

Tables des matières

Résumé	iii
Abstract.....	iv
Tables des matières	v
Liste des tableaux et des figures.....	viii
Liste des abréviations.....	ix
Remerciements	xi
Chapitre 1 : Introduction	1
1. La chromatine eucaryote : une compaction complexe et fonctionnelle.....	2
1.1 L'ADN co-existe avec des protéines au sein de structures complexes	2
1.2 La compaction de la chromatine régule des processus biologiques tels que la transcription.....	2
2. La transcription a lieu dans un contexte chromatinien dynamique	3
2.1 La transcription des gènes codant pour des protéines fonctionnelles est effectuée par l'ARN Polymérase II.....	3
2.2 La transcription est un processus hautement régulé.....	4
3. Les facteurs de transcription agissent de manière intégrée et combinée sur les régions régulatrices.....	7
3.1 L'occupation des enhanceurs est coopérative	7
3.2 L'activation d'un enhanceur a lieu par étapes.....	7
4. Le rôle des facteurs pionniers et tissus-spécifiques dans la sélection des éléments régulateurs qui contrôlent le programme d'expression	8
4.1 Les facteurs pionniers sélectionnent et initient l'activation des enhanceurs	8
4.2 Les facteurs de transcription tissus-spécifiques et signal-dépendants orchestrent les programmes d'expression géniques	9
5. Mediator & Cohésin.....	11
5.1 Le complexe co-activateur Mediator: centre d'interprétation des programmes d'expression génique.....	11
5.2 Le complexe multifonctionnel Cohésin : maintenance de la structure chromatinienne et régulation de l'expression génique.....	12
6. Le rôle de la communication génomique dans la régulation des programmes d'expression géniques.....	13
6.1 La répartition nucléaire des chromosomes n'est pas aléatoire mais fonctionnelle.....	13
6.2 Les enhanceurs forment des <i>clusters</i> dont la perturbation altère la réponse transcriptionnelle.....	14
6.3 Les enhanceurs distaux forment des contacts spécifiques avec les promoteurs de leurs gènes cibles	14
6.4 RNAPII forment des foyers de transcription actifs auxquels peuvent se connecter les interactions enhanceurs-promoteurs pour la co-régulation de plusieurs gènes	15
7. La dérégulation de la transcription dans l'apparition et la progression tumorale – le cancer du sein hormono-dépendant.....	17
7.1 Le Cancer du Sein hormono-dépendant est contrôlé par le facteur de transcription ER (Récepteur aux Estrogènes).....	18
7.2. Stratégies de thérapies anti-estrogéniques et modèles de résistance acquise à ces thérapies transcriptionnelles.....	20
8. Objectifs de recherche	22

Chapitre 2 : FOXA et les facteurs de transcription maîtres recrutent Mediator et Cohesin aux gènes actifs dans les cellules cancéreuses	25
Avant-propos	26
FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells.....	27
Résumé	28
Abstract.....	29
1. Introduction	30
2. Results	32
2.1 Mediator and Cohesin occupy regulatory regions of actively transcribed genes in cancer cells	32
2.2 Proliferation of cancer cells depends on Mediator and Cohesin	33
2.3 Mediator and Cohesin connect with the core transcriptional regulatory circuitry of cancer cells	34
2.4 FOXA and master transcription factors are essential to cancer cell maintenance.....	34
2.5 FOXA and master transcription factors recruit Mediator and Cohesin.....	35
3. Discussion.....	37
4. Figures.....	40
5. Methods.....	46
5.1 Cell Culture	46
5.2 Chromatin Immunoprecipitation	46
5.3 Bioinformatic analyses.....	47
5.4 Lentiviral shRNAs	47
5.5 Proliferation and colony formation assays	47
Acknowledgments	48
Author contributions statement.....	48
Additional information	48
6. Supplementary Information.....	49
6.1 Supplementary Tables.....	49
6.2 Supplementary Figures.....	64
6.3 Supplementary Data Files	68
6.4 Supplementary Experimental Procedures	68
Chapitre 3 : La modulation du programme transcriptionnel et la résistance endocrinienne dans le cancer du sein hormono-dépendant.....	76
Avant-propos	77
Résumé	78
1. Introduction	79
2. Résultats	81
2.1 Validation du modèle d'étude : les traitements anti-estrogéniques n'altèrent pas la prolifération des cellules résistantes.....	81
2.2 Les cellules sensibles et résistantes expriment les mêmes gènes	81
2.3 Un rôle plus secondaire pour ER dans la régulation du programme d'expression des cellules résistantes	82
2.4 Mediator, NIPBL et Cohesin occupent de nouvelles régions régulatrices dans les cellules résistantes	83

2.5 Le rôle d'AP-1 dans la néo-régulation du programme transcriptionnel des cellules devenues résistantes	84
2.6 La sur-expression de FOSL1 modifie la réponse proliférative des cellules sensibles traitées aux anti-estrogènes	85
3. Discussion.....	86
4. Figures.....	89
5. Matériel et Méthodes	94
5.1 Culture cellulaire et traitements hormonaux	94
5.2 Immunoprécipitation de la chromatine	94
5.3 Analyses bioinformatiques.....	95
5.4 Lentivirus ShRNA.....	95
5.5 Essais de prolifération	95
5.6 Extraction d'ARN & RTqPCR.....	96
5.7 Lentivirus ORF.....	96
Chapitre 4 : Conclusion/Discussion	97
1. Les thérapies transcriptionnelles	99
2. Mettre à profit les <i>super-enhancers</i> dans la lutte contre le cancer.....	100
3. Les analyses de contacts tri-dimensionnels à l'échelle génomique offrent une vue d'ensemble des mécanismes qui régulent la transcription normale et aberrante.....	101
4. L'édition du génome: un outil puissant pour combattre le cancer	103
Références	104

Liste des tableaux et des figures

Figure 1.1: La transcription est un processus régulé	6
Figure 1.2: Le rôle des facteurs de transcription dans la sélection et l'activation des enhanceurs	10
Figure 1.3: Le rôle de la communication génomique dans la régulation des programmes d'expression	16
Figure 1.4 : Le récepteur aux estrogènes (ER) contrôle le programme transcriptionnel des cellules du cancer du sein hormono-dépendantes	22
Figure 1.5: Mediator, NIPBL et Cohesin régulent le programme transcriptionnel des cellules normales	24
Figure 2.1: Mediator, Cohesin and NIPBL co-localize at regulatory regions of actively transcribed genes in cancer cells	40
Figure 2.2: Mediator, Cohesin and NIPBL are essential for maintenance of cancer cells properties	42
Figure 2.3: The core circuitry of cancer cells connect with Mediator and Cohesin	43
Figure 2.4: Depletion of the core transcription factors phenocopies the loss of Mediator and Cohesin	44
Figure 2.5: FOXA and master transcription factors are essential to recruit Mediator and Cohesin ...	45
Supplementary Table S1: KEGG pathways enriched with genes occupied by MED1, SMC1A and NIPBL in MCF7, HEPG2 and A549 cells	49
Supplementary Table S2: List of ChIP-Seq datasets used in the study (MCF7)	53
Supplementary Table S2: List of ChIP-Seq datasets used in the study (HEPG2)	56
Supplementary Table S2: List of ChIP-Seq datasets used in the study (A549)	61
Supplementary Table S3: List of shRNA plasmids and sequences used in the study	63
Supplementary Figure S1: shRNA targeting of MED1, SMC1A and NIPBL leads to decreased protein levels	64
Supplementary Figure S2: Validation of the core transcriptional regulatory circuitry in cancer cells	65
Supplementary Figure S3: shRNA targeting of the identified transcription factors leads to decreased protein levels	66
Supplementary Figure S4: Validation of the recruitment of Cohesin and Mediator by FOXA and master transcription factors	67
Figure 3.1: Les cellules résistantes LCC9 se sont affranchies de leur dépendance vis-à-vis du récepteur ER	89
Figure 3.2 : Mediator, NIPBL et Cohesin occupent de nouvelles régions régulatrices dans les cellules résistantes	91
Figure 3.3 : AP-1 est impliqué dans la résistantes endocrinienne des cellules du cancer du sein hormono-dépendant	92
Figure 4 : Modèle de régulation des programmes d'expression génique cancéreux et d'adaptation en réponse à des pressions thérapeutiques d'ordre transcriptionnel	99

Liste des abréviations

3C	Capture de la <u>C</u> onformation des <u>C</u> hromosomes
ADN	<u>A</u> cide <u>D</u> ésoxyribo- <u>N</u> ucléique
ADNc	<u>A</u> DN <u>c</u> odant
AF	<u>A</u> ctivation <u>F</u> unction
AI	<u>I</u> nhibiteur d' <u>A</u> romatases
AP-1	<u>A</u> ctivator <u>P</u> rotein <u>1</u>
APO	<u>A</u> polipoprotéine
AR	<u>R</u> écepteur aux <u>A</u> ndrogènes
ARN	<u>A</u> cide <u>R</u> ibo- <u>N</u> ucléique
ARNPII/PolIII	<u>ARN</u> <u>P</u> olymérase <u>II</u>
ATF3	<u>A</u> ctivating <u>T</u> ranscription <u>F</u> actor <u>3</u>
CCD	<u>D</u> omaine de <u>C</u> ontact <u>C</u> hromatinien
CDK	<u>K</u> inase <u>C</u> ycline- <u>D</u> épendante
CdLS	syndrome <u>C</u> ornelia <u>D</u> e <u>L</u> ange
CEBPB	<u>CCAAT/Enhancer Binding Protein Beta</u>
ChIA-PET	<u>C</u> hromatin <u>I</u> nteractions <u>A</u> nalysis by <u>P</u> aired- <u>E</u> nd <u>S</u> equencing
ChIP-seq	<u>I</u> mmuno- <u>P</u> réciptation de <u>C</u> hromatine suivie de séquençage
CRISPR	<u>C</u> lustered <u>R</u> egularly <u>I</u> nterspaced <u>S</u> hort <u>P</u> alindromic <u>R</u> epeats
CRPCa	<u>C</u> ancer de la <u>P</u> rostate <u>R</u> ésistant à la <u>C</u> astration
CTCF	<u>CCCTC Binding Factor</u>
CTD	<u>D</u> omaine <u>T</u> erminal <u>C</u> arboxyl
Ctrl	Contrôle
E2	<u>E</u> strogènes
EGF	<u>E</u> pidermal <u>G</u> rowth <u>F</u> actor
ER	<u>R</u> écepteur aux <u>E</u> strogènes
ERBB2 (HER2)	<u>Erb-B2 Receptor Tyrosine</u>
ERE	<u>É</u> lements de <u>R</u> éponse aux <u>E</u> strogènes
eRNA	<u>ARN</u> associé à la transcription des <u>e</u> nhancers
ESCs	<u>C</u> ellules <u>S</u> ouches <u>E</u> mbryonnaire
ESR1	<u>E</u> strogen <u>S</u> ignaling <u>R</u> eceptor
FBS	<u>S</u> érum <u>B</u> ovin <u>F</u> oetal
FGFR	<u>F</u> ibroblast <u>G</u> rowth <u>F</u> actor <u>R</u> eceptor
FISH	<u>H</u> ybridation <u>F</u> luorescente <u>I</u> n- <u>S</u> itu
FOS	<u>FBJ Murine Osteosarcoma Viral Oncogene Homolog</u>
FOSL	<u>FOS Like Antigen</u>
FOXA	<u>ForkHead Box A</u>
FUL	<u>F</u> ulvestrant
GAPDH	<u>G</u> lyceraldehyde-3 <u>P</u> hosphate <u>D</u> ehydrogenase
GREB1	<u>G</u> rowth <u>R</u> egulator by <u>E</u> strogen in <u>B</u> reast <u>1</u>
GWAS	<u>G</u> enome- <u>W</u> ide <u>A</u> ssociation <u>S</u> tudies
H3K27Ac	<u>A</u> céthylation de la lysine (<u>K</u>) <u>27</u> de l'histone <u>H3</u>
H3K4Me3	<u>T</u> ri- <u>M</u> éthylation de la lysine (<u>K</u>) <u>4</u> de l'histone <u>H3</u>
HAT	<u>H</u> istone <u>A</u> cétyle- <u>T</u> ransférase
HDAC	<u>H</u> istone <u>D</u> é- <u>A</u> cétyle <u>T</u> ransférase
HDM	<u>H</u> istone <u>D</u> é- <u>M</u> éthylase
HIF	<u>F</u> acteurs <u>I</u> nduits par l' <u>H</u> ypoxie
HMT	<u>H</u> istone <u>M</u> éthyle <u>T</u> ransférase

