development. *Nature*, 371(6495) :333–336.
- Guillet-Deniau, I., Burnol, A.-F., et Girard, J. (1997). Identification and localization of a skeletal muscle serotonin 5-HT_{2A} receptor coupled to the Jak/STAT pathway. *J Biol Chem*, 272(23) :14825–14829.
- Guo, H. S. et Ding, S. W. (2002). A viral protein inhibits the long range signaling activity of the gene silencing signal. *EMBO J*, 21(3) :398–407.
- Gwizdek, C., Ossareh-Nazari, B., Brownawell, A. M., Doglio, A., Bertrand, E., Macara, I. G., et Dargemont, C. (2003). Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J Biol Chem*, 278(8) :5505–5508.
- Ha, I., Wightman, B., et Ruvkun, G. (1996). A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. *Genes Dev*, 10(23) :3041–3050.

- Hagemann, L. J., Peterson, A. J., Weilert, L. L., Lee, R. S., et Tervit, H. R. (1998). In vitro and early in vivo development of sheep gynogenones and putative androgenones. *Mol Reprod Dev*, 50(2) :154–162.
- Hagiwara, Y., Hirai, M., Nishiyama, K., Kanazawa, I., Ueda, T., Sakaki, Y., et Ito, T. (1997). Screening for imprinted genes by allelic message display : identification of a paternally expressed gene impact on mouse chromosome 18. *Proc Natl Acad Sci USA*, 94(17) :9249–9254.
- Haig, D. (1994). Refusing the ovarian time bomb : three viewpoints and a reply. *Trends Genet*, 10(10) :346–347.
- Haig, D. et Westoby, M. (1991). Genomic imprinting in endosperm : its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Phil Trans R Soc Lond*, 333 :1–13.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., et Surani, M. A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev*, 117(1-2) :15–23.
- Haley, B. et Zamore, P. D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol*, 11(7) :599–606.
- Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A., et Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. *Science*, 297(5590) :2232–2237.
- Hamilton, A. J. et Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441) :950–952.
- Hammond, S. M., Bernstein, E., Beach, D., et Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, 404(6775) :293–296.
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., et Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, 293(5532) :1146–1150.
- Hannon, G. J. (2002). RNA interference. *Nature*, 418(6894) :244–251.
- Hardy, R. W., Lindsley, D. L., Livak, K. J., Lewis, B., Siversten, A. L., Joslyn, G. L., Edwards, J., et Bonaccorsi, S. (1984). Cytogenetic analysis of a segment of the *Y* chromosome of *Drosophila melanogaster*. *Genetics*, 107(4) :591–610.
- Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., et Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature*, 405(6785) :486–489.
- Hata, K., Okano, M., Lei, H., et Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development*, 129(8) :1983–1993.
- Hatada, I. et Mukai, T. (1995). Genomic imprinting of p57^{KIP2}, a cyclin-dependent kinase inhibitor, in mouse. *Nat Genet*, 11(2) :204–206.
- Hatada, I., Sugama, T., et Mukai, T. (1993). A new imprinted gene cloned by a methylation-sensitive genome scanning method. *Nucleic Acids Res*, 21(24) :5577–5582.
- Hausler, B. et Somerville, R. L. (1979). Interaction in vivo between strong closely spaced constitutive promoters. *J Mol Biol*, 127(3) :353–356.
- Hawkes, C. et Kar, S. (2004). The insulin-like growth factor-II/mannose-6-phosphate receptor : structure, distribution and function in the central nervous system. *Brain Res Brain Res Rev*, 44(2-3) :117–140.

- Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N., Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M., Komatsubara, H., Shiroishi, T., Moriwaki, K., Katsuki, M., Hatano, N., Sasaki, H., Ueda, T., Mise, N., Takagi, N., Plaas, C., et Chapman, V. M. (1994). Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nat Genet*, 6(1) :33–40.
- He, Z. et Sontheimer, E. J. (2004). "siRNAs and miRNAs" : a meeting report on RNA silencing. *RNA*, 10(8) :1165–1173.
- Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D., et Spector, D. L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*, 107(6) :727–738.
- Henras, A. (2002). *Contribution à l'étude des modes de synthèse et d'action des particules ribonucléoprotéiques de type C/D et H/ACA*. Thèse, Université Toulouse III Paul Sabatier.
- Herman, H., Lu, M., Anggraini, M., Sikora, A., Chang, Y., Yoon, B. J., et Soloway, P. D. (2003). *Trans* allele methylation and paramutation-like effects in mice. *Nat Genet*, 34(2) :199–202.
- Hernandez, A., Lyon, G. J., Schneider, M. J., et St Germain, D. L. (1999). Isolation and characterization of the mouse gene for the type 3 iodothyronine deiodinase. *Endocrinology*, 140(1) :124–130.
- Hershko, A., Razin, A., et Shemer, R. (1999). Imprinted methylation and its effect on expression of the mouse *Zfp127* gene. *Gene*, 234(2) :323–327.
- Hikichi, T., Kohda, T., Kaneko-Ishino, T., et Ishino, F. (2003). Imprinting regulation of the murine *Meg1/Grb10* and human GRB10 genes ; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res*, 31(5) :1398–1406.
- Hipfner, D. R., Weigmann, K., et Cohen, S. M. (2002). The *bantam* gene regulates *Drosophila* growth. *Genetics*, 161(4) :1527–1537.
- Hobert, O. (2004). Common logic of transcription factor and microRNA action. *Trends Biochem Sci*, 29(9) :462–468.
- Holen, T., Amarzguoui, M., Babaie, E., et Prydz, H. (2003). Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway. *Nucleic Acids Res*, 31(9) :2401–2407.
- Holzmann, K., Gerner, C., Korosec, T., Poltl, A., Grimm, R., et Sauermann, G. (1998). Identification and characterization of the ubiquitously occurring nuclear matrix protein NMP 238. *Biochem Biophys Res Commun*, 252(1) :39–45.
- Horike, S., Mitsuya, K., Meguro, M., Kotobuki, N., Kashiwagi, A., Notsu, T., Schulz, T. C., Shirayoshi, Y., et Oshimura, M. (2000). Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. *Hum Mol Genet*, 9(14) :2075–2083.
- Horowitz, H. et Platt, T. (1982). Regulation of transcription from tandem and convergent promoters. *Nucleic Acids Res*, 10(18) :5447–5465.
- Hoshiya, H., Meguro, M., Kashiwagi, A., Okita, C., et Oshimura, M. (2003). Calcr, a brain-specific imprinted mouse calcitonin receptor gene in the imprinted cluster of the proximal region of chromosome 6. *J Hum Genet*, 48(4) :208–211.
- Houbaviy, H. B., Murray, M. F., et Sharp, P. A. (2003). Embryonic stem cell-specific microRNAs. *Dev Cell*, 5(2) :351–358.
- Hüttenhofer, A., Kieffmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.-P., et Brosius, J. (2001). RNomics : an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J*, 20(11) :2943–2953.

- Huang, G. M., Jarmolowski, A., Struck, J. C. R., et Fournier, M. J. (1992). Accumulation of U14 small nuclear RNA in *Saccharomyces cerevisiae* requires box C, box D, and a 5', 3' terminal stem. *Mol Cell Biol*, 12(10) :4456–63.
- Hughes, J. M. et Ares, M. J. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J*, 10(13) :4231–4239.
- Huibregtse, J. M., Scheffner, M., et Howley, P. M. (1993). Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53. *Mol Cell Biol*, 13(2) :775–784.
- Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., et Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, 293(5531) :834–838.
- Hutvagner, G. et Zamore, P. D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 297(5589) :2056–2060.
- Ingelbrecht, I., Van Houdt, H., Van Montagu, M., et Depicker, A. (1994). Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc Natl Acad Sci USA*, 91(22) :10502–10506.
- Ischia, R., Lovisetti-Scamihorn, P., Hogue-Angeletti, R., Wolkersdorfer, M., Winkler, H., et Fischer-Colbrie, R. (1997). Molecular cloning and characterization of NESP55, a novel chromogranin-like precursor of a peptide with 5-HT1B receptor antagonist activity. *J Biol Chem*, 272(17) :11657–11662.
- Ishizuka, A., Siomi, M. C., et Siomi, H. (2002). A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev*, 16(19) :2497–2508.
- Jacobsen, S. E., Running, M. P., et Meyerowitz, E. M. (1999). Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development*, 126(23) :5231–5243.
- Jády, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E., et Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J*, 22(8) :1878–1888.
- Jády, B. E. et Kiss, T. (2000). Characterisation of the U83 and U84 small nucleolar RNAs : two novel 2'-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. *Nucleic Acids Res*, 28(6) :1348–54.
- Jády, B. E. et Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J*, 20(3) :541–551.
- Jenuwein, T. et Allis, C. D. (2001). Translating the histone code. *Science*, 293(5532) :1074–1080.
- Jia, S., Noma, K., et Grewal, S. I. (2004). RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science*, 304(5679) :1971–1976.
- Jiang, S., Hemann, M. A., Lee, M. P., et Feinberg, A. P. (1998). Strain-dependent developmental relaxation of imprinting of an endogenous mouse gene, *Kvlqt1*. *Genomics*, 53(3) :395–399.
- Jin, P., Zarnescu, D. C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T. A., Nelson, D. L., Moses, K., et Warren, S. T. (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci*, 7(2) :113–117.
- Jinno, Y., Yun, K., Nishiwaki, K., Kubota, T., Ogawa, O., Reeve, A. E., et Niikawa, N. (1994). Mosaic and polymorphic imprinting of the WT1 gene in humans. *Nat Genet*, 6(3) :305–309.
- Johnson, D. R. (1974). Hairpin-tail : a case of post-reductional gene action in the mouse egg. *Genetics*, 76(4) :795–805.

- Johnston, R. J. et Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*, 426(6968) :845–849.
- Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B. M., Khochbin, S., et Vourc'h, C. (2004). Stress-induced transcription of satellite III repeats. *J Cell Biol*, 164(1) :25–33.
- Jones, B. K., Levorse, J. M., et Tilghman, S. M. (1998). *Igf2* imprinting does not require its own DNA methylation or *H19* RNA. *Genes Dev*, 12(14) :2200–2207.
- Jones, J. I. et Clemmons, D. R. (1995). Insulin-like growth factors and their binding proteins : biological actions. *Endocr Rev*, 16(1) :3–34.
- Jones, L., Hamilton, A. J., Voinnet, O., Thomas, C. L., Maule, A. J., et Baulcombe, D. C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell*, 11(12) :2291–2301.
- Jones-Rhoades, M. W. et Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell*, 14(6) :787–799.
- Jong, M. T., Carey, A. H., Caldwell, K. A., Lau, M. H., Handel, M. A., Driscoll, D. J., Stewart, C. L., Rinchik, E. M., et Nicholls, R. D. (1999). Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the Prader-Willi syndrome genetic region. *Hum Mol Genet*, 8(5) :795–803.
- Jonkers, J., van Amerongen, R., van der Valk, M., Robanus-Maandag, E., Molenaar, M., Destree, O., et Berns, A. (1999). In vivo analysis of *Frat1* deficiency suggests compensatory activity of *Frat3*. *Mech Dev*, 88(2) :183–194.
- Jost, J.-P. et Bruhat, A. (1997). The formation of DNA methylation patterns and the silencing of genes. *Prog Nucleic Acid Res Mol Biol*, 57 :217–248.
- Jouvenot, Y., Poirier, F., Jami, J., et Paldi, A. (1999). Biallelic transcription of *Igf2* and *H19* in individual cells suggests a post-transcriptional contribution to genomic imprinting. *Curr Biol*, 9(20) :1199–1202.
- Kaffer, C. R., Srivastava, M., Park, K. Y., Ives, E., Hsieh, S., Battle, J., Grinberg, A., Huang, S. P., et Pfeifer, K. (2000). A transcriptional insulator at the imprinted *H19/Igf2* locus. *Genes Dev*, 14(15) :1908–1919.
- Kagitani, F., Kuroiwa, Y., Wakana, S., Shiroishi, T., Miyoshi, N., Kobayashi, S., Nishida, M., Kohda, T., Kaneko-Ishino, T., et Ishino, F. (1997). *Peg5/Neuronatin* is an imprinted gene located on sub-distal chromosome 2 in the mouse. *Nucleic Acids Res*, 25(17) :3428–3432.
- Kagotani, K., Takebayashi, S., Kohda, A., Taguchi, H., Paulsen, M., Walter, J., Reik, W., et Okumura, K. (2002). Replication timing properties within the mouse distal chromosome 7 imprinting cluster. *Biosci Biotechnol Biochem*, 66(5) :1046–1051.
- Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H., et Ropers, H. H. (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat Genet*, 5(1) :74–78.
- Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., Ohlsson, R., et Lobanenko, V. V. (2000). Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive. *Curr Biol*, 10(14) :853–856.
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., et Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, 429(6994) :900–903.
- Kaneko-Ishino, T., Kuroiwa, Y., Miyoshi, N., Kohda, T., Suzuki, R., Yokoyama, M., Viville, S., Barton, S. C., Ishino, F., et Surani, M. A. (1995). *Peg1/Mest* imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat Genet*, 11(1) :52–59.

- Kashiwagi, A., Meguro, M., Hoshiya, H., Haruta, M., Ishino, F., Shibahara, T., et Oshimura, M. (2003). Predominant maternal expression of the mouse *Atp10c* in hippocampus and olfactory bulb. *J Hum Genet*, 48(4) :194–198.
- Kass, S., Tyc, K., Steitz, J. A., et Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell*, 60(6) :897–908.
- Kasschau, K. D. et Carrington, J. C. (1998). A counterdefensive strategy of plant viruses : suppression of posttranscriptional gene silencing. *Cell*, 95(4) :461–470.
- Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., et Carrington, J. C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev Cell*, 4(2) :205–217.
- Kato, M. V., Ikawa, Y., Hayashizaki, Y., et Shibata, H. (1998). Paternal imprinting of mouse serotonin receptor 2A gene *Htr2* in embryonic eye : a conserved imprinting regulation on the *RB/Rb* locus. *Genomics*, 47(1) :146–148.
- Kato, Y., Rideout, W. M. r., Hilton, K., Barton, S. C., Tsunoda, Y., et Surani, M. A. (1999). Developmental potential of mouse primordial germ cells. *Development*, 126(9) :1823–1832.
- Kaufman, M. H., Barton, S. C., et Surani, M. A. (1977). Normal postimplantation development of mouse parthenogenetic embryos to the forelimb bud stage. *Nature*, 265(5589) :53–55.
- Kawasaki, H. et Taira, K. (2003a). Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. *Nature*, 423(6942) :838–842.
- Kawasaki, H. et Taira, K. (2003b). Retraction : Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. *Nature*, 426(6962) :100.
- Kawasaki, H. et Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*, 431(7005) :211–217.
- Kayashima, T., Ohta, T., Niikawa, N., et Kishino, T. (2003). On the conflicting reports of imprinting status of mouse *ATP10a* in the adult brain : strain-background-dependent imprinting ? *J Hum Genet*, 48(9) :492–493.
- Kelley, R. L. (2004). Path to equality strewn with *roX*. *Dev Biol*, 269(1) :18–25.
- Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L., et Kuroda, M. I. (1999). Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. *Cell*, 98(4) :513–522.
- Kelley, R. L., Solovyeva, I., Lyman, L. M., Richman, R., Solovyev, V., et Kuroda, M. I. (1995). Expression of *msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell*, 81(6) :867–877.
- Kennedy, S., Wang, D., et Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature*, 427(6975) :645–649.
- Kennerdell, J. R., Yamaguchi, S., et Carthew, R. W. (2002). RNAi is activated during *Drosophila* oocyte maturation in a manner dependent on *aubergine* and *spindle-E*. *Genes Dev*, 16(15) :1884–1889.
- Keohane, A. M., O’neill, L. P., Belyaev, N. D., Lavender, J. S., et Turner, B. M. (1996). X-Inactivation and histone H4 acetylation in embryonic stem cells. *Dev Biol*, 180(2) :618–630.
- Kermicle, J. (1970). Dependence of the R-mottled aleurone phenotype in maize on the mode of sexual transmission. *Genetics*, 66 :69–85.
- Ketting, R. F., Fischer, S. E. J., Bernstein, E., Sijen, T., Hannon, G. J., et Plasterk, R. H. A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev*, 15(20) :2654–2659.

- Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., et Plasterk, R. H. (1999). *Mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*, 99(2) :133–141.
- Khvorova, A., Reynolds, A., et Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2) :209–216.
- Kikyo, N., Williamson, C. M., John, R. M., Barton, S. C., Beechey, C. V., Ball, S. T., Cattanach, B. M., Surani, M. A., et Peters, J. (1997). Genetic and functional analysis of neuronatin in mice with maternal or paternal duplication of distal Chr 2. *Dev Biol*, 190(1) :66–77.
- Killian, J. K., Byrd, J. C., Jirtle, J. V., Munday, B. L., Stoskopf, M. K., MacDonald, R. G., et Jirtle, R. L. (2000). M6P/IGF2R imprinting evolution in mammals. *Mol Cell*, 5(4) :707–716.
- Killian, J. K., Nolan, C. M., Stewart, N., Munday, B. L., Andersen, N. A., Nicol, S., et Jirtle, R. L. (2001). Monotreme IGF2 expression and ancestral origin of genomic imprinting. *J Exp Zool*, 291(2) :205–212.
- Kim, J., Bergmann, A., Lucas, S., Stone, R., et Stubbs, L. (2004). Lineage-specific imprinting and evolution of the zinc-finger gene *ZIM2*. *Genomics*, 84(1) :47–58.
- Kim, J., Bergmann, A., Wehri, E., Lu, X., et Stubbs, L. (2001). Imprinting and evolution of two Kruppel-type zinc-finger genes, *ZIM3* and *ZNF264*, located in the *PEG3/USP29* imprinted domain. *Genomics*, 77(1-2) :91–98.
- Kim, J., Kollhoff, A., Bergmann, A., et Stubbs, L. (2003). Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, *Peg3*. *Hum Mol Genet*, 12(3) :233–245.
- Kim, J., Lu, X., et Stubbs, L. (1999). *Zim1*, a maternally expressed mouse Kruppel-type zinc-finger gene located in proximal chromosome 7. *Hum Mol Genet*, 8(5) :847–854.
- Kim, J., Noskov, V. N., Lu, X., Bergmann, A., Ren, X., Warth, T., Richardson, P., Kouprina, N., et Stubbs, L. (2000). Discovery of a novel, paternally expressed ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4. *Genome Res*, 10(8) :1138–1147.
- Kimura, Y., Tateno, H., Handel, M. A., et Yanagimachi, R. (1998). Factors affecting meiotic and developmental competence of primary spermatocyte nuclei injected into mouse oocytes. *Biol Reprod*, 59(4) :871–877.
- Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L., et Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science*, 303(5657) :521–523.
- Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B., et Fischer, R. L. (1999). Imprinting of the *MEDEA* polycomb gene in the arabidopsis endosperm. *Plant Cell*, 11(10) :1945–1952.
- Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., et Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes Dev*, 18(10) :1165–1178.
- Kiss, A. M., Jady, B. E., Darzacq, X., Verheggen, C., Bertrand, E., et Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res*, 30(21) :4643–4649.
- Kiss, T. (2002). Small nucleolar RNAs : an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, 109(2) :145–148.
- Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., et Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA : a novel function for small nucleolar RNAs. *Cell*, 85(7) :1077–1088.

- Kiss-László, Z., Henry, Y., et Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J*, 17(3) :797–807.
- Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., et Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. *Nature*, 364(6436) :459–463.
- Klein, D. J., Schmeing, T. M., Moore, P. B., et Steitz, T. A. (2001). The kink-turn : a new RNA secondary structure motif. *EMBO J*, 20(15) :4214–4221.
- Kmita, M. et Duboule, D. (2003). Organizing axes in time and space ; 25 years of colinear tinkering. *Science*, 301(5631) :331–333.
- Knight, S. W. et Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, 293(5538) :2269–2271.
- Knight, S. W. et Bass, B. L. (2002). The role of RNA editing by ADARs in RNAi. *Mol Cell*, 10(4) :809–817.
- Koepsell, H. (1998). Organic cation transporters in intestine, kidney, liver, and brain. *Annu Rev Physiol*, 60 :243–266.
- Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., Park, E. S., Seo, J. S., et Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. *Nature*, 428(6985) :860–864.
- Kono, T., Obata, Y., Yoshimizu, T., Nakahara, T., et Carroll, J. (1996). Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nat Genet*, 13(1) :91–94.
- Koohmaraie, M., Shackelford, S. D., Wheeler, T. L., Lonergan, S. M., et Doumit, M. E. (1995). A muscle hypertrophy condition in lamb (callipyge) : characterization of effects on muscle growth and meat quality traits. *J Anim Sci*, 73(12) :3596–3607.
- Koonin, E. V., Zhou, S., et Lucchesi, J. C. (1995). The chromo superfamily : new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic Acids Res*, 23(21) :4229–4233.
- Kratzer, P. G., Chapman, V. M., Lambert, H., Evans, R. E., et Liskay, R. M. (1983). Differences in the DNA of the inactive X chromosomes of fetal and extraembryonic tissues of mice. *Cell*, 33(1) :37–42.
- Kuroda, M. I., Kernan, M. J., Kreber, R., Ganetzky, B., et Baker, B. S. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell*, 66(5) :935–947.
- Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., et Gage, F. H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell*, 116(6) :779–793.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., et Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev*, 16(22) :2893–2905.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., et Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science*, 294(5543) :853–858.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., et Tuschl, T. (2003). New microRNAs from mouse and human. *RNA*, 9(2) :175–179.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., et Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol*, 12(9) :735–739.
- Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet*, 30(4) :363–364.