HNF	<i>Hepatocyte Nuclear Factor</i>
IgH	<i>Immunoglobulin Heavy locus</i>
JUN	<i>JUN</i> proto-oncogene
Kb	<i>Kilo</i> (paire) de bases
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
KLF	<i>Kruppel-Like Factor</i>
LRH1	<i>Liver Receptor Homolog 1</i>
MAPK	<i>Mitogen-Activated Protein Kinase</i>
MAX	<i>MYC Associated Factor X</i>
MET	<i>MET</i> proto-oncogene, Recepteur Tyrosine Kinase
Mg	<i>Méga</i> (paire) de bases
MYC	<i>v-myc Avian Myelocystomatosis Viral Oncogene Homolog</i>
MYOD1	<i>Myogenic Differentiation 1</i>
NFKB	<i>Nuclear Factor Kappa-B</i>
NIPBL	<i>Nipped-B Like Protein</i>
OCT4 (POU5F1)	<i>POU Class 5 Homeobox 1</i>
ORF	<i>Open Reading Frame</i>
Pb	<i>Paire</i> de bases
PCA	<i>Analyse en Composantes Principales</i>
PDGFB	<i>Platelet Derived Growth Factor Subunit B</i>
PI3KR	<i>Phosphatidylinositol-4,5-biphosphate 3-Kinase Receptor</i>
PIC	<i>Complexe de Pré-Initiation</i>
PPAR	<i>Peroxisome Proliferator Activated Receptor</i>
pTEFb	<i>positive Transcription Elongation Factor beta</i>
PTEN	<i>Phosphatase And Tensin Homolog</i>
qPCR	<i>Réaction de Polymérisation en Chaîne quantitative</i>
RAD21	<i>RAD21 Cohesin Complex Component</i>
RT	<i>Transcription Réverse</i>
SCLC	<i>Small Cells Lung Cancer</i>
SERD	<i>Destructeur Sélectif du Récepteur aux Estrogènes</i>
SERM	<i>Modulateur Sélectif du Récepteur aux Estrogènes</i>
ShRNA	<i>Small hairpin RNA</i>
SMC	<i>Complexe de Maintenance de la Structure des chromosomes</i>
SNP	<i>Polymorphisme Nucléotidique Simple</i>
SOX2	<i>SRY-Box2</i>
SP-1	<i>Specificity Protein 1</i>
STAG2 (SA2)	<i>Stromal Antigen 2</i>
STAT	<i>Signal Transducer and Activator of Transcription</i>
SWI/SNF	<i>SWItch/Sucrose Non-Fermentable</i>
TAD	<i>Domaine d'Association Topologique</i>
TAL1	<i>T-cell Acute Lymphocytic Leukemia 1</i>
TAM	<i>Tamoxifen</i>
TF	<i>Facteur de transcription</i>
TFE1	<i>Trefoil Factor 1</i>
TGF	<i>Transforming Growth Factor</i>
TGFBR	<i>TGF-beta Receptor</i>
TNBC	<i>Triple Negative Breast Cancer</i>
TSS	<i>Transcription Start Site</i>

Remerciements

Je remercie mes ami(e)s et ma famille pour leur patience, leur soutien et leurs conseils tout au long de mon parcours académique à l'Université Laval.

Je remercie ma directrice de programme, Dre Josée Lavoie, ainsi que Chantal Joubert pour leur support depuis ma décision d'effectuer un retour à la maîtrise, jusqu'au dépôt initial de mon mémoire.

Je remercie également l'Université Laval pour les nombreux services de qualité d'aide aux étudiants, notamment Sylvie Dodin, responsable de l'initiative REAGIR.

Je remercie mes collègues pour nos conversations scientifiques et pour l'environnement au sein du laboratoire, particulièrement Maxime pour sa bonne humeur et son enseignement, et Imène pour sa bonté et sa persévérance. Ils furent des modèles qui ont participé à ma formation.

Je remercie les Dr Fabien Lamaze et Eric Fournier pour leur support bioinformatique et leur disponibilité malgré un environnement pas toujours favorable à la formation.

Je remercie le Dr Robert Clarke pour toutes les expériences réalisées dans la lignée cellulaire LCC9. Je remercie également le Dr Mathieu Laplante pour la mise à disposition de sa banque de vecteurs lentiviraux de sur-expression.

Je remercie mon directeur de recherche pour l'apprentissage d'une façon de penser et de raisonner toujours plus large, pour la culture scientifique qu'il m'a permis de recevoir, ainsi que pour l'opportunité de participer à la rédaction et à la soumission d'un article. Je le remercie également pour sa flexibilité et son ouverture d'esprit.

Je remercie finalement et non moins le Comité Étudiants de m'avoir donné la chance de m'investir dans les activités académiques du Centre de Recherche, ainsi que pour leur confiance tout au long de mon mandat de co-responsable.

Chapitre 1 : Introduction

1. La chromatine eucaryote : une compaction complexe et fonctionnelle

1.1 L'ADN co-existe avec des protéines au sein de structures complexes

La découverte de l'ADN comme molécule primordiale de la génétique, celle qui comporte toute l'information nécessaire au fonctionnement d'un organisme, a immédiatement focalisé l'attention sur sa structure ¹. L'ADN comporte deux chaînes polynucléotidiques, constituées d'une suite continue de bases azotées (ATCG), qui s'enroulent l'une autour de l'autre pour former une double hélice complémentaire anti-parallèle, et dont certains arrangements nucléotidiques codent pour des gènes, permettant d'exercer toutes les fonctions biologiques d'un organisme une fois correctement exprimés ².

Dans la cellule, l'ADN est associé à des protéines ². Chaque molécule d'ADN, accompagnée de ses protéines, correspond à un chromosome. L'association globale ADN - protéines est appelé chromatine ³. La majorité des protéines associées à l'ADN sont des histones ⁴. Deux copies de chaque histone H2A, H2B, H3 et H4 servent à former le cœur protéique autour duquel l'ADN nucléosomal s'enroule. L'histone H1 est quant à elle inter-nucléosomique. Cette association ADN - histones permet la compaction de l'ADN, et participe à plusieurs fonctions importantes. Tout d'abord, la forme compacte de l'ADN lui permet d'être contenu dans le noyau de la cellule eucaryote. En effet, l'ADN d'une cellule humaine fait 2 mètres de long alors que le diamètre du noyau devant le contenir ne fait que 10 à 15 μm de diamètre ⁵. Ensuite, cet empaquetage le protège de certaines altérations. En effet, l'ADN sous sa forme chromosomique est extrêmement stable ⁶. De plus, seul l'ADN sous forme chromosomique peut être transmis efficacement aux cellules filles lorsque la cellule se divise ^{7,8}. Enfin, le chromosome confère une organisation générale particulière à chaque molécule d'ADN, gouvernant l'expression des gènes ⁹. D'autres protéines sont associées, notamment de nombreuses protéines de liaison à l'ADN, permettant certains processus biologiques d'avoir lieu, entre autres la transcription, qui permet l'expression des gènes codés par l'ADN.

1.2 La compaction de la chromatine régule des processus biologiques tels que la transcription

La première étape de compaction de l'ADN, le nucléosome, l'associe aux protéines histones ⁴. Mais cette compaction limite son accessibilité, empêchant notamment la fixation de protéines impliquées dans la transcription de l'ADN ¹⁰. Les cellules eucaryotes tirent profit de ces propriétés inhibitrices de la chromatine pour réguler l'expression de leurs gènes ¹¹. Le remaniement local de

certaines nucléosomes permet à des régions spécifiques de l'ADN d'interagir avec des protéines. Il s'agit d'un processus dynamique localisé ¹².

Bien qu'il existe une corrélation entre le nombre de gènes et la complexité d'un organisme, il semblerait que ce soit plutôt la densité génique qui serait reliée à cette complexité. Par exemple, l'espèce humaine contient environ 20 000 gènes, autant que le vers *c. elegans*. Mais sa densité génique est 30 fois inférieure ². En fait, l'ADN humain contient presque 80% de régions intergéniques, comprenant des séquences régulatrices, dédiées au contrôle et à la régulation de la transcription, dont la nature et le fonctionnement possèdent une grande complexité ¹³.

2. La transcription a lieu dans un contexte chromatinien dynamique

2.1 La transcription des gènes codant pour des protéines fonctionnelles est effectuée par l'ARN Polymérase II

La transcription chez les eucaryotes est prise en charge par des ARN Polymérases, et nécessitent l'action de facteurs généraux de la transcription. L'ARN Polymérase II (ARNPII) est responsable de la transcription de la majorité des gènes, en fait essentiellement tous les gènes codant pour des protéines ². Pour transcrire un gène, l'ARNPII effectue une série de trois étapes : initiation, élongation et terminaison ¹⁴.

L'initiation de la transcription commence avec le recrutement de l'ARNPII et des facteurs d'initiation sur une courte séquence d'ADN en amont d'un gène qui entoure le TSS (*Transcription Start Site*), le promoteur ¹⁵. L'appariement des bases, qui constitue la double hélice nucléotidique de l'ADN, est rompu. Le brin « sens », ou codant, sert alors de matrice pour l'enzyme ARNPII qui assemble une chaîne d'ARN nouvellement synthétisée. C'est la molécule d'ARN qui servira ensuite de matrice pour le processus de traduction, qui permettra à la cellule de produire la protéine fonctionnelle correspondante ². Une fois que l'ARNPII a synthétisé un court fragment d'ARN (approximativement 10 bases), elle passe à l'étape d'élongation. Elle dissocie la chaîne d'ARN nouvellement synthétisée au fur et à mesure de sa progression, et corrige les potentielles erreurs créées. La terminaison constitue en la libération de la molécule d'ARN produite ¹⁴.

L'initiation représente l'étape la plus critique de la transcription et c'est pourquoi elle est soumise à une régulation complexe et très dynamique. *In vitro*, les facteurs généraux de la transcription représentent tout ce qui est nécessaire avec l'ARNPII pour initier la transcription sur un

ADN nu. *In vivo* cependant, les facteurs généraux de la transcription ne sont pas suffisants à eux seuls pour déclencher une expression significative ¹⁶.

2.2 La transcription est un processus hautement régulé

La transcription des cellules eucaryotes est, comme chez les bactéries, régulée par des activateurs et des répresseurs : des protéines liant l'ADN, qui facilitent ou empêchent l'initiation de la transcription de gènes spécifiques en réponse à des signaux appropriés ¹¹. Cependant, les actions de ces protéines régulatrices sont complexifiées par les caractéristiques additionnelles des cellules et des gènes eucaryotes.

2.2.1 Les nucléosomes sont modifiés pour réguler l'accessibilité de l'ADN à la machinerie transcriptionnelle

Chez les eucaryotes, l'ADN est enveloppé dans les nucléosomes. Ceux-ci constituent un frein à la transcription, qui n'est pas rencontré chez les bactéries, mais leurs modifications offrent de nouvelles opportunités de régulation ¹⁰. En effet, les histones au sein de ces nucléosomes sont sujettes à diverses modifications qui affectent la disposition et la compacité des nucléosomes, et la flexibilité de la structure de la chromatine (**Figure 1.1A**) ¹⁷.

Les lysines situées dans les queues des histones sont fréquemment modifiées par des groupes acétyles ou méthyles. De façon similaire, les sérines font l'objet de modifications par les phosphates. La nature ainsi que la position de ces modifications ont un rôle essentiel dans leur interprétation ¹⁸. Un changement évident est réalisé par l'acétylation et la phosphorylation, qui réduisent la charge positive totale des queues des histones. La perte de charge positive réduit l'affinité des queues pour les charges négatives de la molécule d'ADN, ayant pour conséquence une ouverture et donc une plus grande accessibilité de la chromatine à ces régions ¹⁹. Les queues modifiées des histones peuvent aussi servir à recruter des protéines qui possèdent des domaines particuliers de reconnaissance spécifiques aux différentes modifications (**Figure 1.1A**) ²⁰. Ces protéines peuvent être de nouvelles enzymes de modifications des queues d'histones qui vont pouvoir modifier les nucléosomes adjacents, ou encore des complexes de remodelage des nucléosomes.

Les complexes de remodelage des nucléosomes facilitent les mouvements de ces derniers le long de l'ADN et contribuent ainsi à réguler la structure de la chromatine aux endroits du génome où ils sont recrutés (**Figure 1.1A**) ²¹. Ils peuvent être composés de deux à plus de dix sous-unités, et possèdent une activité catalytique ²². La combinaison entre les modifications des queues d'histones

et le remodelage des nucléosomes résulte de façon localisée en un profond changement dans l'accessibilité de l'ADN.

2.2.2 Le rôle des séquences régulatrices dans la régulation transcriptionnelle

Une autre différence entre les eucaryotes et les procaryotes réside dans le nombre de protéines régulatrices qui contrôlent un gène typique. Cela est reflété par le nombre et la disposition des sites de liaison des régulateurs associés à un gène, et démontre une plus vaste intégration de signaux chez ces organismes ². Ces séquences régulatrices peuvent être localisées à des dizaines de milliers voire des centaines de milliers de nucléotides à partir du promoteur, aussi bien en amont qu'en aval (**Figure 1.1B**) ^{23,24}.