- Lai, E. C., Tomancak, P., Williams, R. W., et Rubin, G. M. (2003). Computational identification of *Drosophila* microRNA genes. *Genome Biol*, 4(7) :R42.
- Landers, M., Bancescu, D. L., Le Meur, E., Rougeulle, C., Glatt-Deeley, H., Brannan, C., Muscatelli, F., et Lalande, M. (2004). Regulation of the large (~ 1000 kb) imprinted murine *Ube3a* antisense transcript by alternative exons upstream of *Snurf/Snrpn*. *Nucleic Acids Res*, 32(11) :3480–3492.
- Lau, N. C., Lim, L. P., Weinstein, E. G., et Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294(5543) :858–862.
- Lee, C. G., Chang, K. A., Kuroda, M. I., et Hurwitz, J. (1997). The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J*, 16(10) :2671–2681.
- Lee, C. Y., Lee, A., et Chanfreau, G. (2003a). The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of *S. cerevisiae* box C/D snoRNAs. *RNA*, 9(11) :1362–1370.
- Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., et Ishino, F. (2002a). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development*, 129(8) :1807–1817.
- Lee, J. T. (2000). Disruption of imprinted X inactivation by parent-of-origin effects at *Tsix*. *Cell*, 103(1) :17–27.
- Lee, J. T., Davidow, L. S., et Warshawsky, D. (1999a). *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nat Genet*, 21(4) :400–404.
- Lee, J. T. et Lu, N. (1999). Targeted mutagenesis of *Tsix* leads to nonrandom X inactivation. *Cell*, 99(1) :47–57.
- Lee, J. T., Lu, N., et Han, Y. (1999b). Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. *Proc Natl Acad Sci USA*, 96(7) :3836–3841.
- Lee, M. H., Reynisdottir, I., et Massague, J. (1995). Cloning of p57^{KIP2}, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. *Genes Dev*, 9(6) :639–649.
- Lee, M. P., DeBaun, M. R., Mitsuya, K., Galonek, H. L., Brandenburg, S., Oshimura, M., et Feinberg, A. P. (1999c). Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc Natl Acad Sci USA*, 96(9) :5203–5208.
- Lee, R. C. et Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 294(5543) :862–864.
- Lee, R. C., Feinbaum, R. L., et Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5) :843–854.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et Kim, V. N. (2003b). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956) :415–419.
- Lee, Y., Jeon, K., Lee, J. T., Kim, S., et Kim, V. N. (2002b). MicroRNA maturation : stepwise processing and subcellular localization. *EMBO J*, 21(17) :4663–4670.
- Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., et Kim, V. N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, Sous presse.
- Lee, Y. J., Park, C. W., Hahn, Y., Park, J., Lee, J., Yun, J. H., Hyun, B., et Chung, J. H. (2000). *Mit1/Lb9* and *Copg2*, new members of mouse imprinted genes closely linked to *Peg1/Mest(1)*. *FEBS Lett*, 472(2-3) :230–234.

- Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., et Carthew, R. W. (2004b). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117(1) :69–81.
- Lefebvre, L., Viville, S., Barton, S. C., Ishino, F., Keverne, E. B., et Surani, M. A. (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. *Nat Genet*, 20(2) :163–169.
- Leff, S. E., Brannan, C. I., Reed, M. L., Ozcelik, T., Francke, U., Copeland, N. G., et Jenkins, N. A. (1992). Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nat Genet*, 2(4) :259–264.
- Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., et Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol*, 13(14) :1192–1200.
- Lercher, M. J. et Hurst, L. D. (2003). Imprinted chromosomal regions of the human genome have unusually high recombination rates. *Genetics*, 165(3) :1629–1632.
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., et Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7) :787–798.
- Li, E., Beard, C., et Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, 366(6453) :362–365.
- Li, H., Li, W. X., et Ding, S. W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science*, 296(5571) :1319–1321.
- Li, H. D., Zagorski, J., et Fournier, M. J. (1990). Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 10(3) :1145–1152.
- Li, L., Keverne, E. B., Aparicio, S. A., Ishino, F., Barton, S. C., et Surani, M. A. (1999). Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science*, 284(5412) :330–333.
- Liang, W. Q. et Fournier, M. J. (1995). U14 base-pairs with 18S rRNA : a novel snoRNA interaction required for rRNA processing. *Genes Dev*, 9(19) :2433–2443.
- Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., et Bartel, D. P. (2003a). Vertebrate microRNA genes. *Science*, 299(5612) :1540.
- Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., et Bartel, D. P. (2003b). The microRNAs of *Caenorhabditis elegans*. *Genes Dev*, 17(8) :991–1008.
- Lin, S.-P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaillé, J., et Ferguson-Smith, A. C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet*, 35(1) :97–102.
- Lindbo, J. A. et Dougherty, W. G. (1992a). Pathogen-derived resistance to a potyvirus : immune and resistant phenotypes in transgenic tobacco expressing altered forms of a potyvirus coat protein nucleotide sequence. *Mol Plant Microbe Interact*, 5(2) :144–153.
- Lindbo, J. A. et Dougherty, W. G. (1992b). Untranslatable transcripts of the tobacco etch virus coat protein gene sequence can interfere with tobacco etch virus replication in transgenic plants and protoplasts. *Virology*, 189(2) :725–733.
- Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M., et Dougherty, W. J. (1993). Induction of a highly specific antiviral state in transgenic plants : Implications of gene expression and virus resistance. *Plant Cell*, 5(12) :1749–1759.

- Lingel, A., Simon, B., Izaurralde, E., et Sattler, M. (2003). Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature*, 426(6965) :465–469.
- Lipardi, C., Wei, Q., et Paterson, B. M. (2001). RNAi as random degradative PCR : siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell*, 107(3) :297–307.
- Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D., Carrington, J. C., Doerge, R. W., Colot, V., et Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998) :471–476.
- Liu, F. et Roth, R. A. (1995). Grb-IR : a SH2-domain-containing protein that binds to the insulin receptor and inhibits its function. *Proc Natl Acad Sci USA*, 92(22) :10287–10291.
- Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., et Hannon, G. J. (2004a). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, 305(5689) :1437–1441.
- Liu, J. P., Baker, J., Perkins, A. S., Robertson, E. J., et Efstratiadis, A. (1993). Mice carrying null mutations of the genes encoding insulin-like growth factor I (*Igf-1*) and type 1 IGF receptor (*Igf1r*). *Cell*, 75(1) :59–72.
- Liu, L., Grainger, J., Canizares, M. C., Angell, S. M., et Lomonosoff, G. P. (2004b). Cowpea mosaic virus RNA-1 acts as an amplicon whose effects can be counteracted by a RNA-2-encoded suppressor of silencing. *Virology*, 323(1) :37–48.
- Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P., et Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science*, 301(5641) :1921–1925.
- Liu, Y., Mochizuki, K., et Gorovsky, M. A. (2004c). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc Natl Acad Sci USA*, 101(6) :1679–1684.
- Livak, K. J. (1984). Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics*, 107(4) :611–634.
- Llave, C., Kasschau, K. D., Rector, M. A., et Carrington, J. C. (2002a). Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, 14(7) :1605–1619.
- Llave, C., Xie, Z., Kasschau, K. D., et Carrington, J. C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*, 297(5589) :2053–2056.
- Lowe, T. M. et Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science*, 283(5405) :1168–71.
- Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., et Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, 303(5654) :95–98.
- Lund, G., Messing, J., et Viotti, A. (1995). Endosperm-specific demethylation and activation of specific alleles of α -tubulin genes of *Zea mays* L. *Mol Gen Genet*, 246(6) :716–722.
- Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J., et Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc Natl Acad Sci USA*, 97(19) :10637–10642.
- Lyman, L. M., Copps, K., Rastelli, L., Kelley, R. L., et Kuroda, M. I. (1997). *Drosophila* male-specific lethal-2 protein : structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics*, 147(4) :1743–1753.
- Lynch, C. et Tristem, M. (2003). A co-opted *gypsy*-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. *Curr Biol*, 13(17) :1518–1523.

- Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*, 190 :372–373.
- Lyon, M. F. et Glenister, P. H. (1977). Factors affecting the observed number of young resulting from adjacent-2 disjunction in mice carrying a translocation. *Genet Res*, 29(1) :83–92.
- MacDonald, H. R. et Wevrick, R. (1997). The necdin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. *Hum Mol Genet*, 6(11) :1873–1878.
- Maden, B. E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol*, 39 :241–303.
- Maden, B. E., Corbett, M. E., Heeney, P. A., Pugh, K., et Ajuh, P. M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie*, 77(1-2) :22–29.
- Maden, B. E. H. (1998). *Modification and Editing of RNA*, chapitre 24, pages 421 – 440. ASM Press. « Intracellular locations of RNA-modifying enzymes ».
- Maison, C., Bailly, D., Peters, A. H., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., et Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet*, 30(3) :329–334.
- Mak, W., Baxter, J., Silva, J., Newall, A. E., Otte, A. P., et Brockdorff, N. (2002). Mitotically stable association of polycomb group proteins Eed and Enx1 with the inactive X chromosome in trophoblast stem cells. *Curr Biol*, 12(12) :1016–1020.
- Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., et Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science*, 303(5658) :666–669.
- Mallory, A. C., Dugas, D. V., Bartel, D. P., et Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr Biol*, 14(12) :1035–1046.
- Mann, J. R. et Lovell-Badge, R. H. (1984). Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. *Nature*, 310(5972) :66–67.
- Mann, M. et Varmuza, S. (1994). Refusing the ovarian time bomb : three viewpoints and a reply. *Trends Genet*, 10(10) :348–349.
- Marahrens, Y., Loring, J., et Jaenisch, R. (1998). Role of the *Xist* gene in X chromosome choosing. *Cell*, 92(5) :657–664.
- Marahrens, Y., Panning, B., Dausman, J., Strauss, W., et Jaenisch, R. (1997). *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev*, 11(2) :156–166.
- Marker, C., Zemmann, A., Terhörst, T., Kiefmann, M., Kastenmayer, J. P., Green, P., Bachellerie, J.-P., Brosius, J., et Hüttenhofer, A. (2002). Experimental RNomics : identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol*, 12(23) :2002–2013.
- Martens, J. A., Laprade, L., et Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature*, 429(6991) :571–574.
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., et Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 110(5) :563–574.
- Massé, E., Escorcía, F. E., et Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev*, 17(19) :2374–2383.
- Massenet, S., Mougin, A., et Branlant, C. (1998). *Modification and editing of RNA*, chapitre Posttranscriptional modifications in the U small nuclear RNAs, pages 201–227. ASM Press.