Les enhancers sont de petits segments d'ADN, qui mesurent typiquement quelques centaines de paires de bases, et qui servent de plateformes opérationnelles pour recruter des facteurs de transcription, à travers de courtes séquences nucléotidiques spécifiques appelées motifs, qui présentent un fort degré de dégénérescence de reconnaissance ²⁵⁻²⁸. Des enhancers alternatifs lient différents groupes de régulateurs et contrôlent l'expression d'un même gène à différents moments et à différents endroits en réponse à différents signaux ²⁹.

D'autres séquences régulatrices, appelées isolateurs ou éléments frontières, sont retrouvées entre les enhancers et les promoteurs (**Figure 1.1B**). Ces isolateurs bloquent l'activation du promoteur induite par les activateurs liés aux enhancers, et garantissent que les activateurs ne fonctionnent pas aveuglément ³⁰⁻³². Un isolateur agit comme une barrière vis-à-vis de l'influence d'éléments en *cis* voisins actifs, et peut ainsi éviter certaines interactions tri-dimensionnelles entre des enhancers et des promoteurs pour réguler négativement une activation transcriptionnelle non désirée par la cellule. La protéine CTCF est retrouvée sur la très grande majorité de ces éléments et est responsable de leur activité isolatrice ³⁰⁻³². A l'échelle tri-dimensionnelle, CTCF permet ainsi la compartimentalisation de l'activation génique, et donc de l'activité transcriptionnelle, par sa capacité à réguler de façon spatiale les boucles topologiques régulatrices au sein de la chromatine.

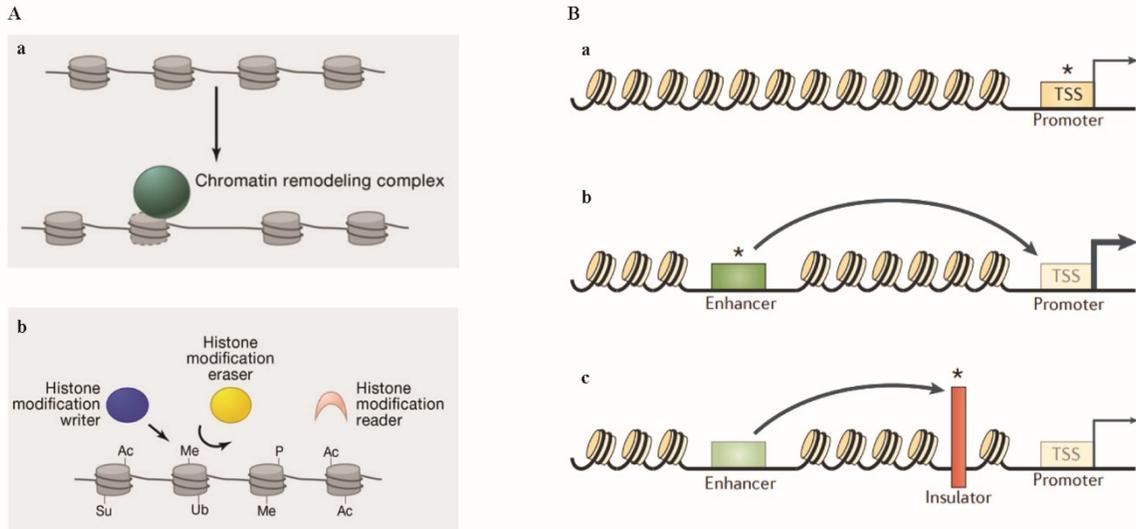


Figure 1.1: La transcription est un processus régulé.

(A) La structure des nucléosomes est régulée de façon dynamique par des protéines. (a) La structure de la chromatine est régulée par des complexes de remodelage ATP-dépendants. Ceux-ci peuvent mobiliser un nucléosome et créer des régions ouvertes dans la chromatine, permettant aux régulateurs de la transcription et à la machinerie transcriptionnelle un meilleur accès aux séquences contenues dans l'ADN. **(b)** L'activité transcriptionnelle est influencée par des protéines qui reconnaissent et modifient des histones au sein des nucléosomes. Certaines protéines ont la capacité d'effectuer des modifications covalentes sur des résidus des queues d'histone (*writers*). À l'inverse, d'autres protéines peuvent au contraire les « effacer » (*erasers*). Certaines protéines peuvent également reconnaître des modifications spécifiques grâce à des domaines de reconnaissance particuliers (*reader*). Ces fonctions ne sont pas mutuellement exclusives au sein d'une même protéine de régulation de la structure des nucléosomes. Il existe plusieurs modifications enzymatiques des résidus des queues d'histones : l'acétylation (Ac), la méthylation (Me), la phosphorylation (P), mais aussi la sumoylation (Su) et l'ubiquitination (Ub). Ces modifications enzymatiques sont effectuées par des protéines de différentes familles selon la nature des résidus qu'elles sont susceptibles de modifier (Tiré de Lee et al. ³³).

(B) Les séquences régulatrices : rôle des promoteurs, enhanceurs et isolateurs. (a) Les promoteurs résident habituellement dans les 10 à 100 pb qui entourent le TSS (*Transcription Start Site*) et recrutent la machinerie transcriptionnelle basale composée de ARNPII et des facteurs généraux de la transcription. Un même gène, exprimé conjointement dans plusieurs types cellulaires, utilise souvent le même promoteur. **(b)** Les enhanceurs sont des régions d'ADN qui recrutent les facteurs de transcription et les co-facteurs. Ils peuvent activer la transcription au niveau des promoteurs. Contrairement aux séquences promotrices, les enhanceurs peuvent être localisés jusqu'à 1 Mb du TSS de leur gène cible. **(c)** Les insulateurs limitent l'activation des promoteurs par les enhanceurs, en prévenant leur communication génomique. Chacune de ces séquences peut arborer des mutations (*) qui affectent l'expression génique, résultant en l'apparition et la progression de maladies dont le cancer (Tiré de Sur & Taipale ³⁰).

3. Les facteurs de transcription agissent de manière intégrée et combinée sur les régions régulatrices

3.1 L'occupation des enhanceurs est coopérative

Le génome humain code pour 2 000 à 3 000 facteurs de transcription, dont quelques 200 à 400 sont conjointement exprimés au sein d'une cellule, souvent à de très fortes concentrations cellulaires³⁴. Leur faible affinité de séquence pour les motifs présents au sein des enhanceurs suggère que leur mécanisme fonctionnel est bien plus complexe qu'une reconnaissance de séquence unique. Les enhanceurs contiennent en outre des *clusters* de différents sites de reconnaissance à différents facteurs de transcription, une propriété qui a d'ailleurs servi à les définir (**Figure 1.2**)³⁵.

L'observation de la complexité d'occupation génomique de ces facteurs a amené au modèle de réseau transcriptionnel interconnecté³⁶. En effet, les facteurs de transcription n'agissent pas seuls, leurs fonctions devraient être considérées d'une manière intégrée (**Figure 1.2**). L'occupation combinatoire des facteurs de transcription peut entraîner différents types de réponses transcriptionnelles, dépendamment de la façon dont ces facteurs interagissent entre eux³⁷, et est régulée de façon spatio-temporelle³⁸⁻⁴¹. Dans certains cas simples, l'activité d'un enhanceur est proportionnelle à la concentration d'un facteur individuel⁴². L'occupation coopérative amène une relation non-linéaire entre la concentration d'un facteur et son degré d'occupation génomique sur les éléments régulateurs. Cette coopération est très souvent associée à des communications protéiques⁴³⁻⁴⁵ (**Figure 1.2A**). Ces interactions peuvent être directes, ou bien indirectes, via des composants conservés de la machinerie transcriptionnelle par exemple, ou encore via la restructuration locale de l'ADN (**Figure 1.2A**).

3.2 L'activation d'un enhanceur a lieu par étapes

Le recrutement et la fixation de facteurs de transcription sur la chromatine est un processus collaborateur et hiérarchique¹⁶ (**Figure 1.2**). L'activation des enhanceurs commence avec le recrutement de certains facteurs de transcription et le remodelage local des nucléosomes²⁹. Le profil des modifications épigénétiques au sein des enhanceurs, ainsi que leur occupation génomique, a permis de les classifier et de prédire leur état d'activation dans différents types cellulaires ou à un instant donné^{13,46}. Un enhanceur inactif est inaccessible pour les facteurs de transcription car enveloppé dans une structure chromatiniennne très compacte impliquant des histones non modifiées. Certains enhanceurs sont marqués quand ils présentent des régions pauvres en nucléosomes et sont occupés par

certains facteurs de transcription, mais nécessitent un signal supplémentaire pour exercer leur activité régulatrice, le recrutement de facteurs additionnels, et celui de co-facteurs, résultant en l'activation de la transcription. Les enhancers présentent des profils épigénétiques spécifiques de leur état d'activation. La grande majorité des enhancers arbore des mono et di méthylation de la lysine K4 de l'histone H3 (H3K4me1 et H3K4me2). L'acétylation de la lysine K27 de l'histone H3 (H3K27Ac) est caractéristique d'un enhancer actif, de même que la présence de l'ARNPII ²⁹ (**Figure 1.2B**).

La relation entre l'occupation génomique et l'expression des gènes proches est relativement faible. Bien que certains événements d'occupation ne produisent pas une réponse transcriptionnelle immédiate, ils peuvent avoir un rôle fonctionnel dans le remodelage de la structure chromatinienne, le positionnement des nucléosomes, et le marquage de certains enhancers, qui peuvent influencer l'expression latente de ces gènes ³⁵.

4. Le rôle des facteurs pionniers et tissus-spécifiques dans la sélection des éléments régulateurs qui contrôlent le programme d'expression

4.1 Les facteurs pionniers sélectionnent et initient l'activation des enhancers

Les événements successifs depuis le marquage des enhancers jusqu'à leur activation peuvent avoir d'importants rôles fonctionnels sur la précision de l'exécution des programmes d'expression géniques. La nature de l'état d'un enhancer est initialement contrôlée par l'action de facteurs de transcription pionniers (**Figure 1.2B**) ^{29,47}. Ces facteurs particuliers peuvent fonctionner comme déterminants cruciaux pour la spécificité cellulaire car ils contrôlent la sélection des programmes transcriptionnels et des régions régulatrices impliquées dans l'exécution de ces programmes. Ils ont la capacité de lier l'ADN sous sa forme inaccessible aux autres facteurs, c'est-à-dire de reconnaître des motifs au sein des nucléosomes, et de recruter des complexes de remodelage de la chromatine afin de découvrir l'élément régulateur ⁴⁸⁻⁵⁰. Par exemple, le facteur pionnier FOXA possède un domaine de liaison à l'ADN en hélice très particulier, similaire à l'histone H1, qui lui confère l'habilité de lier la chromatine sous sa forme compactée. Les facteurs pionniers Oct4, Sox2 et Klf4 ont également une nette préférence de liaison aux régions couvertes de nucléosomes, et leur expression ectopique entraîne la reprogrammation des cellules somatiques en cellules pluripotentes ⁵¹. Ces facteurs apportent donc la spécificité de séquence qui offre aux complexes de remodelage généraux une spécificité d'activation d'enhancers particuliers et participent activement à la reprogrammation cellulaire.

En plus du repositionnement des nucléosomes, la protection vis-à-vis de la méthylation de l'ADN au sein des enhancers marqués semble jouer un rôle dans l'occupation des enhancers par les facteurs de transcription. En effet, la méthylation des cytosines de l'ADN influence généralement négativement le recrutement des facteurs transcriptionnels ⁵².

L'occupation de la chromatine par les facteurs pionniers n'implique pas souvent une réponse immédiate. En effet, son rôle principal passe par l'initiation d'une série de changements qui permettent à un enhancer de recruter subséquentiellement les facteurs de transcription complémentaires (**Figure 1.2B**) ¹⁶.

4.2 Les facteurs de transcription tissus-spécifiques et signal-dépendants orchestrent les programmes d'expression géniques

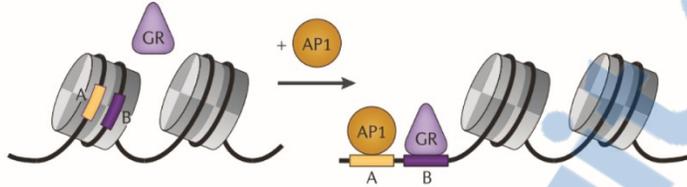
Les facteurs de transcription sont responsables de l'orchestration spatiale et temporelle des programmes d'expression géniques ³⁵. Les facteurs pionniers marquent les enhancers pour lesquels les autres facteurs de transcription, par exemple signaux-dépendants, auront plus d'accessibilité pour la chromatine et permettront d'établir de plus larges régions pauvres en nucléosomes en réponse à un signal particulier (**Figure 1.2B**) ^{29,53}. La sélection et la fonction des éléments de régulation est donc contrôlée par la combinaison d'action des facteurs pionniers et signaux-spécifiques qui interagissent entre eux ^{54,55} (**Figure 1.2**). Des études classiques de gain/perte de fonction prouvent que ces facteurs orchestrent l'identité des tissus dans lesquels ils sont sélectivement exprimés. Par exemple, la surexpression de MYOD1 dans des fibroblastes est capable de les convertir en cellules musculaires. La perte de Oct4, à l'inverse, conduit à la disparition de la population de cellules pluripotentes dans l'embryon de mammifère. Les gènes régulés par une forte densité de facteurs de transcription tissus-spécifiques sont préférentiellement associés à l'apparition de maladies et peuvent ainsi être identifiés à travers les types cellulaires par la cartographie épigénétique des enhancers actifs ^{56,57}.