- Matzke, M. A., Aufsatz, W., Kanno, T., Mette, M. F., et Matzke, A. J. (2002). Homology-dependent gene silencing and host defense in plants. *Adv Genet*, 46 :235–275.
- Matzke, M. A. et Matzke, A. J. (2004). Planting the seeds of a new paradigm. *PLoS Biol*, 2(5) :E133.
- Maxwell, E. S. et Fournier, M. J. (1995). The small nucleolar RNAs. *Annu Rev Biochem*, 35 :897–934.
- McGrath, J. et Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37(1) :179–183.
- McLaren, R. J. et Montgomery, G. W. (1999). Genomic imprinting of the insulin-like growth factor 2 gene in sheep. *Mamm Genome*, 10(6) :588–591.
- McNally, E. M., Ly, C. T., et Kunkel, L. M. (1998). Human ε -sarcoglycan is highly related to α -sarcoglycan (adhalin), the limb girdle muscular dystrophy 2D gene. *FEBS Lett*, 422(1) :27–32.
- Meguro, M., Kashiwagi, A., Mitsuya, K., Nakao, M., Kondo, I., Saitoh, S., et Oshimura, M. (2001a). A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome. *Nat Genet*, 28(1) :19–20.
- Meguro, M., Mitsuya, K., Nomura, N., Kohda, M., Kashiwagi, A., Nishigaki, R., Yoshioka, H., Nakao, M., Oishi, M., et Oshimura, M. (2001b). Large-scale evaluation of imprinting status in the Prader-Willi syndrome region : an imprinted direct repeat cluster resembling small nucleolar RNA genes. *Hum Mol Genet*, 10(4) :383–394.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., et Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*, 15(2) :185–197.
- Meller, V. H. (2000). Dosage compensation : making 1X equal 2X. *Trends Cell Biol*, 10(2) :54–59.
- Meller, V. H., Gordadze, P. R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R. L., et Kuroda, M. I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. *Curr Biol*, 10(3) :136–143.
- Meller, V. H. et Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J*, 21(5) :1084–1091.
- Mermoud, J. E., Costanzi, C., Pehrson, J. R., et Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J Cell Biol*, 147(7) :1399–1408.
- Mermoud, J. E., Popova, B., Peters, A. H., Jenuwein, T., et Brockdorff, N. (2002). Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. *Curr Biol*, 12(3) :247–251.
- Mette, M. F., van der Winden, J., Matzke, M., et Matzke, A. J. (2002). Short RNAs can identify new candidate transposable element families in Arabidopsis. *Plant Physiol*, 130(1) :6–9.
- Mette, M. F., van der Winden, J., Matzke, M. A., et Matzke, A. J. (1999). Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in *trans*. *EMBO J*, 18(1) :241–8.
- Meyer, E., Butler, A., Dubrana, K., Duharcourt, S., et Caron, F. (1997). Sequence-specific epigenetic effects of the maternal somatic genome on developmental rearrangements of the zygotic genome in *Paramecium primaurelia*. *Mol Cell Biol*, 17(7) :3589–3599.
- Meyer, E. et Garnier, O. (2002). Non-Mendelian inheritance and homology-dependent effects in ciliates. *Adv Genet*, 46 :305–337.

- Migeon, B. R. et Do, T. T. (1979). In search of non-random X inactivation : studies of fetal membranes heterozygous for glucose-6-phosphate dehydrogenase. *Am J Hum Genet*, 31(5) :581–585.
- Migeon, B. R., Lee, C. H., Chowdhury, A. K., et Carpenter, H. (2002). Species differences in TSIX/*Tsix* reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet*, 71(2) :286–293.
- Mitsuya, K., Meguro, M., Lee, M. P., Katoh, M., Schulz, T. C., Kugoh, H., Yoshida, M. A., Niikawa, N., Feinberg, A. P., et Oshimura, M. (1999). *LIT1*, an imprinted antisense RNA in the human *KvLQT1* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum Mol Genet*, 8(7) :1209–1217.
- Miyoshi, N., Kuroiwa, Y., Kohda, T., Shitara, H., Yonekawa, H., Kawabe, T., Hasegawa, H., Barton, S. C., Surani, M. A., Kaneko-Ishino, T., et Ishino, F. (1998). Identification of the *Meg1/Grb10* imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. *Proc Natl Acad Sci USA*, 95(3) :1102–1107.
- Miyoshi, N., Wagatsuma, H., Wakana, S., Shiroishi, T., Nomura, M., Aisaka, K., Kohda, T., Surani, M. A., Kaneko-Ishino, T., et Ishino, F. (2000). Identification of an imprinted gene, *Meg3/Gtl2* and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes Cells*, 5(3) :211–220.
- Mizuno, Y., Sotomaru, Y., Katsuzawa, Y., Kono, T., Meguro, M., Oshimura, M., Kawai, J., Tomaru, Y., Kiyosawa, H., Nikaido, I., Amanuma, H., Hayashizaki, Y., et Okazaki, Y. (2002). *Asb4*, *Ata3*, and *Dcn* are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem Biophys Res Commun*, 290(5) :1499–1505.
- Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., et Simon, J. A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*, 111(2) :197–208.
- Mochizuki, K., Fine, N. A., Fujisawa, T., et Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*, 110(6) :689–699.
- Mochizuki, K. et Gorovsky, M. A. (2004). Small RNAs in genome rearrangement in *Tetrahymena*. *Curr Opin Genet Dev*, 14(2) :181–187.
- Monk, M., Boubelik, M., et Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 99(3) :371–382.
- Moon, Y. S., Smas, C. M., Lee, K., Villena, J. A., Kim, K. H., Yun, E. J., et Sul, H. S. (2002). Mice lacking paternally expressed *Pref-1/Dlk1* display growth retardation and accelerated adiposity. *Mol Cell Biol*, 22(15) :5585–5592.
- Moore, T. (1994). Refusing the ovarian time bomb : three viewpoints and a reply. *Trends Genet*, 10(10) :347–348.
- Moore, T., Constância, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H., et Reik, W. (1997). Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. *Proc Natl Acad Sci USA*, 94(23) :12509–12514.
- Moore, T. et Haig, D. (1991). Genomic imprinting in mammalian development : a parental tug-of-war. *Trends Genet*, 7(2) :45–49.
- Morey, C., Arnaud, D., Avner, P., et Clerc, P. (2001). *Tsix*-mediated repression of *Xist* accumulation is not sufficient for normal random X inactivation. *Hum Mol Genet*, 10(13) :1403–1411.

- Morey, C., Navarro, P., Debrand, E., Avner, P., Rougeulle, C., et Clerc, P. (2004). The region 3' to *Xist* mediates X chromosome counting and H3 Lys-4 dimethylation within the *Xist* gene. *EMBO J*, 23(3) :594–604.
- Morisaki, H., Hatada, I., Morisaki, T., et Mukai, T. (1998). A novel gene, ITM, located between p57^{KIP2} and IPL, is imprinted in mice. *DNA Res*, 5(4) :235–240.
- Morris, K. V., Chan, S. W., Jacobsen, S. E., et Looney, D. J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, 305(5688) :1289–1292.
- Moss, E. G., Lee, R. C., et Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell*, 88(5) :637–646.
- Mougey, E. B., O'Reilly, M., Osheim, Y., Miller, O. L., Beyer, A., et Sollner-Webb, B. (1993). The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev.*, 7(8) :1609–1619.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., et Dreyfuss, G. (2002). miRNPs : a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev*, 16(6) :720–728.
- Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T. A., et Vaucheret, H. (2000). *Arabidopsis* *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, 101(5) :533–542.
- Murrell, A., Heeson, S., et Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat Genet*, 36(8) :889–893.
- Nakanishi, H., Obaishi, H., Satoh, A., Wada, M., Mandai, K., Satoh, K., Nishioka, H., Matsuura, Y., Mizoguchi, A., et Takai, Y. (1997). Neurabin : a novel neural tissue-specific actin filament-binding protein involved in neurite formation. *J Cell Biol*, 139(4) :951–961.
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., et Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683) :386–389.
- Narayan, P. et Rottman, F. M. (1992). Methylation of mRNA. *Adv Enzymol Relat Areas Mol Biol*, 65 :255–85.
- Narayanan, A., Lukowiak, A., Jádý, B. E., Dragon, F., Kiss, T., Terns, R. M., et Terns, M. P. (1999a). Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J*, 18(18) :5120–5130.
- Narayanan, A., Speckmann, W., Terns, R., et Terns, M. P. (1999b). Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell*, 10(7) :2131–2147.
- Nelson, P. T., Hatzigeorgiou, A. G., et Mourelatos, Z. (2004). miRNP : mRNA association in polyribosomes in a human neuronal cell line. *RNA*, 10(3) :387–394.
- Newman, D. R., Kuhn, J. F., Shanab, G. M., et Maxwell, E. S. (2000). Box C/D snoRNA-associated proteins : two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, 6(6) :861–879.
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., et Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet*, 23(1) :58–61.
- Nicholls, R. D. et Knepper, J. L. (2001). Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet*, 2 :153–175.

- Nicoloso, M., Caizergues-Ferrer, M., Michot, B., Azum, M.-C., et Bachellerie, J.-P. (1994). U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. *Mol Cell Biol*, 14(9) :5766–76.
- Nicoloso, M., Qu, L.-H., Michot, B., et Bachellerie, J.-P. (1996). Intron-encoded, antisense small nucleolar RNAs : the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol*, 260(2) :178–95.
- Niewmierzycka, A. et Clarke, S. (1999). S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J Biol Chem*, 274(2) :814–24.
- Nishikura, K. (2001). A short primer on RNAi : RNA-directed RNA polymerase acts as a key catalyst. *Cell*, 107(4) :415–418.
- Niswender, C. M., Sanders-Bush, E., et Emeson, R. B. (1998). Identification and characterization of RNA editing events within the 5-HT_{2C} receptor. *Ann N Y Acad Sci*, 861 :38–48.
- Nolan, C. M., Killian, J. K., Petitte, J. N., et Jirtle, R. L. (2001). Imprint status of M6P/IGF2R and IGF2 in chickens. *Dev Genes Evol*, 211(4) :179–183.
- Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L. G., Hume, D. A., Hayashizaki, Y., Tomita, M., RIKEN GER Group, et GSL Members (2003). Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res*, 13(6B) :1301–1306.
- Nykänen, A., Haley, B., et Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, 107(3) :309–321.
- Obata, Y., Kaneko-Ishino, T., Koide, T., Takai, Y., Ueda, T., Domeki, I., Shiroishi, T., Ishino, F., et Kono, T. (1998). Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis. *Development*, 125(8) :1553–1560.
- Ogawa, Y. et Lee, J. T. (2003). *Xite*, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol Cell*, 11(3) :731–743.
- Ogura, A., Suzuki, O., Tanemura, K., Mochida, K., Kobayashi, Y., et Matsuda, J. (1998). Development of normal mice from metaphase I oocytes fertilized with primary spermatocytes. *Proc Natl Acad Sci USA*, 95(10) :5611–5615.
- Ogura, T. et Wilkinson, A. J. (2001). AAA⁺ superfamily ATPases : common structure–diverse function. *Genes Cells*, 6(7) :575–597.
- Oh, H., Park, Y., et Kuroda, M. I. (2003). Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes Dev*, 17(11) :1334–1339.
- Ohler, U., Yekta, S., Lim, L. P., Bartel, D. P., et Burge, C. B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA*, 10(9) :1309–1322.
- Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., et Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science*, 303(5658) :644–649.
- Okamura, K., Ishizuka, A., Siomi, H., et Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev*, 18(14) :1655–1666.
- Olsen, P. H. et Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol*, 216(2) :671–680.
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R., et Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science*, 288(5465) :517–522.

- O'Neill, M. J., Ingram, R. S., Vrana, P. B., et Tilghman, S. M. (2000). Allelic expression of IGF2 in marsupials and birds. *Dev Genes Evol*, 210(1) :18–20.
- Ono, R., Shiura, H., Aburatani, H., Kohda, T., Kaneko-Ishino, T., et Ishino, F. (2003). Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. *Genome Res*, 13(7) :1696–1705.
- Onyango, P., Miller, W., Lehoczy, J., Leung, C. T., Birren, B., Wheelan, S., Dewar, K., et Feinberg, A. P. (2000). Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res*, 10(11) :1697–1710.
- Ooi, S. L., Samarsky, D. A., Fournier, M. J., et Boeke, J. D. (1998). Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme : intron length effects and activity of a precursor snoRNA. *RNA*, 4(9) :1096–1110.
- Oudejans, C. B., Westerman, B., Wouters, D., Gooyer, S., Leegwater, P. A., van Wijk, I. J., et Sleutels, F. (2001). Allelic IGF2R repression does not correlate with expression of antisense RNA in human extraembryonic tissues. *Genomics*, 73(3) :331–337.
- Pal-Bhadra, M., Bhadra, U., et Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol Cell*, 9(2) :315–327.
- Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A., et Elgin, S. C. (2004). Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science*, 303(5658) :669–672.
- Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., et Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955) :257–263.
- Palauqui, J. C., Elmayan, T., De Borne, F. D., Cr  t  , P., Charles, C., et Vaucheret, H. (1996). Frequencies, timing, and spatial patterns of co-suppression of nitrate reductase and nitrite reductase in transgenic *Tobacco* plants. *Plant Physiol*, 112(4) :1447–1456.
- Paldi, A., Gyapay, G., et Jami, J. (1995). Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. *Curr Biol*, 5(9) :1030–1035.
- Palmer, M. J., Mergner, V. A., Richman, R., Manning, J. E., Kuroda, M. I., et Lucchesi, J. C. (1993). The male-specific lethal-one (*msl-1*) gene of *Drosophila melanogaster* encodes a novel protein that associates with the X chromosome in males. *Genetics*, 134(2) :545–557.
- Panning, B., Dausman, J., et Jaenisch, R. (1997). X chromosome inactivation is mediated by *Xist* RNA stabilization. *Cell*, 90(5) :907–916.
- Papaefthimiou, I., Hamilton, A., Denti, M., Baulcombe, D., Tsagris, M., et Tabler, M. (2001). Replicating potato spindle tuber viroid RNA is accompanied by short RNA fragments that are characteristic of post-transcriptional gene silencing. *Nucleic Acids Res*, 29(11) :2395–2400.
- Pardo-Manuel de Villena, F., de la Casa-Esper  n, E., et Sapienza, C. (2000). Natural selection and the function of genome imprinting : beyond the silenced minority. *Trends Genet*, 16(12) :573–579.
- Park, W., Li, J., Song, R., Messing, J., et Chen, X. (2002a). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol*, 12(17) :1484–1495.
- Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I., et Meller, V. H. (2002b). Extent of chromatin spreading determined by *roX* RNA recruitment of MSL proteins. *Science*, 298(5598) :1620–1623.
- Parrish, S., Fleenor, J., Xu, S., Mello, C., et Fire, A. (2000). Functional anatomy of a dsRNA trigger : differential requirement for the two trigger strands in RNA interference. *Mol Cell*, 6(5) :1077–1087.

- Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., et Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808) :86–89.
- Paulsen, M., El-Maarri, O., Engemann, S., Strödicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., et Walter, J. (2000). Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum Mol Genet*, 9(12) :1829–1841.
- Peculis, B. A. (1997). The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol Cell Biol*, 17(7) :3702–3713.
- Peculis, B. A. et Steitz, J. A. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the *Xenopus* oocyte. *Cell*, 73(6) :1233–1245.
- Pederson, T. (1998). The plurifunctional nucleolus. *Nucleic Acids Res*, 26(17) :3871–6.
- Pellizzoni, L., Crosio, C., Campioni, N., Loreni, F., et Pierandrei-Amaldi, P. (1994). Different forms of U15 snoRNA are encoded in the introns of the ribosomal protein S1 gene of *Xenopus laevis*. *Nucleic Acids Res*, 22(22) :4607–4613.
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., et Brockdorff, N. (1996). Requirement for *Xist* in X chromosome inactivation. *Nature*, 379(6561) :131–137.
- Perk, J., Makedonski, K., Lande, L., Cedar, H., Razin, A., et Shemer, R. (2002). The imprinting mechanism of the Prader-Willi/Angelman regional control center. *EMBO J*, 21(21) :5807–5814.
- Peters, A. H., Mermoud, J. E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., et Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet*, 30(1) :77–80.
- Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., Williamson, C. M., et Kelsey, G. (1999). A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. *Proc Natl Acad Sci USA*, 96(7) :3830–3835.
- Pfeffer, S., Dunoyer, P., Heim, F., Richards, K. E., Jonard, G., et Ziegler-Graff, V. (2002). P0 of beet Western yellows virus is a suppressor of posttranscriptional gene silencing. *J Virol*, 76(13) :6815–6824.
- Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C., et Tuschl, T. (2004). Identification of virus-encoded microRNAs. *Science*, 304(5671) :734–736.
- Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., et Sontheimer, E. J. (2004). A Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell*, 117(1) :83–94.
- Piras, G., El Kharroubi, A., Kozlov, S., Escalante-Alcalde, D., Hernandez, L., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., et Stewart, C. L. (2000). *Zac1* (*Lot1*), a potential tumor suppressor gene, and the gene for ϵ -sarcoglycan are maternally imprinted genes : identification by a subtractive screen of novel uniparental fibroblast lines. *Mol Cell Biol*, 20(9) :3308–3315.
- Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., Held, W. A., Hayashizaki, Y., et Chapman, V. M. (1996). Identification of *Grf1* on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat Genet*, 14(1) :106–109.
- Plasterk, R. H. (2002). RNA silencing : the genome's immune system. *Science*, 296(5571) :1263–1265.

- Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., et Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science*, 300(5616) :131–135.
- Pondel, M. (2000). Calcitonin and calcitonin receptors : bone and beyond. *Int J Exp Pathol*, 81(6) :405–422.
- Proudfoot, N. J. (1986). Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature*, 322(6079) :562–565.
- Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., et Rådmark, O. (2002a). Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J*, 21(21) :5864–5874.
- Provost, P., Silverstein, R. A., Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., Wright, A., Samuelsson, B., Rådmark, O., et Ekwall, K. (2002b). Dicer is required for chromosome segregation and gene silencing in fission yeast cells. *Proc Natl Acad Sci USA*, 99(26) :16648–16653.
- Qian, N., Frank, D., O'Keefe, D., Dao, D., Zhao, L., Yuan, L., Wang, Q., Keating, M., Walsh, C., et Tycko, B. (1997). The IPL gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. *Hum Mol Genet*, 6(12) :2021–2029.
- Qu, L.-H., Henras, A., Lu, Y.-J., Zhou, H., Zhou, W.-X., Zhu, Y.-Q., Zhao, J., Henry, Y., Caizergues-Ferrer, M., et Bachellerie, J.-P. (1999). Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol*, 19(2) :1144–58.
- Qu, L.-H., Henry, Y., Nicoloso, M., Michot, B., Azum, M.-C., Renalier, M.-H., Caizergues-Ferrer, M., et Bachellerie, J.-P. (1995). U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA. *Nucleic Acids Res*, 23(14) :2669–76.
- Qu, L.-H., Nicoloso, M., Michot, B., Azum, M.-C., Caizergues-Ferrer, M., Renalier, M.-H., et Bachellerie, J.-P. (1994). U21, a novel small nucleolar RNA with a 13 nt. complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals. *Nucleic Acids Res*, 22(20) :4073–81.
- Rastan, S. (1982). Primary non-random X-inactivation caused by controlling elements in the mouse demonstrated at the cellular level. *Genet Res*, 40(2) :139–147.
- Rastan, S. et Robertson, E. J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp Morphol*, 90 :379–388.
- Reed, J. C., Kasschau, K. D., Prokhnovsky, A. I., Gopinath, K., Pogue, G. P., Carrington, J. C., et Dolja, V. V. (2003). Suppressor of RNA silencing encoded by Beet yellows virus. *Virology*, 306(2) :203–209.
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*, 15(3) :326–331.
- Reik, W. et Walter, J. (2001). Genomic imprinting : parental influence on the genome. *Nat Rev Genet*, 2(1) :21–32.
- Reinhart, B. J. et Bartel, D. P. (2002). Small RNAs correspond to centromere heterochromatic repeats. *Science*, 297(5588) :1831.
- Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., et Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403(6772) :901–906.

- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., et Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev*, 16(13) :1616–1626.
- Renalier, M.-H., Nicoloso, M., Qu, L. H., et Bachellerie, J.-P. (1996). SnoRNA U21 is also intron-encoded in *Drosophila melanogaster* but in a different host-gene as compared to warm-blooded vertebrates. *FEBS Lett*, 379(3) :212–216.
- Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., et Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, 110(4) :513–520.
- Richard, P., Darzacq, X., Bertrand, E., Jádý, B. E., Verheggen, C., et Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J*, 22(16) :4283–4293.
- Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., et Wolffe, A. P. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet*, 25(3) :338–342.
- Robinson, W. P. et Lalonde, M. (1995). Sex-specific meiotic recombination in the Prader-Willi/Angelman syndrome imprinted region. *Hum Mol Genet*, 4(5) :801–806.
- Rountree, M. R., Bachman, K. E., et Baylin, S. B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet*, 25(3) :269–277.
- Rudenko, G. N., Ono, A., et Walbot, V. (2003). Initiation of silencing of maize *MuDR/Mu* transposable elements. *Plant J*, 33(6) :1013–1025.
- Rudert, F., Bronner, S., Garnier, J.-M., et Dolle, P. (1995). Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. *Mamm Genome*, 6(2) :76–83.
- Runte, M., Hüttenhofer, A., Groß, S., Kiefmann, M., Horsthemke, B., et Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum Mol Genet*, 10(23) :2687–2700.
- Russell, L. B. (1963). Mammalian X-chromosome action : inactivation limited in spread and region of origin. *Science*, 140 :976–978.
- Sabin-Teyssier, E. (2002). MBI-68 is a cytoplasmic small nucleolar RNA. *J Artifacts*, 49(35) :512–1024.
- Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T., et Li, E. (2000). X inactivation in the mouse embryo deficient for *Dnmt1* : distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol*, 225(2) :294–303.
- Samarsky, D. A., Fournier, M. J., Singer, R. H., et Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J*, 17(13) :3747–3757.
- Sandell, L. L., Guan, X. J., Ingram, R., et Tilghman, S. M. (2003). Gtm, a creatine synthesis enzyme, is imprinted in mouse placenta. *Proc Natl Acad Sci USA*, 100(8) :4622–4627.
- Sarot, E., Payen-Groschêne, G., Bucheton, A., et Péliisson, A. (2004). Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics*, 166(3) :1313–1321.
- Saxena, S., Jonsson, Z. O., et Dutta, A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem*, 278(45) :44312–44319.
- Schmidt, J. V., Matteson, P. G., Jones, B. K., Guan, X. J., et Tilghman, S. M. (2000). The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes Dev*, 14(16) :1997–2002.
- Schoenherr, C. J., Levorse, J. M., et Tilghman, S. M. (2003). CTCF maintains differential methylation at the *Igf2/H19* locus. *Nat Genet*, 33(1) :66–69.