La transcription des gènes eucaryotes requiert l'orchestration précise d'interactions complexes entre une myriade de protéines et de séquences régulatrices ⁵⁸. La régulation de la transcription est accomplie en grande partie par les enhancers, qui contiennent de multiples sites de reconnaissance pour une grande variété de facteurs de transcription. Les enhancers peuvent activer la transcription indépendamment de leur localisation, distance ou orientation par rapport au promoteur d'un gène cible ²⁵. Cependant, ces programmes ne sont interprétés qu'à travers le recrutement orchestré de co-activateurs par ces facteurs de transcription ⁵⁹.

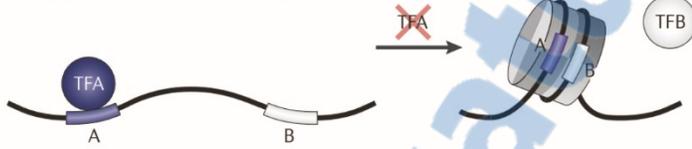
A
a Co-binding to common cofactors or common complexes
 (transcriptional synergy)



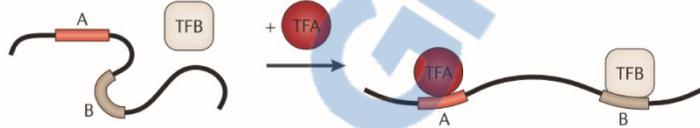
b Activating chromatin remodelling
 (e.g. a pioneer TF)



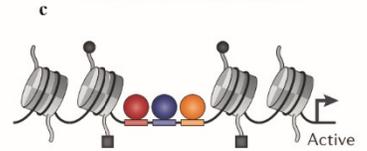
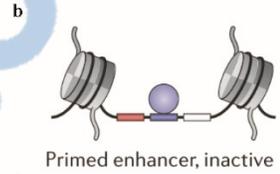
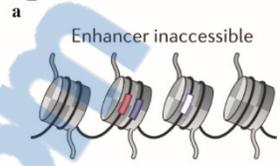
c Blocking nucleosome repositioning
 (e.g. 'assisted loading' or 'collaborative competition')



d Architectural role
 (e.g. DNA bending)



B



Rapport-Grain.com

Figure 1.2: Le rôle des facteurs de transcription dans la sélection et l'activation des enhanceurs.

(A) Les facteurs de transcription sont recrutés de manière coopérative sur les enhanceurs. (a)

Deux ou plusieurs facteurs de transcription peuvent co-occuper le même élément régulateur et recruter un co-activateur commun (par exemple, p300) ou alors différents composants d'un complexe protéique de régulation (comme Mediator), ce qui peut résulter en une augmentation d'affinité de chaque facteur pour l'enhancer, ou stabiliser leur recrutement sur la chromatine. **(b)** Certains facteurs de transcription peuvent agir de façon coopérative via l'activation du remodelage de la chromatine. Les motifs de reconnaissance de TFA et TFB sont localisés dans une région riche en nucléosomes. Le recrutement de TFA sur son motif permet le repositionnement des nucléosomes, résultant en l'exposition de la séquence de recrutement de TFB et du recrutement de ce dernier sur la chromatine. TFA joue un rôle de facteur pionnier et assiste le recrutement de TFB, sans nécessairement besoin d'interaction directe entre les deux facteurs. **(c)** En occupant une région, un facteur de transcription peut aussi limiter la compaction nucléosomale, et peut ainsi faciliter le recrutement d'autres facteurs de transcription sur des séquences adjacentes. Ce modèle pourrait servir à marquer certains enhanceurs de manière passive, en « laissant » volontairement un facteur occuper un élément régulateur qui risque d'être à nouveau sollicité. **(d)** Certains facteurs de transcription peuvent induire des « plis » locaux de l'ADN, résultant en une augmentation d'affinité pour le recrutement d'autres facteurs de transcription (Tiré de Spitz et al ³⁵).

(B) L'activation d'un enhancer a lieu par étapes. (a) Certains enhanceurs sont, par exemple à des stades précoces du développement, associés à une densité nucléosomale importante qui empêche l'accès des séquences régulatrices pour le recrutement de facteurs de transcription. **(b)** L'expression d'un facteur pionnier, par exemple à un stade transitoire du développement ou lors de la reprogrammation cellulaire, peut initier le remodelage de la chromatine, le repositionnement des nucléosomes, et peut ainsi marquer un enhancer pour une activation future. **(c)** Dépendamment du contexte cellulaire, le recrutement de facteurs de transcription particuliers additionnels sur un enhancer marqué peut résulter en son activation, l'association de caractéristiques épigénétiques activatrices, et le recrutement de la machinerie transcriptionnelle sur le promoteur du gène associé (Tiré de Spitz et al. ³⁵).

5. Mediator & Cohesin

5.1 Le complexe co-activateur Mediator: centre d'interprétation des programmes d'expression génique

Les facteurs généraux de la transcription sont directement connectés à l'enzyme ARNPII et impliqués dans la formation et la fonction d'un complexe de pré-initiation (PIC) ¹⁴. De nombreux co-activateurs, recrutés par la combinaison de facteurs de transcription associés aux régions régulatrices, facilitent cette pénétrance initiale de la barrière chromatinienne, pour rendre la chromatine accessible à la machinerie transcriptionnelle, et stimulent l'initiation de la transcription ⁵.

Pour être correctement interprété, le signal de régulation doit être proprement relayé à la machinerie transcriptionnelle afin de transcrire les gènes codants. C'est le rôle du complexe Mediator, dont le modèle général est celui selon lequel les signaux de régulation seraient traduits à travers sa

capacité à connecter les régions régulatrices qui reçoivent ces signaux avec l'ARNPII afin de permettre la formation stable du PIC ^{60,61}.

Mediator est un complexe multiprotéique qui présente une composition en sous-unités variable selon le contexte cellulaire ⁶². Différents facteurs de transcription, qui dépendent de l'induction de différentes voies de signalisation, peuvent interagir avec des sous-unités spécifiques de Mediator pour orchestrer leurs réponses biologiques ⁶³⁻⁶⁵. Mediator est dynamique non seulement au niveau de sa composition mais aussi dans sa structure. En effet, celle-ci est très flexible, et l'interaction avec d'autres protéines peut complètement la modifier ^{62,66}. Mediator agirait en tant que senseur et assimiler une multitude de signaux, pour délivrer le message correspondant, traduit et calibré, à la machinerie transcriptionnelle, et fonctionnerait comme un centre de traduction et d'interprétation pour de nombreux programmes d'expression ⁶⁰.

Il a été démontré que Mediator participait au maintien des régions déplétées en nucléosomes ⁶⁷. Des études de communication génomique, entre les enhanceurs et les promoteurs, durant l'activation de la réponse aux récepteurs nucléaires, ont suggéré un rôle topologique de formation ou maintien de ces communications pour Mediator ⁶⁸. Depuis, Mediator a été associé fonctionnellement avec Cohesin dans la régulation des programmes d'expression tissus-spécifiques ⁶⁹.

5.2 Le complexe multifonctionnel Cohesin : maintenance de la structure chromatinienne et régulation de l'expression génique

Le complexe de maintenance de la structure des chromosomes, Cohesin, est composé des sous-unités SMC1, SMC3, RAD21 et STAG2, et est un constituant majeur des chromosomes mitotiques et interphasiques. Cohesin est surtout connue pour ses fonctions dans la cohésion des chromatides sœurs suite à la réplication de l'ADN ⁷⁰. Cette cohésion est essentielle pour la ségrégation des chromosomes ainsi que pour la réparation des cassures double brins de l'ADN ⁷¹. De nombreuses études confèrent également un rôle pour Cohesin dans la régulation des gènes ainsi que dans le maintien de la structure de la chromatine ⁷².

Une évidence suggérant un rôle pour Cohesin dans la régulation génique est dans la démonstration par ChIP-seq que Cohesin et son partenaire NIPBL (responsable du recrutement de Cohesin sur la chromatine) occupent préférentiellement les régions actives du génome de la drosophile, présentant une forte co-localisation avec ARNPII ⁷³. Certaines pathologies ont été associées à des mutations dans des sous-unités de Cohesin ainsi que dans NIPBL, notamment dans le

syndrome de Cornelia de Lange (CdLS). Les auteurs démontrent que les anomalies développementales observées ne sont pas le résultat de défauts aberrants de cohésion mais reflètent plutôt des fonctions distinctes des protéines de cohésion, probablement dans la régulation des gènes ⁷⁴.

CTCF fonctionne comme un insulateur protéique *in vitro* ⁷⁵. Son rôle *in vivo* a d'abord été prouvé sur le *locus* endogène de la B-Globine ⁷⁶, puis à l'échelle génomique, avec l'occupation de dizaines de milliers de sites sur le génome ^{77,78}. Des études de ChIP-seq ont identifié les sites d'occupation génomique de Cohésin sur le génome humain, et ont montré une forte colocalisation avec la protéine insulatrice CTCF, présentant certaines interdépendances fonctionnelles ^{79,80}. Il a été proposé que Cohésin participe activement aux fonctions insulatrices de CTCF.

Cohésin a une fonction centrale dans le maintien des programmes d'expression. Avec son rôle bivalent dans la régulation génique ainsi que dans la structure de la chromatide, Cohésin est capable de former des boucles topologiques et orchestre ainsi la communication génomique fonctionnelle à de multiples échelles.

6. Le rôle de la communication génomique dans la régulation des programmes d'expression géniques

6.1 La répartition nucléaire des chromosomes n'est pas aléatoire mais fonctionnelle

L'étude de la structure des chromosomes et leurs interactions dynamiques avec d'autres composants nucléaires connaît un essor croissant avec les récentes avancées technologiques. L'imagerie microscopique par fluorescence (FISH) permet de visualiser directement le comportement tri-dimensionnel de certains *loci* par marquage génomique spécifique ⁸¹. Les nombreux résultats obtenus à l'aide de cette technique ont par exemple montré que les chromosomes n'étaient pas répartis aléatoirement dans l'espace nucléaire, mais qu'ils présentaient une certaine organisation préférentielle, intra ou péri nucléaire, en fonction du nombre de gènes qu'ils comprenaient ⁸². Cette technique de FISH a par la suite permis la découverte de contacts inter et intra chromosomiques, entre régions transcriptionnellement actives, au centre du noyau. Ces résultats ont vu naître une relation entre la topologie des chromosomes et leurs fonctions transcriptionnelles, avec la notion de territoires chromosomiques ⁸³. La localisation d'un gène dans un territoire chromosomique semble en effet influencer son accessibilité pour des fonctions nucléaires spécifiques comme la transcription ⁹.

6.2 Les enhancers forment des *clusters* dont la perturbation altère la réponse transcriptionnelle

Récemment, de nouvelles technologies, plus résolutive et quantitative, permettant l'analyse des interactions médiées par la chromatine, ont permis de mieux comprendre les mécanismes de régulation transcriptionnelle à l'échelle non plus seulement linéaire, mais tri-dimensionnelle, connectant ainsi fonctionnellement le contrôle de la structure chromatinienne avec la régulation des programmes d'expression génique⁸⁴. Des expériences de 3C (Capture de la Conformation Chromatinienne) ont permis d'analyser la structure de la chromatine et ses interactions tri-dimensionnelles à certains *loci* et ont ainsi démontré que plusieurs éléments enhancers étaient capables de s'associer dans l'espace nucléaire⁸⁵. Ce *clustering* d'enhancers est d'ailleurs facilité par le complexe de maintenance de la structure de la chromatine, Cohesin. La moitié des gènes dérégulés en réponse à la réduction de Cohesin résident au sein des structures formées par l'interaction des enhancers entre eux, suggérant un rôle pour Cohesin dans la régulation génique, via le *clustering* spatio-fonctionnelle des éléments régulateurs

6.3 Les enhancers distaux forment des contacts spécifiques avec les promoteurs de leurs gènes cibles

La communication spécifique, fonctionnelle et spatio-temporelle, entre un enhancer et son promoteur correspondant, est essentielle pour la régulation orchestrée et l'exécution précise des programmes d'expression géniques (**Figure 1.3A**). Par définition, les enhancers contribuent de façon additive et parfois redondante à l'expression de leurs gènes cibles³⁵. Jusqu'à récemment, peu d'outils étaient disponibles pour lier les enhancers à leurs promoteurs cibles. Les enhancers, découverts il y a une trentaine d'années, ont longtemps été prédits en utilisant différents profils génomiques : par exemple la présence de motifs pour des facteurs transcriptionnels et leur conservation, l'occupation réelle observée notamment en ChIP-seq sur le génome, l'accessibilité et l'état de la structure de la chromatine, ou encore les modifications des queues amino-terminales des protéines histones (**Figure 1.3A**)²⁶.