- Scholzen, T., Solursh, M., Suzuki, S., Reiter, R., Morgan, J. L., Buchberg, A. M., Siracusa, L. D., et Iozzo, R. V. (1994). The murine decorin. Complete cDNA cloning, genomic organization, chromosomal assignment, and expression during organogenesis and tissue differentiation. *J Biol Chem*, 269(45) :28270–28281.
- Schramke, V. et Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science*, 301(5636) :1069–1074.
- Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., et Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2) :199–208.
- Schwarz, D. S., Hutvagner, G., Haley, B., et Zamore, P. D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol Cell*, 10(3) :537–548.
- Schwarz, D. S., Tomari, Y., et Zamore, P. D. (2004). The RNA-induced silencing complex is a Mg^{2+} -dependent endonuclease. *Curr Biol*, 14(9) :787–791.
- Seitz, H., Royo, H., Bortolin, M.-L., Lin, S.-P., Ferguson-Smith, A. C., et Cavaillé, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Research*, 14(9) :1741–1748.
- Seitz, H., Youngson, N., Lin, S.-P., Dalbert, S., Paulsen, M., Bachellerie, J.-P., Ferguson-Smith, A. C., et Cavaillé, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet*, 34(3) :261–262.
- Sharma, K. et Tollervey, D. (1999). Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol*, 19(9) :6012–6019.
- Sharman, G. B. (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature*, 230(5291) :231–232.
- Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G. C., Rastan, S., et Brockdorff, N. (1997). Stabilization of *Xist* RNA mediates initiation of X chromosome inactivation. *Cell*, 91(1) :99–107.
- Shi, H., Djikeng, A., Tschudi, C., et Ullu, E. (2004). Argonaute protein in the early divergent eukaryote *Trypanosoma brucei* : control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol*, 24(1) :420–427.
- Shinagawa, T. et Ishii, S. (2003). Generation of *Ski*-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter. *Genes Dev*, 17(11) :1340–1345.
- Shou, C., Farnsworth, C. L., Neel, B. G., et Feig, L. A. (1992). Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor for Ras p21. *Nature*, 358(6384) :351–354.
- Shuster, M., Dhar, M. S., Olins, A. L., Olins, D. E., Howell, C. Y., Gollin, S. M., et Chaillet, J. R. (1998). Parental alleles of an imprinted mouse transgene replicate synchronously. *Dev Genet*, 23(4) :275–284.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K. L., Parrish, S., Timmons, L., Plasterk, R. H., et Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, 107(4) :465–476.
- Sijen, T. et Plasterk, R. H. (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*, 426(6964) :310–314.
- Sijen, T., Wellink, J., Hiriart, J. B., et Van Kammen, A. (1996). RNA-mediated virus resistance : Role of repeated transgenes and delineation of targeted regions. *Plant Cell*, 8(12) :2277–2294.
- Silhavy, D., Molnar, A., Luciola, A., Szittyá, G., Hornyik, C., Tavazza, M., et Burgyan, J. (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO J*, 21(12) :3070–3080.

- Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P., et Brockdorff, N. (2003). Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell*, 4(4) :481–495.
- Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J., et Plasterk, R. H. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol*, 12(15) :1317–1319.
- Simmler, M. C., Cattanaach, B. M., Rasberry, C., Rougeulle, C., et Avner, P. (1993). Mapping the murine *Xce* locus with (CA)_n repeats. *Mamm Genome*, 4(9) :523–530.
- Simon, I., Tenzen, T., Reubinooff, B. E., Hillman, D., McCarrey, J. R., et Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature*, 401(6756) :929–932.
- Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., et Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell*, 5(4) :659–669.
- Sleutels, F. et Barlow, D. P. (2002). The origins of genomic imprinting in mammals. *Adv Genet*, 46 :119–163.
- Sleutels, F., Tjon, G., Ludwig, T., et Barlow, D. P. (2003). Imprinted silencing of *Slc22a2* and *Slc22a3* does not need transcriptional overlap between *Igf2r* and *Air*. *EMBO J*, 22(14) :3696–3704.
- Sleutels, F., Zwart, R., et Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, 415(6873) :810–813.
- Smas, C. M. et Sul, H. S. (1993). Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell*, 73(4) :725–734.
- Smilnich, N. J., Day, C. D., Fitzpatrick, G. V., Caldwell, G. M., Lossie, A. C., Cooper, P. R., Smallwood, A. C., Joyce, J. A., Schofield, P. N., Reik, W., Nicholls, R. D., Weksberg, R., Driscoll, D. J., Maher, E. R., Shows, T. B., et Higgins, M. J. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci USA*, 96(14) :8064–8069.
- Smit, M., Segers, K., Carrascosa, L. G., Shay, T., Baraldi, F., Gyapay, G., Snowden, G., Georges, M., Cockett, N., et Charlier, C. (2003). Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics*, 163(1) :453–456.
- Smith, C. J., Watson, C. F., Bird, C. R., Ray, J., Schuch, W., et Grierson, D. (1990). Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants. *Mol Gen Genet*, 224(3) :477–481.
- Smith, C. M. et Steitz, J. A. (1998). Classification of *gas5* as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snorna host genes. *Mol Cell Biol*, 18(12) :6897–909.
- Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D., et Lucchesi, J. C. (2000). The *drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol*, 20(1) :312–318.
- Smith, H. A., Swaney, S. L., Parks, T. D., Wernsman, E. A., et Dougherty, W. G. (1996). Transgenic plant virus resistance mediated by untranslatable sense RNAs : expression, regulation, and fate of nonessential RNAs. *Plant Cell*, 6(10) :1441–1453.
- Smith, R. J., Dean, W., Konfortova, G., et Kelsey, G. (2003). Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res*, 13(4) :558–569.

- Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. *Annu Rev Genet*, 22 :127–146.
- Solter, D. (1994). Refusing the ovarian time bomb : three viewpoints and a reply. *Trends Genet*, 10(10) :346.
- Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., et Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol*, 10(12) :1026–1032.
- Song, J. J., Smith, S. K., Hannon, G. J., et Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC Slicer activity. *Science*, 305(5689) :1434–1437.
- Spahn, L. et Barlow, D. P. (2003). An ICE pattern crystallizes. *Nat Genet*, 35(1) :11–12.
- Srivastava, M., Hsieh, S., Grinberg, A., Williams-Simons, L., Huang, S. P., et Pfeifer, K. (2000). *H19* and *Igf2* monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of *H19*. *Genes Dev*, 14(10) :1186–1195.
- Stahl, G., Bidou, L., Rousset, J.-P., et Cassan, M. (1995). Versatile vectors to study recoding : conservation of rules between yeast and mammalian cells. *Nucleic Acids Res*, 23(9) :1557–1560.
- Stark, A., Brennecke, J., Russell, R. B., et Cohen, S. M. (2003). Identification of *Drosophila* MicroRNA Targets. *PLoS Biol*, 1(3) :E60.
- Steitz, J. A. et Tycowski, K. T. (1995). Small RNA chaperones for ribosome biogenesis. *Science*, 270(5242) :1626–1627.
- Steward, O. et Schuman, E. M. (2003). Compartmentalized synthesis and degradation of proteins in neurons. *Neuron*, 40(2) :347–359.
- Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., et Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell*, 73(1) :61–71.
- Stuckenholz, C., Meller, V. H., et Kuroda, M. I. (2003). Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in *Drosophila melanogaster*. *Genetics*, 164(3) :1003–1014.
- Sugawara, M., Nakanishi, T., Fei, Y. J., Martindale, R. G., Ganapathy, M. E., Leibach, F. H., et Ganapathy, V. (2000). Structure and function of ATA3, a new subtype of amino acid transport system A, primarily expressed in the liver and skeletal muscle. *Biochim Biophys Acta*, 1509(1-2) :7–13.
- Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., Cha, K. Y., Chung, H. M., Yoon, H. S., Moon, S. Y., Kim, V. N., et Kim, K. S. (2004). Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*, 270(2) :488–498.
- Surani, M. A., Barton, S. C., et Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308(5959) :548–550.
- Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J., et Schultz, R. M. (2004a). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol*, 269(1) :276–285.
- Svoboda, P., Stein, P., Filipowicz, W., et Schultz, R. M. (2004b). Lack of homologous sequence-specific DNA methylation in response to stable dsRNA expression in mouse oocytes. *Nucleic Acids Res*, 32(12) :3601–3606.
- Svoboda, P., Stein, P., et Schultz, R. M. (2001). RNAi in mouse oocytes and preimplantation embryos : effectiveness of hairpin dsRNA. *Biochem Biophys Res Commun*, 287(5) :1099–1104.

- Szabó, P., Tang, S. H., Rentsendorj, A., Pfeifer, G. P., et Mann, J. R. (2000). Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function. *Curr Biol*, 10(10) :607–610.
- Szabó, P. E., Pfeifer, G. P., et Mann, J. R. (2004a). Parent-of-origin-specific binding of nuclear hormone receptor complexes in the *H19-Igf2* imprinting control region. *Mol Cell Biol*, 24(11) :4858–4868.
- Szabó, P. E., Tang, S. H., Silva, F. J., Tsark, W. M., et Mann, J. R. (2004b). Role of CTCF binding sites in the *Igf2/H19* imprinting control region. *Mol Cell Biol*, 24(11) :4791–4800.
- Szeto, I. Y., Li, L. L., et Surani, M. A. (2000). *Ocat*, a paternally expressed gene closely linked and transcribed in the opposite direction to *Peg3*. *Genomics*, 67(2) :221–227.
- Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., et Mello, C. C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell*, 99(2) :123–132.
- Takada, S., Paulsen, M., Tevendale, M., Tsai, C. E., Kelsey, G., Cattanach, B. M., et Ferguson-Smith, A. C. (2002). Epigenetic analysis of the *Dlk1-Gtl2* imprinted domain on mouse chromosome 12 : implications for imprinting control from comparison with *Igf2-H19*. *Hum Mol Genet*, 11(1) :77–86.
- Takada, S., Tevendale, M., Baker, J., Georgiades, P., Campbell, E., Freeman, T., Johnson, M. H., Paulsen, M., et Ferguson-Smith, A. C. (2000). *Delta-like* and *gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr Biol*, 10(18) :1135–1138.
- Takagi, N. et Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, 256(5519) :640–642.
- Takagi, N., Sugawara, O., et Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma*, 85(2) :275–286.
- Takeda, A., Sugiyama, K., Nagano, H., Mori, M., Kaido, M., Mise, K., Tsuda, S., et Okuno, T. (2002). Identification of a novel RNA silencing suppressor, NSs protein of *Tomato spotted wilt virus*. *FEBS Lett*, 532(1-2) :75–79.
- Tanaka, M., Puchyr, M., Gertsenstein, M., Harpal, K., Jaenisch, R., Rossant, J., et Nagy, A. (1999). Parental origin-specific expression of *Mash2* is established at the time of implantation with its imprinting mechanism highly resistant to genome-wide demethylation. *Mech Dev*, 87(1-2) :129–142.
- Tang, G., Reinhart, B. J., Bartel, D. P., et Zamore, P. D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev*, 17(1) :49–63.
- Tang, T. H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J., et Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci USA*, 99(11) :7536–7541.
- Taverna, S. D., Coyne, R. S., et Allis, C. D. (2002). Methylation of histone H3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell*, 110(6) :701–711.
- Thorvaldsen, J. L., Duran, K. L., et Bartolomei, M. S. (1998). Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev*, 12(23) :3693–3702.
- Thorvaldsen, J. L., Mann, M. R., Nwoko, O., Duran, K. L., et Bartolomei, M. S. (2002). Analysis of sequence upstream of the endogenous *H19* gene reveals elements both essential and dispensable for imprinting. *Mol Cell Biol*, 22(8) :2450–2462.

- Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., et Hurt, E. C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell*, 72(3) :443–57.
- Tomari, Y., Du, T., Haley, B., Schwarz, D. S., Bennett, R., Cook, H. A., Koppetsch, B. S., Theurkauf, W. E., et Zamore, P. D. (2004). RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell*, 116(6) :831–841.
- Tonkin, L. A. et Bass, B. L. (2003). Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science*, 302(5651) :1725.
- Tsai, C. E., Lin, S.-P., Ito, M., Takagi, N., Takada, S., et Ferguson-Smith, A. C. (2002). Genomic imprinting contributes to thyroid hormone metabolism in the mouse embryo. *Curr Biol*, 12(14) :1221–1226.
- Tsai, T. F., Jiang, Y. H., Bressler, J., Armstrong, D., et Beaudet, A. L. (1999). Paternal deletion from *Snrpn* to *Ube3a* in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader-Willi syndrome. *Hum Mol Genet*, 8(8) :1357–1364.
- Tufarelli, C., Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., et Higgs, D. R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet*, 34(2) :157–165.
- Tycowski, K. T., Shu, M. D., et Steitz, J. A. (1994). Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science*, 266(5190) :1558–1561.
- Tycowski, K. T., Shu, M. D., et Steitz, J. A. (1996a). A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, 379(6564) :464–6.
- Tycowski, K. T., Smith, C. M., Shu, M. D., et Steitz, J. A. (1996b). A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proc Natl Acad Sci USA*, 93 :14480–14485.
- Tycowski, K. T. et Steitz, J. A. (2001). Non-coding snoRNA host genes in *Drosophila* : expression strategies for modification guide snoRNAs. *Eur J Cell Biol*, 80(2) :119–125.
- Tycowski, K. T., You, Z.-H., Graham, P. J., et Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell*, 2(5) :629–38.
- Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., Fujimoto, H., Shibata, H., Hayashizaki, Y., et Sasaki, H. (2000). The paternal methylation imprint of the mouse *H19* locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells*, 5(8) :649–659.
- Uetsuki, T., Takagi, K., Sugiura, H., et Yoshikawa, K. (1996). Structure and expression of the mouse *necdin* gene. Identification of a postmitotic neuron-restrictive core promoter. *J Biol Chem*, 271(2) :918–924.
- Vaistij, F. E., Jones, L., et Baulcombe, D. C. (2002). Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell*, 14(4) :857–867.
- Van Blokland, R., Van der Geest, N., Mol, J. N. M., et Kooter, J. M. (1994). Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. *Plant J*, 6(6) :861–877.
- van De Sluis, B., Rothuizen, J., Pearson, P. L., van Oost, B. A., et Wijmenga, C. (2002). Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum Mol Genet*, 11(2) :165–173.
- Vance, V. et Vaucheret, H. (2001). RNA silencing in plants – defense and counterdefense. *Science*, 292(5525) :2277–2280.

- Varmuza, S. et Mann, M. (1994). Genomic imprinting – defusing the ovarian time bomb. *Trends Genet*, 10(4) :118–123.
- Varraut, A., Ciani, E., Apiou, F., Bilanges, B., Hoffmann, A., Pantaloni, C., Bockaert, J., Spengler, D., et Journot, L. (1998). hZAC encodes a zinc finger protein with antiproliferative properties and maps to a chromosomal region frequently lost in cancer. *Proc Natl Acad Sci USA*, 95(15) :8835–8840.
- Vazquez, F., Gasciolli, V., Crété, P., et Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol*, 14(4) :346–351.
- Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K., et Slack, F. J. (2004). The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3'UTR. *Genes Dev*, 18(2) :132–137.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., et Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, 303(5658) :672–676.
- Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., et Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, 6(6) :1331–1342.
- Vielle-Calzada, J.-P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M. A., et Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic *DDM1* activity. *Genes Dev*, 13(22) :2971–2982.
- Vitali, P., Royo, H., Seitz, H., Bachellerie, J.-P., Hüttenhofer, A., et Cavaillé, J. (2003). Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res*, 31(22) :6543–6551.
- Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends Genet*, 17(8) :449–459.
- Voinnet, O., Lederer, C., et Baulcombe, D. C. (2000). A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell*, 103(1) :157–167.
- Voinnet, O., Pinto, Y. M., et Baulcombe, D. C. (1999). Suppression of gene silencing : a general strategy used by diverse DNA and RNA viruses of plants. *Proc Natl Acad Sci USA*, 96(24) :14147–14152.
- Voinnet, O., Vain, P., Angell, S., et Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*, 95(2) :177–187.
- Volpe, T., Schramke, V., Hamilton, G. L., White, S. A., Teng, G., Martienssen, R. A., et Allshire, R. C. (2003). RNA interference is required for normal centromere function in fission yeast. *Chromosome Res*, 11(2) :137–146.
- Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., et Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297(5588) :1833–1837.
- Vu, T. H., Li, T., et Hoffman, A. R. (2004). Promoter-restricted histone code, and not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. *Hum Mol Genet*, 13(19) :2233–2245.
- Wade, P. A., Geggion, A., Jones, P. L., Ballestar, E., Aubry, F., et Wolffe, A. P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet*, 23(1) :62–66.
- Wang, H., Boisvert, D., Kim, K. K., Kim, R., et Kim, S. H. (2000). Crystal structure of a fibrillar homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J*, 19(3) :317–23.

- Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J. C., Nagy, A., et Magnuson, T. (2001a). Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet*, 28(4) :371–375.
- Wang, J. F., Zhou, H., Chen, Y. Q., Luo, Q. J., et Qu, L. H. (2004a). Identification of 20 microRNAs from *Oryza sativa*. *Nucleic Acids Res*, 32(5) :1688–1695.
- Wang, Q., Curran, M. E., Splawski, I., Burn, T. C., Millholland, J. M., VanRaay, T. J., Shen, J., Timothy, K. W., Vincent, G. M., de Jager, T., Schwartz, P. J., Toubin, J. A., Moss, A. J., Atkinson, D. L., Landes, G. M., Connors, T. D., et Keating, M. T. (1996). Positional cloning of a novel potassium channel gene : KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet*, 12(1) :17–23.
- Wang, Y., Joh, K., Masuko, S., Yatsuki, H., Soejima, H., Nabetani, A., Beechey, C. V., Okinami, S., et Mukai, T. (2004b). The mouse *Murr1* gene is imprinted in the adult brain, presumably due to transcriptional interference by the antisense-oriented *U2af1-rs1* gene. *Mol Cell Biol*, 24(1) :270–279.
- Wang, Y., Zhang, W., Jin, Y., Johansen, J., et Johansen, K. M. (2001b). The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in *Drosophila*. *Cell*, 105(4) :433–443.
- Ward, D. F. et Murray, N. E. (1979). Convergent transcription in bacteriophage lambda : interference with gene expression. *J Mol Biol*, 133(2) :249–266.
- Warnecke, P. M., Mann, J. R., Frommer, M., et Clark, S. J. (1998). Bisulfite sequencing in preimplantation embryos : DNA methylation profile of the upstream region of the mouse imprinted *H19* gene. *Genomics*, 51(2) :182–190.
- Wassenegger, M. (2002). Gene silencing-based disease resistance. *Transgenic Res*, 11(6) :639–653.
- Wassenegger, M. et Péliissier, T. (1998). A model for RNA-mediated gene silencing in higher plants. *Plant Mol. Biol.*, 37(2) :349–362.
- Waterhouse, P. M., Graham, M. W., et Wang, M. B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci USA*, 95(23) :13959–13964.
- Watkins, N. J., Dickmanns, A., et Lührmann, R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol*, 22(23) :8342–8352.
- Watkins, N. J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., et Lührmann, R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, 103(3) :457–466.
- Weber, M., Hagège, H., Murrell, A., Brunel, C., Reik, W., Cathala, G., et Forné, T. (2003). Genomic imprinting controls matrix attachment regions in the *Igf2* gene. *Mol Cell Biol*, 23(24) :8953–8959.
- Weinstein, L. S., Liu, J., Sakamoto, A., Xie, T., et Chen, M. (2004). GNAS : Normal and abnormal functions. *Endocrinology*, Sous presse.
- Weisstein, A. E. et Spencer, H. G. (2003). The evolution of genomic imprinting via variance minimization : an evolutionary genetic model. *Genetics*, 165(1) :205–222.
- Wentworth, B. M., Schaefer, I. M., Villa-Komaroff, L., et Chirgwin, J. M. (1986). Characterization of the two nonallelic genes encoding mouse preproinsulin. *J Mol Evol*, 23(4) :305–312.
- Wevrick, R. et Francke, U. (1997). An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome (IPW) gene. *Hum Mol Genet*, 6(2) :325–332.

- Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E., et Plasterk, R. H. (2003). The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet*, 35(3) :217–218.
- Wightman, B., Burglin, T. R., Gatto, J., Arasu, P., et Ruvkun, G. (1991). Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes Dev*, 5(10) :1813–1824.
- Wightman, B., Ha, I., et Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, 75(5) :855–862.
- Wilkins, J. F. et Haig, D. (2003). What good is genomic imprinting : the function of parent-specific gene expression. *Nat Rev Genet*, 4(5) :359–368.
- Williamson, C. M., Ball, S. T., Nottingham, W. T., Skinner, J. A., Plagge, A., Turner, M. D., Powles, N., Hough, T., Papworth, D., Fraser, W. D., Maconochie, M., et Peters, J. (2004). A cis-acting control region is required exclusively for the tissue-specific imprinting of *Gnas*. *Nat Genet*, 36(8) :894–899.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M. J., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T., et Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407(6802) :327–339.
- Wroe, S. F., Kelsey, G., Skinner, J. A., Bodle, D., Ball, S. T., Beechey, C. V., Peters, J., et Williamson, C. M. (2000). An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse *Gnas* locus. *Proc Natl Acad Sci USA*, 97(7) :3342–3346.
- Wu-Scharf, D., Jeong, B., Zhang, C., et Cerutti, H. (2000). Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science*, 290(5494) :1159–1162.
- Wutz, A., Rasmussen, T. P., et Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat Genet*, 30(2) :167–174.
- Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., et Barlow, D. P. (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature*, 389(6652) :745–749.
- Wyss, M. et Kaddurah-Daouk, R. (2000). Creatine and creatinine metabolism. *Physiol Rev*, 80(3) :1107–1213.
- Xie, Z., Kasschau, K. D., et Carrington, J. C. (2003). Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr Biol*, 13(9) :784–789.
- Xu, P., Vernooy, S. Y., Guo, M., et Hay, B. A. (2003). The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol*, 13(9) :790–795.
- Xu, Y., Goodyer, C. G., Deal, C., et Polychronakos, C. (1993). Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochem Biophys Res Commun*, 197(2) :747–754.
- Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., et Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, 426(6965) :468–474.
- Yang, D., Lu, H., et Erickson, J. W. (2000). Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos. *Curr Biol*, 10(19) :1191–1200.
- Yang, Y., Li, T., Vu, T. H., Ulaner, G. A., Hu, J. F., et Hoffman, A. R. (2003). The histone code regulating expression of the imprinted mouse *Igf2r* gene. *Endocrinology*, 144(12) :5658–5670.
- Yao, M. C., Fuller, P., et Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, 300(5625) :1581–1584.