Le ChIA-PET combine les approches de ChIP-seq et de 3C pour identifier les interactions chromatiniennes médiées par une protéine en particulier⁸⁴. Des études ont développé des méthodes intégratives pour analyser l'interactome global des contacts enhancers-promoteurs dans plusieurs lignées cellulaires^{86,87}. Ils ont notamment démontré que les interactions enhancers-promoteurs avaient un degré de spécificité cellulaire plus fort que les enhancers pris individuellement (**Figure**

1.3A). De plus, les promoteurs fortement exprimés de manière tissu-spécifique ont tendance à être régulés par de multiples enhanceurs, qui présentent un degré de conservation plus faible. Enfin, le complexe Cohesin est responsable de la formation et du maintien de ces boucles topologiques fonctionnelle et ce de manière CTCF indépendante (**Figure 1.3A**).

6.4 RNAPII forment des foyers de transcription actifs auxquels peuvent se connecter les interactions enhanceurs-promoteurs pour la co-régulation de plusieurs gènes

Les progrès techniques en FISH et les approches moléculaires ont permis d'identifier des centaines voire des milliers de foyers dans lesquels sont concentrés les transcrits naissants ainsi que la forme active de l'ARNPII, dépendamment du type cellulaire et de son état de différenciation (**Figure 1.3B**)⁸⁸. Ces foyers font 45 à 100 10^{-9} mètres de diamètre et sont riches en protéines. Le modèle propose que ces usines à transcription se forment comme conséquence de l'activation de l'expression de gènes fortement induits et que ces usines sont ensuite capables de recruter d'autres gènes. Les interactions enhanceur-promoteur montrent une forte spécificité tissulaire et ont lieu dans ces foyers de concentration de l'ARNPII⁸⁹. La transcription de gènes co-régulés a lieu dans le contexte d'interactions et de contacts chromosomiques qui forment des complexes multigéniques. Ces contacts, ainsi que l'activité transcriptionnelle qui en résulte, sont perdus lors de la déplétion de facteurs de transcription et de protéines de structure de la chromatine⁹⁰.

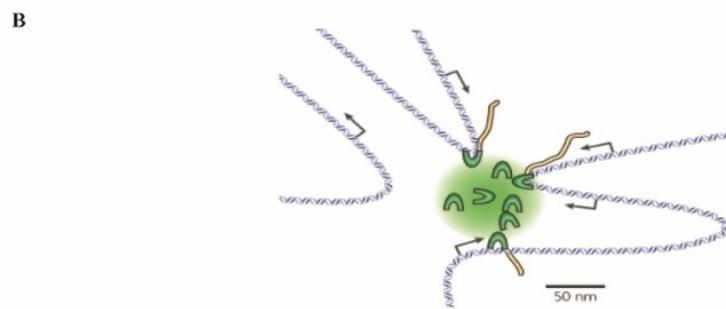
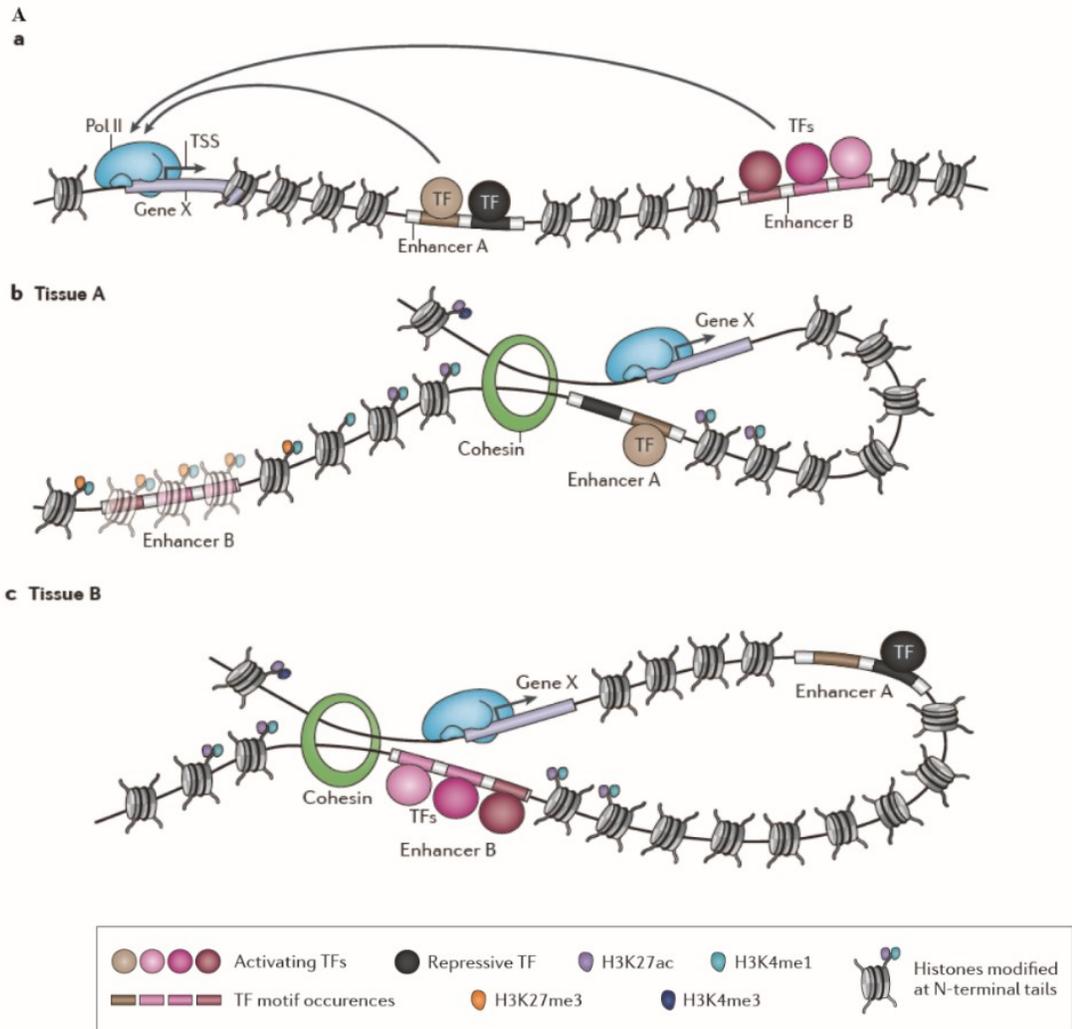


Figure 1.3: Le rôle de la communication génomique dans la régulation des programmes d'expression.

(A) Les enhancers forment des contacts chromatinien **tissus-spécifiques** (a) Les enhancers peuvent être localisés à presque n'importe quelle distance du TSS de leurs gènes cibles, et peuvent contrôler un même promoteur selon le contexte cellulaire, ce qui rend leur caractérisation fonctionnelle difficile. **(b) (c)** Dans un tissu donné, les enhancers actifs (A en **(b)**, B en **(c)**) sont occupés par des facteurs de transcription tissus-spécifiques et sont rapprochés physiquement de leurs gènes cibles par la formation de boucles topologiques dans l'ADN, dont la formation dépend de Cohésin et d'autres complexes protéiques. Les enhancers et les promoteurs actifs sont caractérisés par une déplétion nucléosomale. Les nucléosomes bordant les régions ouvertes sont modifiés épigénétiquement. Par exemple, l'acétylation de la lysine 4 de l'histone H3 (H3K27Ac) est typique d'un enhancer actif. Les enhancers inactifs sont plutôt caractérisés par des modifications répressives des queues d'histones **(b)** ou la présence de répresseurs transcriptionnels **(c)** (Tiré de Shlyueva et al. ²⁶).

(B) Le modèle d'usine à transcription pour le *clustering* spatio-fonctionnel des enhancers et la co-régulation génique. Ce modèle représente une usine à transcription avec un diamètre d'environ $70 \cdot 10^{-9}$ nM. Cette usine contient plusieurs molécules d'ARNPII (croissants verts). Les ARN naissants sont représentés en jaunes. Les flèches indiquent le sens de la transcription. Les gènes transcrits sur les mêmes chromosomes ou sur différents chromosomes peuvent s'associer avec la machinerie transcriptionnelle au sein de ces usines à transcription pour une co-régulation génique (Tiré de Sutherland et al. ⁸⁸).

7. La dérégulation de la transcription dans l'apparition et la progression tumorale – le cancer du sein hormono-dépendant

Les récentes avancées génomiques ont permis d'identifier la plupart des gènes, sinon tous, fréquemment mutés dans la plupart des cancers. Ces études ont révélé que sur les 25 000 gènes contenus dans l'ADN humain, plus de 500 sont liés au cancer ^{91,92}. En fait, plus d'un tiers de ces gènes associés au cancer sont soit des protéines kinases qui contrôlent les voies de transduction intracellulaires, soit des régulateurs de la transcription qui régulent l'expression génique en aval de ces voies ³⁰. Ce dernier groupe de gènes peut coder pour des facteurs de transcription, des co-régulateurs transcriptionnels, des régulateurs de la structure chromatinienne, mais aussi des ARN non codants qui interagissent avec ces régions régulatrices ³³.

Des études d'association *genome-wide* (GWAS) révèlent en fait que la grande majorité des mutations survenant dans l'apparition et la progression des cancers résident non pas dans les régions codantes mais dans les régions intergéniques du génome humain, particulièrement au sein de régions comprenant des éléments régulateurs potentiels ^{93,94}. Ces mutations peuvent altérer la fonction d'un enhancer en *cis* 1) en augmentant son nombre de copies, résultant en l'amplification de son activité, 2) en effectuant des réarrangements structuraux pouvant changer la spécificité d'un enhancer pour les promoteurs (en amont d'oncogènes par exemple), ou 3) en altérant l'occupation des facteurs de

transcription³⁰. Ces mutations peuvent générer des *super-enhancers* proches d'oncogènes^{95,96}. Par exemple, dans de multiples cellules cancéreuses, des *super-enhancers* sont retrouvés proches de l'oncogène *MYC*, et ce de façon tumeur-spécifique. Dans le myélome multiple, c'est une translocation qui déplace le *super-enhancer* associé normalement à l'expression du gène *IgH* adjacent au promoteur de *MYC*. Dans la leucémie aiguë lymphoblastique, c'est la surexpression du facteur de transcription TAL1 qui est associée à la formation d'un *super-enhancer* dans le *locus* de l'oncogène. Dans le cancer pulmonaire, une amplification implique un *super-enhancer* qui couvre le gène *MYC* et ses régions régulatrices. Des mutations peuvent également briser des éléments frontières isolateurs⁹⁷ et établir ainsi un état de régulation constitutivement actif qui dirige la prolifération incontrôlée des cellules cancéreuses (voir **Chapitre 4**). À l'inverse, les mutations dites en *trans*, comme l'expression aberrante de facteurs de transcription, discutée plus haut, qui contrôlent les enhancers, résultent de façon similaire à leur dérégulation d'activité.

7.1 Le Cancer du Sein hormono-dépendant est contrôlé par le facteur de transcription ER (Récepteur aux Estrogènes)

Le cancer du sein est l'un des cancers les plus fréquents chez la femme, sinon le premier, d'après les statistiques de la Société Canadienne du Cancer. Mais il est aussi l'un des plus hétérogènes, sa classification ne cessant d'être continuellement mise à jour⁹⁸. Le cancer du sein est souvent initié par une accumulation d'hormones estrogéniques dans l'organisme, on parle alors de cancers du sein hormono-dépendants. La première association entre l'apparition du cancer du sein et l'activité hormonale a été décrite en 1896⁹⁹.

Les estrogènes sont des hormones principalement ovariennes, dont E2 est la forme circulante la plus abondante¹⁰⁰. Les cellules luminales épithéliales constituent le site majeur d'action de E2, car elles contiennent le récepteur qui médie son action, le récepteur aux estrogènes (ER), et sont donc responsables de l'apparition du cancer du sein hormono-dépendant¹⁰¹. Il existe deux formes du récepteur ER, c'est-à-dire deux protéines codées par deux gènes différents, ER α et ER β ¹⁰². Le rôle de ce second récepteur estrogénique est pourtant mal défini, notamment dans l'apparition du cancer et de l'acquisition de résistances aux traitements, avec des modes d'action tantôt synergiques et additifs, tantôt opposés¹⁰³. De plus, son niveau d'expression est faible comparé à ER α dans le tissu mammaire normal et cancéreux, c'est pourquoi nous limiterons l'appellation de ce dernier à ER. Les cancers du sein hormono-dépendants ont souvent un bon pronostic en clinique, il suffit alors de cibler le facteur de transcription responsable de la croissance tumorale.