- Yekta, S., Shih, I.-h., et Bartel, D. P. (2004). MicroRNA-directed cleavage of *HOXB8* mRNA. *Science*, 304(5670) :594–596.
- Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., et Rotman, G. (2003). Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol*, 21(4) :379–386.
- Yelina, N. E., Savenkov, E. I., Solovyev, A. G., Morozov, S. Y., et Valkonen, J. P. (2002). Long-distance movement, virulence, and RNA silencing suppression controlled by a single protein in hordei- and potyviruses : complementary functions between virus families. *J Virol*, 76(24) :12981–12991.
- Yevtodiyenko, A., Carrn, M. S., Patel, N., et Schmidt, J. V. (2002). Analysis of candidate imprinted genes linked to *Dlk1-Gtl2* using a congenic mouse line. *Mamm Genome*, 13(11) :633–638.
- Yevtodiyenko, A., Steshina, E. Y., Farner, S. C., Levorse, J. M., et Schmidt, J. V. (2004). A 178-kb BAC transgene imprints the mouse *Gtl2* gene and localizes tissue-specific regulatory elements. *Genomics*, 84(2) :277–287.
- Yi, R., Qin, Y., Macara, I. G., et Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, 17(24) :3011–3016.
- Yi-Brunozzi, H. Y., Easterwood, L. M., Kamilar, G. M., et Beal, P. A. (1999). Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2. *Nucleic Acids Res*, 27(14) :2912–7.
- Yoder, J. A., Walsh, C. P., et Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet*, 13(8) :335–340.
- Yoon, B. J., Herman, H., Sikora, A., Smith, L. T., Plass, C., et Soloway, P. D. (2002). Regulation of DNA methylation of *Rasgrf1*. *Nat Genet*, 30(1) :92–96.
- Yu, S., Gavrilova, O., Chen, H., Lee, R., Liu, J., Pacak, K., Parlow, A. F., Quon, M. J., Reitman, M. L., et Weinstein, L. S. (2000). Paternal versus maternal transmission of a stimulatory G-protein alpha subunit knockout produces opposite effects on energy metabolism. *J Clin Invest*, 105(5) :615–623.
- Yu, S., Yu, D., Lee, E., Eckhaus, M., Lee, R., Corria, Z., Accili, D., Westphal, H., et Weinstein, L. S. (1998). Variable and tissue-specific hormone resistance in heterotrimeric Gs protein α -subunit ($G_s\alpha$) knockout mice is due to tissue-specific imprinting of the $G_s\alpha$ gene. *Proc Natl Acad Sci USA*, 95(15) :8715–8720.
- Yu, Y.-T., Scharl, E. C., Smith, C. M., et Steitz, J. A. (1999). *The RNA world*, chapitre The growing world of small ribonucleoproteins, pages 487–524. Cold Spring Harbor Laboratory Press, 2ème édition.
- Yuan, G., Klämbt, C., Bachellerie, J.-P., Brosius, J., et Hüttenhofer, A. (2003). RNomics in *Drosophila melanogaster* : identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res*, 31(10) :2495–2507.
- Zamore, P. D. (2002). Ancient pathways programmed by small RNAs. *Science*, 296(5571) :1265–1269.
- Zamore, P. D., Tuschl, T., Sharp, P. A., et Bartel, D. P. (2000). RNAi : double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1) :25–33.
- Zeng, Y. et Cullen, B. R. (2003). Sequence requirements for micro RNA processing and function in human cells. *RNA*, 9(1) :112–123.
- Zeng, Y., Yi, R., et Cullen, B. R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci USA*, 100(17) :9779–9784.

- Zhang, H., Kolb, F. A., Brondani, V., Billy, E., et Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J*, 21(21) :5875–5885.
- Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., et Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1) :57–68.
- Zhao, J., Hyman, L., et Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes : mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*, 63(2) :405–445.
- Zhou, H., Zhao, J., Yu, C. H., Luo, Q. J., Chen, Y. Q., Xiao, Y., et Qu, L. H. (2004). Identification of a novel box C/D snoRNA from mouse nucleolar cDNA library. *Gene*, 327(1) :99–105.
- Zhou, S., Yang, Y., Scott, M. J., Pannuti, A., Fehr, K. C., Eisen, A., Koonin, E. V., Fouts, D. L., Wrightsman, R., Manning, J. E., et Lucchesi, J. C. (1995). Male-specific lethal 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO J*, 14(12) :2884–2895.
- Zwart, R., Sleutels, F., Wutz, A., Schinkel, A. H., et Barlow, D. P. (2001). Bidirectional action of the *Igf2r* imprint control element on upstream and downstream imprinted genes. *Genes Dev*, 15(18) :2361–2366.

Index

A

adipocytes 37, 75, 147, 148
 androgénètes 22
 Angiospermes 22
Arabidopsis thaliana .. 22, 68, 76, 77, 83, 89,
 154
 Archæ 48
 ARN C/D introniques 47, 53
 ARN_m 5-HT_{2C} 57, 150
 asymétrie des petits ARN double brin ... 65

B

bombe à retardement ovarienne 37

C

cellules germinales primordiales 25–27
 centre d’empreinte 23, 28, 29, 32, 34, 35
 chaperones ribonucléiques 51
 clivage des ARN pré-ribosomiques 49
 corpuscules de Cajal voir scaRNA

D

déméthylation active 23, 26
 différenciation macronucléaire 85
 disomies uniparentales 21
 distinction des chromosomes parentaux .. 40
 DMR secondaire 28, 29
 DMR voir région différentiellement
 méthylée

E

Eed 95
 effacement de l’empreinte 25
Escherichia coli 62
 estimation du nombre de gènes soumis à l’em-
 preinte 24
 extinction de parasites génomiques .. 38, 79,
 151

F

facilitation de l’évolution 38

G

gènes HOX 76

guidage de la 2’-O-méthylation 49

I

identification de gènes soumis à l’empreinte
 21
 îlot CpG 33, 93, 98
 inosine 57, 90
 interférence transcriptionnelle voir occlusion
 de promoteur
 invalidation de Dicer 63, 77, 80, 88

L

lecture de l’empreinte 29
 localisation de la maturation des microARN
 65
 loci soumis à l’empreinte 23
 locus *Dlk1/Gtl2* 36
 locus *Igf2/H19* 34
 locus *Igf2r/Air* 32
 locus *Kcnq1* 33
 locus PWS/AS 35

M

maintien de l’empreinte 29
 Maïs 22, 80
 Mammifères placentaires 22, 162
 Marsupiaux 22, 92, 152, 162
Meg3 36
 méthyl-transférase. 25, 27, 46, 82, 84, 89, 95
 mise en place de l’empreinte 26
 modulation de l’expression des gènes 39

N

nature de l’empreinte 25

O

occlusion de promoteur 98

P

Paramécie 87
 parthénogénètes 21, 22, 38
 pluricellularité 78
 polymérase à ARN dépendant de l’ARN voir
 RdRP

<i>Pref-1</i>	36
protéines associées aux ARN C/D	46
PTGS	83

R

RdRP	68, 80, 89
Recherche de complémentarités <i>in silico</i>	72, 142, 174
recherche de gènes de snoRNA C/D	115
région différenciellement méthylée	25, 32–34, 122
réplication asynchrone	23
répresseur viral du RNAi	79, 80
RISC	63, 68, 71
RITS	83
rôle de la 2'-O-méthylation	50

S

scaRNA	50
séquençage du génome humain	20
sérotonine	voir ARN _m 5-HT _{2C}
similitudes entre microARN et siRNA	62
snoRNA et scaRNA H/ACA	50
stRNA	75
syndrome d'Angelman	35
syndrome de Beckwith-Wiedemann	33
syndrome de Prader-Willi	35, 58

T

taux de recombinaison	24
<i>Tetrahymena</i>	85
théorie du conflit	39
transplantation de pronuclei	21

Genomic imprinting and small non-coding RNAs

Abstract

Genomic imprinting results in a differential expression of both alleles of some genes, according to their parental origin. This study led to the identification of many mammalian small non-coding RNA genes (C/D RNA and microRNA genes) which are subjected to genomic imprinting, and to a preliminary characterization of their expression. Whereas most C/D RNAs participate to ribosomal RNA or small nuclear RNA biogenesis, the ones we describe don't seem to; microRNAs are usually post-transcriptional repressors of specific genes : among the microRNAs we describe, two could repress a retrotransposon. Potential functions for those small non-coding RNA genes (in development control, retrotransposon silencing and genomic imprinting mechanism) are discussed.

Hervé SEITZ

Empreinte génomique parentale et petits ARN non-codants

Directeur de thèse : Jérôme CAVAILLÉ

Soutenance publique le 25 octobre 2004, salle de conférences du bâtiment IBCG.

Résumé

Le phénomène d’empreinte génomique parentale se traduit par une expression différentielle des deux allèles de certains gènes, en fonction de leur origine parentale. Nos travaux ont abouti à la découverte de nombreux gènes de petits ARN non-codants de Mammifères (ARN C/D et microARN) soumis à l’empreinte génomique parentale, et à une première caractérisation de leur expression. Alors que la plupart des ARN C/D participent à la biogenèse des ARN ribosomiques et des petits ARN nucléaires, ceux que nous décrivons en semblent incapables ; les microARN, quant à eux, sont habituellement des répresseurs post-transcriptionnels de gènes spécifiques : parmi les microARN que nous décrivons, deux pourraient ainsi réprimer un rétrotransposon. Les fonctions éventuelles de ces gènes de petits ARN non-codants (dans le contrôle du développement, la répression de ce rétrotransposon, et le mécanisme de l’empreinte génomique parentale) sont discutées.

Mots-clés

ARN non-codants ; empreinte génomique parentale ; microARN ; ARN C/D

Discipline : Biologie moléculaire

Laboratoire de rattachement : Laboratoire de biologie moléculaire eucaryote du C.N.R.S. (U.M.R. 5099), I.E.F.G., 118 route de Narbonne, 31062 Toulouse CEDEX 4