Les voies d'action de ER sont largement décrites dans de nombreuses revues de littérature ¹⁰³⁻¹⁰⁵. Nous en ferons ici un bref résumé, en mettant l'emphase sur sa voie d'action canonique génomique (**Figure 1.4A**). En effet, ER est un récepteur nucléaire dont les voies d'actions sont multiples et complexes, ce qui complique la compréhension des phénotypes de résistance acquises aux traitements hormonaux employés. Dans la cellule, le récepteur ER est codé par le gène *ESR1*. La protéine ER possède deux domaines d'activation, AF1 et AF2 ^{106,107}. AF2 est intégré au domaine de fixation du ligand et son activité est ligand-dépendante. AF1 est régulé par phosphorylation, par les acteurs des voies MAPK. L'action de ces domaines d'activation dépend du contexte cellulaire ¹⁰⁸.

La voie d'action canonique de ER implique son activation par son ligand, l'estrogène présent dans la cellule, qui va induire l'accumulation du récepteur actif dans le noyau (**Figure 1.4A**). En effet, ER est normalement associé à des protéines chaperonnes dans un état inactif. La fixation de E2 va permettre cette dissociation¹⁰⁹. Cette liaison va alors favoriser le changement de conformation optimal du récepteur, et sa dimérisation, pour reconnaître et lier les séquences arborant son motif de liaison à l'ADN, ERE (*Estrogen Responsive Elements*)¹¹⁰ (**Figure 1.4A**). La fixation ER sur la chromatine va initier le recrutement de nombreux co-facteurs (**Figure 1.4A**) ^{111,112}, comme celui d'enzymes de modifications des queues amino-terminales d'histones de la famille des HAT (*Histones Acetyl-Transferases*) ainsi que celui de protéines de remodelage de la structure chromatinienne comme le complexe SWI/SNF. ER recrute également des co-régulateurs transcriptionnels au niveau des régions régulatrices et promotrices de ses gènes cibles pour finalement permettre le recrutement de la machinerie transcriptionnelle et l'initiation de l'expression des gènes régulés par E2 (**Figure 1.4A**). Parmi ces co-activateurs, il a été montré que MED1 jouait un rôle important dans la régulation de la transcription induite par ER ^{63,64}.

Un autre modèle d'action de ER est celui où le récepteur agit en tant que co-régulateur transcriptionnel, en liant de manière indirecte cette fois le génome, par exemple via les facteurs de transcription AP-1, NFκB ou encore SP-1 (**Figure 1.4A**). ER stabilise le complexe transcriptionnel ainsi organisé et peut recruter des facteurs additionnels ¹⁰⁴. ER peut également subir certaines modifications post-traductionnelles, notamment la phosphorylation par interaction avec les voies de signalisation en aval des récepteurs aux facteurs de croissance (**Figure 1.4A**). ER ainsi modifié peut alors activer la transcription de manière ligand indépendante, mais également influencer les voies de signalisation de manière non-génomique, il s'agit de son troisième mode d'action, et il est le plus mal connu à ce jour ¹⁰⁴. La forme activée de ER associée à la chromatine est rapidement ubiquitinylée puis dégradée par le protéasome ^{113,114}.

Plus récemment, des études originales ont permis une meilleure compréhension à l'échelle génomique de la régulation transcriptionnelle induite par ER. La caractérisation du *cistrome* de ER en ChIP-seq ¹¹⁵, ainsi que l'intégration des progrès en analyses 3D du génome occupé par ER et ARNPII ¹¹⁶, ont permis de comprendre l'interactome chromatinien de ER et les mécanismes de régulation génique impliqués : la régulation de la transcription médiée par ER implique des interactions chromatiniennes entre des régions séparées par de longues distances génomiques ainsi que l'action du facteur de transcription pionnier FOXA1 ^{47,117-119}.

7.2. Stratégies de thérapies anti-estrogéniques et modèles de résistance acquise à ces thérapies transcriptionnelles

La thérapie privilégiée des cancers du sein hormono-dépendants, particulièrement ceux ayant formé des métastases, est la thérapie endocrinienne. La détermination du niveau d'expression de ER accompagnant l'apparition de chaque cancer est alors cruciale pour savoir si le patient va bénéficier de la thérapie endocrinienne ou non. Les revues décrites recensent également les mécanismes d'action exercés par l'application des traitements endocriniens ¹⁰³⁻¹⁰⁵.

Il existe plusieurs familles de traitements hormonaux (**Figure 1.4B**). Les SERMs (*Selective Estrogen Receptor Modulators*), comme le Tamoxifen, utilisé depuis 1973 ¹²⁰, agissent au niveau du site de fixation du ligand par compétition d'affinité. Ils inhibent le domaine d'activation AF2 du récepteur, en induisant un changement de conformation non fonctionnel qui va séquestrer le domaine dans un état inactif ¹²¹. Les SERMs antagonisent donc l'activation du facteur de transcription au niveau de ses gènes cibles (**Figure 1.4B**), mais peuvent agir comme agonistes dans certains tissus ou dans certaines conditions, compliquant leur utilisation et augmentant le risque de cancers secondaires (notamment endométrial ¹²²) et le risque de résistances acquises. Les Inhibiteurs d'Aromatase (AI) sont des inhibiteurs de la transformation de précurseurs stéroïdiens en substrats estrogéniques ¹²³. Les AI agissent par inhibition ou par compétition d'affinité au niveau du site de fixation de l'enzyme sur ses substrats. Ils réduisent localement et considérablement la production de E2 (**Figure 1.4B**). Enfin, le Fulvestrant, dont l'utilisation en clinique a été approuvée en 2002, fait partie de la troisième génération de traitements anti-estrogéniques. Il appartient à la famille des SERDs (*Selective Estrogen Receptor Destroyers*). C'est un antagoniste complet puissant qui inhibe conjointement les domaines d'activation AF1 et AF2 par compétition d'affinité et réduit la demi-vie de ER (**Figure 1.4B**) ^{124,125}.

L'utilisation adjuvante et néo-adjuvante de ces inhibiteurs hormonaux a montré de nets progrès dans la survie des patientes atteintes de cancer du sein hormono-dépendant ¹²⁶.

Malheureusement, chez la grande majorité des patientes traitées pendant plusieurs années, des phénotypes de résistances s'acquièrent et les traitements endocriniens deviennent alors inefficaces, ouvrant le domaine médical spécialisé à de nombreuses recherches moléculaires et pharmacologiques.

Dans la plupart des cas rapportés de résistance aux traitements endocriniens, ER est toujours présent, et son activité continue de réguler la croissance tumorale ¹²⁷. De nombreuses revues de littérature résument les principaux mécanismes potentiellement impliqués dans la résistance endocrinienne ^{128,129}. L'acquisition de la résistance est un mécanisme très hétérogène ¹³⁰ qui présente des acquisitions de résistance aux traitements employés par étape ¹³¹.

Au niveau moléculaire, plusieurs modèles, non mutuellement exclusifs, ont été proposés et/ou démontrés pour expliquer l'acquisition de la résistance. Des mutations du gène codant pour ER et ses partenaires ont été rapportées, augmentant la sensibilité du récepteur pour de faibles concentrations circulantes d'estrogènes dans le cas de la résistance acquise aux AIs par exemple, ou encore modifiant le contenu de la cellule en co-régulateurs pour contourner les effets du Tamoxifen, mais ces mutations restent extrêmement rares ¹⁰⁵.

La majorité des modèles accusent les modifications post-traductionnelles de ER, par l'activation oncogénique des voies de signalisation mitogéniques en aval des récepteurs aux facteurs de croissance, qui résultent en une transactivation de ER de façon ligand-indépendante (**Figure 1.4B**) ¹²⁹. Dans 30% des cancers du sein, le récepteur HER2 (codé par le gène ERBB2) est sur-exprimé. Les études réalisées par les laboratoires des Dr Lupien et Moerkens ^{132,133} proposent un mécanisme de résistance impliquant la sur-activation de HER2 : le rôle de ER serait conservé mais EGF induirait un *cistrome* différent, suffisant pour réverser l'inhibition de la prolifération cellulaire induite par l'action du Tamoxifen. La suppression du suppresseur de tumeur PTEN, associée à de nombreux cancers notamment le cancer du sein, augmente la signalisation via PI3K/AKT, avec pour même conséquence la transactivation de ER ^{134,135}.

Des essais cliniques sont en cours, notamment l'utilisation d'inhibiteurs de kinases (**Figure 1.4B**). L'utilisation de ces inhibiteurs, de façon combinatoire avec les traitements hormonaux existantes, a déjà montré des résultats encourageants ^{136,137}. De nombreux laboratoires se sont intéressés aux profils génomiques d'occupation de ER et de ARNPII en réponse aux différents traitements et ont montré un changement d'éléments *cis*-régulés, offrant un potentiel de découverte majeure aux études génomiques ¹³⁸.

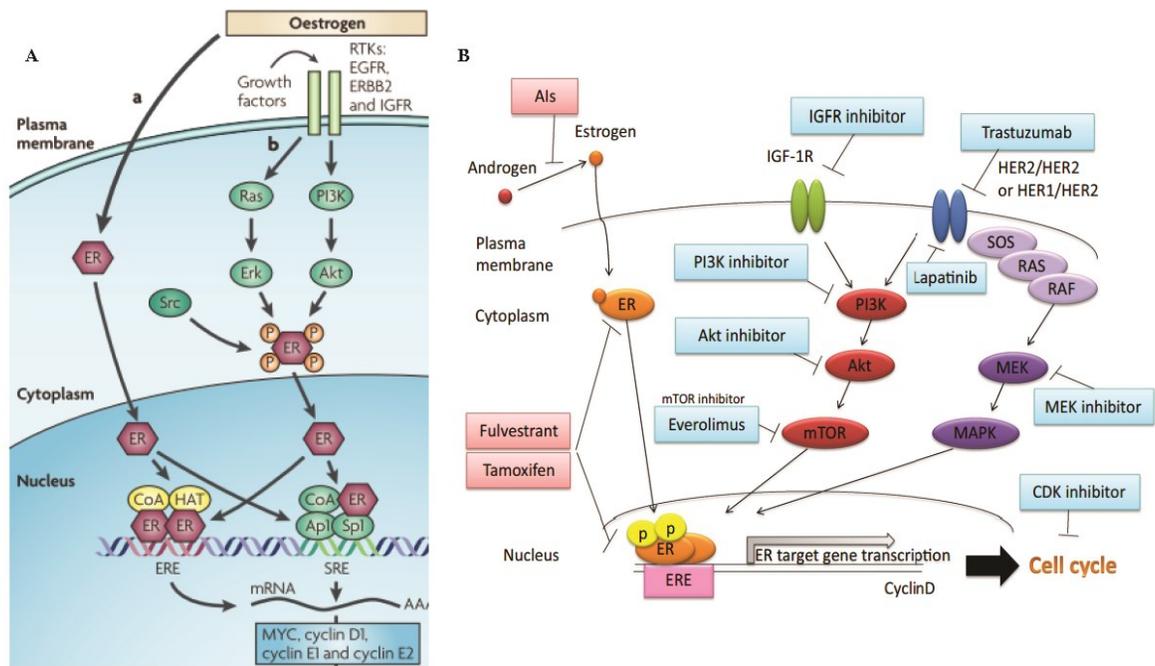


Figure 1.4 : Le récepteur aux estrogènes (ER) contrôle le programme transcriptionnel des cellules du cancer du sein hormono-dépendantes.

(A) Voies d'action génomiques de ER. (a) Dans la voie génomique canonique, les estrogènes lient le récepteur aux estrogènes (ER) et activent l'expression génique, soit par recrutement direct du dimère ER à des motifs de reconnaissance spécifiques ERE, sur des complexes incluant des co-activateurs et des HATs, soit par recrutement indirect et par des interactions protéiques avec d'autres facteurs de transcription, particulièrement les membres de la famille AP-1 et SP-1. (b) ER peut aussi être activé en réponse à l'activation de voies de signalisation en aval des récepteurs aux tyrosine-kinases, comme EGFR, IGFR ou encore ERBB2 (HER2), par les sérine/thréonine-kinases telles que ERK ou AKT, résultant en l'activation ligand-indépendante de ER (Adapté de Musgrove et al. ¹²⁹).

(B) Cibles thérapeutiques moléculaires du cancer du sein hormono-dépendant. Le cancer du sein hormono-dépendant est traité avec des agents anti-estrogéniques tels que les AIs, qui inhibent la synthèse de E2, le Tamoxifen (SERM) ou encore le Fulvestrant (SERD) qui antagonisent l'action de ER. L'apparition de nombreuses résistances aux traitements endocriniens a entraîné l'utilisation d'inhibiteurs des voies de signalisation responsables de la transactivation ligand-indépendante de ER, ou encore d'inhibiteurs de gènes mitogéniques en aval de la réponse estrogénique (Adapté de Hayashi et al. ¹³⁰).

8. Objectifs de recherche

Les mécanismes de régulation qui établissent et contrôlent les programmes d'expression géniques présentent une très grande complexité de par leur spécificité, leur coopérativité, leur redondance, et impliquent de nombreux contacts fonctionnels à de multiple échelles. Leur

interprétation est donc limitée à la compréhension de ces mécanismes, notamment dans le cas de maladies, comme le cancer, ou ils sont dérégulés ³³.

Contrairement aux facteurs de transcription, les co-activateurs ont des rôles plus conservés et généraux, et sont essentiels pour l'interprétation correcte et fonctionnelle des programmes transcriptionnels ^{59,60}. Les travaux précédents de mon directeur de recherche ont montré que le co-activateur Mediator, le complexe de maintenance de la structure chromosomique Cohesin, ainsi que son facteur de fixation à la chromatine NIPBL, contrôlaient conjointement l'exécution des programmes d'expression tissus-spécifiques des cellules normales ⁶⁹ (**Figure 1.5**).

La dérégulation des programmes transcriptionnels est une caractéristique principale du cancer ³⁰. Au sein du laboratoire, les questions d'intérêt général sont i) quels sont les déterminants majeurs de la mise en place et du maintien des programmes transcriptionnels et ii) les cellules cancéreuses suivent-elles les mêmes lois qui régissent la régulation génique dans les cellules normales.

Les résultats présentés dans le **Chapitre 2** de ce mémoire semblent suggérer que la régulation des programmes transcriptionnels des cellules cancéreuses suit les mêmes règles que pour les cellules normales et implique les mêmes acteurs transcriptionnels, soulignant un nouveau rôle des facteurs pionniers de la famille FOXA dans le recrutement des co-activateurs Mediator et Cohesin.

Nous avons ensuite voulu appliquer ces découvertes à la problématique de la résistance endocrinienne, une manifestation plus agressive du cancer du sein hormono-dépendant suite à des traitements ciblant la transcription, pour identifier les facteurs impliqués dans cette progression cancéreuse et comprendre comment le programme transcriptionnel et sa régulation s'adaptent pour survivre à des pressions de sélection d'ordre transcriptionnelles.

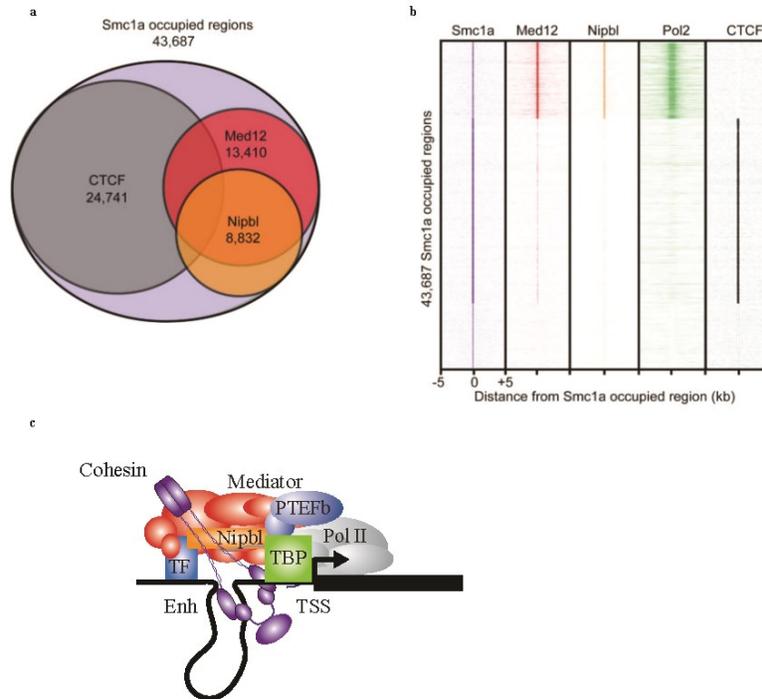


Figure 1.5: Mediator, NIPBL et Cohesin régulent le programme transcriptionnel des cellules normales.

(a) Diagramme de Venn illustrant le pourcentage de régions génomiques co-occupées par MED12 (Mediator), SMC1A (Cohesin), NIPBL et CTCF. **(b)** Cartographie des régions génomiques occupées par chaque facteur. SMC1A, NIPBL et MED12 co-occupent des régions généralement riches en ARNPII. SMC1A montre une distribution bi-modale associée ou bien à NIPBL, MED12 et ARNPII, ou bien associée à la protéine d'architecture CTCF. Pour chaque région occupée par SMC1A, l'occupation de MED12, NIPBL, ARNPII et CTCF est indiquée dans une fenêtre de 10 kb, centrée sur la région SMC1A (Tiré de Kagey et al. ⁶⁹). **(c)** Les complexes Mediator et Cohesin, et la protéine NIPBL contrôlent conjointement les programmes d'expression génique des cellules normales, par la formation de boucles topologiques, l'association fonctionnelle des enhancers et de leurs promoteurs respectifs, et le recrutement de la machinerie transcriptionnelle aux TSS ⁶⁹.

**Chapitre 2 : FOXA et les facteurs de transcription maîtres
recrutent Mediator et Cohesin aux gènes actifs dans les cellules
cancéreuses**

Avant-propos

L'article inséré dans ce mémoire a été accepté et publié par le journal d'édition Scientific Reports le 14 octobre 2016.

Les auteurs Gaëlle Bourriquen, Michèle Fournier, Maxime Côté, Vicky Caron et Steve Bilodeau ont conçu et réalisé les expériences. Les auteurs Fabien Claude-Lamaze, Eric Fournier, Charles Joly-Beauparlant, Arnaud Droit et Steve Bilodeau ont conçu et effectué les analyses bioinformatiques et computationnelles. L'auteur Stéphane Gobeil a fourni les constructions plasmidiques pour les ShRNA. Steve Bilodeau a rédigé le manuscrit.

Co-première auteure de l'article inséré, je suis responsable des expériences effectuées dans la lignée cellulaire MCF7 et du montage de la plupart des figures principales et supplémentaires. J'ai participé aux stratégies d'analyses bioinformatiques ainsi qu'à la relecture du texte principal, la rédaction des légendes ainsi que du matériel et méthodes.

Le manuscrit a subi de légères modifications d'ordre typographique pour son insertion au sein de ce mémoire. Il est inséré dans sa langue de soumission, en anglais.

FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells

Michèle Fournier^{1, 2¶}, Gaëlle Bourriquen^{1, 2¶}, Fabien C. Lamaze^{1, 2}, Maxime C. Côté^{1, 2}, Éric Fournier^{1, 2}, Charles Joly-Beauparlant², Vicky Caron^{1, 2}, Stéphane Gobeil^{2, 3}, Arnaud Droit^{2, 3} and Steve Bilodeau^{1, 2, 4*}

¹ Centre de Recherche sur le Cancer de l'Université Laval

² Centre de recherche du CHU de Québec – Université Laval

³ Département de médecine moléculaire, Faculté de Médecine, Université Laval, Québec, Canada

⁴ Département de biologie moléculaire, biochimie médicale et pathologie, Faculté de Médecine, Université Laval, Québec, Canada

¶These authors contributed equally to this work.

* Corresponding author

E-mail: steve.bilodeau@crchuq.ulaval.ca

Centre de recherche du CHU de Québec – Université Laval

9 McMahon Street

Québec City, Canada

G1R 2J6

Phone: (418) 525-4444 ext.15550

Fax: (418) 691-5439

Résumé

Le contrôle du programme d'expression des gènes est essentiel au maintien de l'identité et des fonctions biologiques d'une cellule. Les complexes Mediator et Cohésin ont été établis comme co-facteurs centraux contrôlant le programme transcriptionnel des cellules normales. Cependant, leur distribution, recrutement et importance n'ont pas été complètement démontrés dans les cellules cancéreuses. Mediator et Cohésin occupent les régions régulatrices de gènes activement transcrits et maintiennent la prolifération et le potentiel tumorigène des cellules cancéreuses. A travers l'intégration de données publiques, nous avons montré que les facteurs de transcription pionniers FOXA, ainsi que les facteurs de transcription maîtres spécifiques, établissent le réseau transcriptionnel des cellules cancéreuses. Leur perte phénotypie celle de Mediator et Cohésin. Enfin, ces facteurs semblent essentiels au recrutement de Mediator et Cohésin sur le génome. Notre étude propose que le maintien des cellules cancéreuses dépende du recrutement de Mediator et Cohésin par les facteurs FOXA et maîtres.

Abstract

Controlling the transcriptional program is essential to maintain the identity and the biological functions of a cell. The Mediator and Cohesin complexes have been established as central cofactors controlling the transcriptional program in normal cells. However, the distribution, recruitment and importance of these complexes in cancer cells have not been fully investigated. Here we show that FOXA and master transcription factors are part of the core transcriptional regulatory circuitry of cancer cells and are essential to recruit Mediator and Cohesin. Indeed, Mediator and Cohesin occupied the enhancer and promoter regions of actively transcribed genes and maintained the proliferation and colony forming potential. Through integration of publically available ChIP-Seq datasets, we predicted the core transcriptional regulatory circuitry of each cancer cell. Unexpectedly, for all cells investigated, the pioneer transcription factors FOXA1 and/or FOXA2 were identified in addition to cell-specific master transcription factors. Loss of both types of transcription factors phenocopied the loss of Mediator or Cohesin. Lastly, the master and pioneer transcription factors were essential to recruit Mediator and Cohesin to regulatory regions of actively transcribed genes. Our study proposes that maintenance of the cancer cell state is dependent on recruitment of Mediator and Cohesin through FOXA and master transcription factors.

1. Introduction

Transcription factors enforce the cell-specific transcriptional program by binding regulatory elements dispersed throughout the genome³⁵. Indeed, their unique ability to recognize specific DNA sequences makes them the central piece of cell state control. While hundreds of transcription factors are expressed in a cell, only a handful are essential to establish and maintain the transcriptional program^{139–141}. A landmark demonstration established decades ago that expression of the transcription factor MYOD1 was alone sufficient to turn fibroblasts into myoblasts¹⁴². Since then, many transcription factors have been added to the list of the reprogramming transcription factors including the various combinations of OCT4, SOX2, KLF4 and MYC to force somatic cells into an induced pluripotent cell state^{143,144}. Those dominant transcription factors enforcing a specific transcriptional program are typically part of the core transcriptional regulatory circuitry of a cell³³. These core transcriptional regulatory circuitries include master transcription factors, like MYOD1 and OCT4, which are expressed at high levels relative to other transcription factors, are essential to cell state maintenance and positively regulate cell-type-specific genes in addition to their own expression^{33,145,146}. Another class of transcription factors important for embryonic development and reprogramming are pioneer transcription factors¹⁴⁷. These transcription factors, like FOXA1, possess the ability to bind DNA sequences wrapped around nucleosomes and positively or negatively influence the transcriptional program^{47,51,148}. Overall, the intricate interplays between transcription factors define the transcriptional program and, therefore, the cellular state.

Transcription factors exert their control on gene expression through recruitment of a wide variety of cofactors, chromatin modifiers and regulators of the chromosome structure^{33,35,59}. Ultimately, transcription factors are important to recruit the RNA Polymerase II (Pol II) machinery and to regulate elongation^{149,150}. Bridging the transcription factors with Pol II is the role of the coactivator complex Mediator⁶⁰. Indeed, Mediator serves as a central scaffold in the pre-initiation complex to regulate Pol II activity. Along with the Mediator complex, the Cohesin complex, which is essential to maintain sister chromatids cohesion during cell division⁷⁰, and the cohesin loading factor NIPBL are essential to form cell-type-specific connections between enhancers and promoters⁶⁹. Other transcription factors, like MYC, promotes Pol II pause-release by recruiting the P-TEFb complex at active promoters¹⁵¹. Therefore, a combination of transcription factors is required at enhancer and promoter regions to specify the transcriptional program of a cell. Identifying the combinations of transcription factors governing the transcriptional program of each cell provides the core information to control cell state.

The gene expression program controls normal and aberrant biological functions of a cell. Indeed, to reach a disease state, cells often modify their transcriptional program to acquire new identities and functions³³. Interestingly, modulation of regulatory regions is emerging as a key feature of many human diseases^{152,153}. For example, the vast majority of single nucleotide polymorphisms (SNPs) associated with cancer are found in the non-coding fraction of the genome with a large subset inside enhancer regions¹⁵⁴. Furthermore, modulation of enhancer regions is a promising cancer treatment^{95,96}. Additionally, mutations in transcriptional regulators occupying enhancer regions, including transcription factors and cofactors, are causative of many diseases¹⁵⁵. Among them, mutations in Mediator, Cohesin and the Cohesin loader NIPBL have been associated to developmental syndromes as well as different forms of cancer^{156,157}. While it is not clear how these mutations lead to a disease state, they point to the importance of proper control of regulatory regions to maintain a normal cell state.

Here we show that FOXA pioneer transcription factors are, with master transcription factors, the core transcriptional regulatory circuitry of cancer cells. These transcription factors are essential to recruit the Mediator and Cohesin complexes to regulatory regions of actively transcribed genes. Taken together, our results suggest that Mediator and Cohesin are connected to the core transcriptional regulatory circuitry through FOXA and master transcription factors to maintain the cancer cell state.

2. Results

2.1 Mediator and Cohesin occupy regulatory regions of actively transcribed genes in cancer cells

Maintenance of the transcriptional program depends on transcription factors and cofactors including Mediator and Cohesin. Indeed, Mediator and Cohesin occupy enhancer regions with master transcription factors to control the transcriptional program and identity of embryonic stem cells⁶⁹. To define the contribution of Mediator and Cohesin to the regulation of the cancer transcriptional program, we established the genome-wide occupancy of MED1 (Mediator), SMC1A (Cohesin) and the cohesin loader NIPBL. Three cancer cell lines were profiled using chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq): MCF7 (breast), HEPG2 (liver) and A549 (lung). Principal component analyses (PCA) confirmed that the distribution profiles of NIPBL and MED1 were positively correlated ($R=0.53$, 0.45 and 0.53 for MCF7, HEPG2 and A549, respectively, all p -values $< 10e-100$) across the genome of each cell type (**Figure 2.1A**) in accordance with previous reports⁶⁹. Furthermore, Cohesin, which is known to occupy regions with NIPBL and Mediator, but also the transcription factor CTCF^{79,80}, showed a bimodal distribution in all three cell lines (**Figure 2.1A**). Therefore, for subsequent analyses, SMC1A-occupied regions were divided into CTCF-occupied (with CTCF) and not (no CTCF). Detailed examination of the ChIP-Seq density profiles confirmed the presence of Mediator, Cohesin and NIPBL at the transcription start site (TSS) and well-characterized enhancers regions of genes associated with breast, liver and lung functions (**Figure 2.1B**). These results suggest that similar to observations made in primary cells⁶⁹, Mediator and Cohesin co-occupy the genome at regulatory regions of actively transcribed genes in cancer cells.

The mammalian genome is segmented into functional elements orchestrating transcriptional events and gene expression¹⁵⁸. To extend the occupancy profiles of Mediator, Cohesin and NIPBL genome-wide, we compared the ChIP-Seq data with predicted functional elements. For MCF7 cells, histone marks were combined (see **Supplementary Information**) to define the different chromatin states while an 18-state model built using ChromHMM was available for A549 and HEPG2^{46,158}. MED1, SMC1A and NIPBL occupied active enhancers and promoters in all three cell types (**Figure 2.1C, D**). Lowly transcribed regions including heterochromatin and polycomb-occupied sites were not enriched. While the functional elements occupied by Mediator, Cohesin and NIPBL are similar in the three cancer cells investigated, the coordinates of these elements are different. Indeed, regions occupied by MED1, SMC1A (no CTCF) and NIPBL in MCF7 were different from those found in

HEPG2 and A549 (**Figure 2.1E**). However, regions occupied by CTCF and SMC1A (SMC1A with CTCF) were highly redundant between cell lines in accordance with previous reports¹⁵⁹. Taken together, these results suggest a conserved function of Mediator, Cohesin and NIPBL in the control of regulatory elements throughout the genome of cancer cells.

2.2 Proliferation of cancer cells depends on Mediator and Cohesin

As in normal cells, the gene networks controlling the identity and functions of each cancer cell is unique. To determine the nature of the genes occupied by Mediator, Cohesin and NIPBL, we investigated the associated gene networks in MCF7, HEPG2 and A549. KEGG pathways analyses revealed the association of MED1, SMC1A and NIPBL with genes reflecting the normal and disease functions of the tissues (**Figure 2.2A and Supplementary Table S1**). For example, in MCF7 (breast) cells, genes implicated in the oxytocin signaling pathway were enriched. For HEPG2 (liver) cells, genes associated with hepatic functions like PPAR signaling, complement and coagulation cascades as well as fatty acids metabolism were strongly enriched. In A549 (lung) cells, pathways associated with cell adhesion and TGF-beta signaling were identified. Interestingly, among the direct target genes, many known oncogenes including FGFR2, MAX and PIK3R1 in MCF7, PDGFB, PPARG and HIF1A in HEPG2 and MYC, MET and EGFR in A549 were identified. These observations suggest a role of Mediator and Cohesin in the maintenance of cancer cell properties through a cell-specific network of genes.

The presence of a common transcriptional mechanism centered on Mediator, Cohesin and NIPBL could be a unifying feature of all cancer cells. Whether or not cancer cells are dependent on these cofactors to maintain their proliferative state is not known. Therefore, we quantified the cellular response of MCF7, HEPG2 and A549 with proliferation and soft-agar colony formation assays in MED1, SMC1A and NIPBL loss-of-function experiments. For each target, two independent shRNAs were validated for efficiency by Western blot and used in functional assays (**Supplementary Figure S1**). In all three cell lines, similar results were obtained (**Figure 2.2B, C**). Indeed, loss of MED1, SMC1A or NIPBL led to a global decrease in proliferation (**Figure 2.2B**). During the 7-day experiment, cells were not dividing, but were maintained. To determine if these changes affected the colony formation potential of cancer cells, we used soft-agar colony formation assays. Loss of MED1, SMC1A and NIPBL reduced the colony formation potential of MCF7, HEPG2 and A549 (**Figure 2.2C**). Indeed, all cells transduced with control shRNAs produced colonies while their number were significantly reduced with decreasing levels of MED1, SMC1A and NIPBL. Altogether, these results support a role of Mediator and Cohesin in the maintenance of cancer cell properties.

2.3 Mediator and Cohesin connect with the core transcriptional regulatory circuitry of cancer cells

Cofactors are recruited to regulatory elements by transcription factors. To determine the transcription factor(s) recruiting Mediator and Cohesin, we collected publicly available ChIP-Seq datasets for transcription factors (**Supplementary Table S2**) and computed pairwise genome-wide relationships for each cell line. Transcription factors were ranked based on their percentage of overlap with MED1, SMC1A and NIPBL across the genome of each cancer cell (**Figure 2.3A**). The top 20 was selected for further filtering. To refine the list of candidates, we performed an *in silico* enrichment of the binding motifs found in the HOCOMOCO database in regions occupied by MED1, SMC1A and NIPBL. Indeed, transcription factors are recruited to functional elements in *cis* through recognition of a DNA sequence but also in *trans* through protein-protein interactions¹¹². The 20 most frequent DNA motifs found in regions occupied by MED1, SMC1A and NIPBL were cross-referenced with the list of 20 co-occupying transcription factors (**Figure 2.3A**). In each cell line, multiple transcription factors were identified. In MCF7, ERa, FOXA1, FOSL2 and JUND were found, while HNF4A, FOXA1, FOXA2 and CEBPB were identified in HEPG2 and FOSL2, FOXA1, FOXA2, JUND and ATF3 in A549. These results suggest that a combination of transcription factors is responsible for the recruitment of Mediator and Cohesin at regulatory regions.

Whether the identified transcription factors formed a core transcriptional regulatory circuitry in each cancer cell is unknown. To determine the connections between the transcription factors and their corresponding genes, we built the transcriptional network using the ChIP-Seq data for each transcription factor (**Figure 2.3B and Supplementary Figure S2**). Interestingly, the transcription factors were interconnected in all three cell types (**Figure 2.3B**). In MCF7, ERa was found at the promoter of *ESR1*, *FOSL2*, *FOXA1* and *JUND*. In return, FOSL2, FOXA1 and JUND occupied the promoter of *ESR1* (**Figure 2.3C and Supplementary Figure S2**). The connections for MCF7, HEPG2 and A549 are illustrated in **Figure 2.3B, C**. These results suggest that the transcription factors underlying the recruitment of Mediator and Cohesin are part of the core transcriptional regulatory circuitry of cancer cells.

2.4 FOXA and master transcription factors are essential to cancer cell maintenance

To prioritize candidates for functional studies, we wanted to identify the central and most important transcription factors in each network. Interestingly, many of the identified transcription factors are or are presumed to be master and pioneer transcription factors. For example, ERa and

HNF4A were implicated with the core transcriptional regulatory circuitry of breast and liver cells^{47,160}. A shared characteristic of master transcription factors is that they form positive autoregulatory loops, regulating their own gene¹³⁹. Therefore, we identified the transcription factors occupying their own regulatory regions with Mediator, Cohesin and NIPBL. A limited number of transcription factors fulfilled that characteristic and they became our leading candidates for controlling the cancer cell state. These transcription factors included ERa in MCF7; HNF4A and FOXA2 in HEPG2 and FOSL2 in A549 (**Figure 2.3B, C**). Since, the FOXA family of pioneer transcription factors was identified in all cell lines, FOXA1 in MCF7 and FOXA2 in A549 were also included for further characterization. It is to be noted that many of these transcription factors are associated with Mediator and Cohesin in various cell types. Indeed, ERa and HNF4A were shown to interact with Mediator¹⁶¹⁻¹⁶³. In addition, ERa was co-purified with Cohesin subunits¹⁶⁴. Furthermore, FOXA1 and FOXA2 were found to recruit coactivators and interact with other transcription factors at functional enhancers^{112,165,166}.

If the identified transcription factors are essential to the function of Mediator, Cohesin and NIPBL, they should phenocopy each other in functional loss-of-function assays. Indeed, we have shown that Mediator and Cohesin are essential to the maintenance of the cancer cell properties (**Figure 2.2**). Therefore, we used the same cell proliferation and colony formation assays with validated shRNAs targeting each identified transcription factors (**Supplementary Figure S3**). In all cases, loss of the transcription factor reduced the proliferation and the colony formation potential of cancer cells (**Figure 2.4A, B**). Indeed, lower levels of ERa and FOXA1 in MCF7, HNF4A and FOXA2 in HEPG2 and FOSL2 and FOXA2 in A549 reduced the proliferation rate (**Figure 2.4A**) and the colony formation potential (**Figure 2.4B**). The results demonstrate that the identified core transcriptional regulatory circuitry in each cell is essential to maintain cancer cells.

2.5 FOXA and master transcription factors recruit Mediator and Cohesin

Integration of the ChIP-Seq results and the functional experiments suggest that Mediator and Cohesin are recruited to regulatory regions of actively transcribed genes by the core transcriptional regulatory circuitry. To functionally validate our results, we focused our analysis on MCF7 cells. Pioneer transcription factors like FOXA1 are known to bind the condensed chromatin prior to transcriptional activation⁵¹. On the other hand, master transcription factors like ERa tend to occupy nucleosome-free regions⁵¹. We entertained two models: 1) an all-or-nothing model where the presence of many transcription factors is required to stabilize large cofactors complexes like Mediator and Cohesin and 2) a sequential model where the transcription factors and cofactors are orderly recruited. To determine the best model, we proceeded with loss-of-function experiments for ERa and

FOXA1 followed by ChIP of MED1 and SMC1A (**Figure 2.5**). The same shRNA as previously described (**Figure 2.4 and Supplementary Figure S3**) were used. Molecularly, loss-of-function of the master transcription factor ERa and the pioneer transcription factor FOXA1 led to decreased recruitment of MED1 and SMC1A at enhancers and promoters of active genes *GREB1* (**Figure 2.5A**) and *TTF1* (**Figure 2.5B**). The total protein levels of the cofactors were not affected by the decreased expression of ERa and FOXA1 (**Supplementary Figure S4**). Therefore, we conclude that a combination of transcription factors is required to recruit and stabilize large protein complexes like Mediator and Cohesin to regulatory regions of actively transcribed regions.

3. Discussion

While each cell presents unique biological characteristics, our results propose a shared common ground in transcriptional regulation. Indeed, Mediator, Cohesin and NIPBL occupy enhancer and promoter regions of actively transcribed genes in all primary⁶⁹ and cancer cells (**Figure 2.1**) tested so far. In addition, cancer cells are no exception to the simple rule stating that transcription factors provide cell specificity to the transcriptional program. Our results establish a model where a combination of FOXA and master transcription factors are part of the core transcriptional regulatory circuitry and essential to maintain the cancer cell properties (**Figures 2.3 and 2.4**). Regulatory regions are occupied by dozens and potentially more transcription factors in each cell¹⁶⁷. Surprisingly, loss of a single transcription factor affected the recruitment of Mediator and Cohesin (**Figure 2.5**) pointing toward a combinatorial effect in the stabilization and function of Mediator and Cohesin. This observation is corroborated by most reprogramming experiments where a combination of transcription factors is required to reach a new cell state³³. However, we cannot exclude the possibility that the similar phenotypes observed in Mediator, Cohesin and transcription factor loss-of-function experiments (**Figures 2.2 and 2.4**) could be attributable to independent mechanisms including cell cycle defects. Therefore, our results suggest a non-redundant role of transcription factors at regulatory regions and postulate that FOXA and master transcription factors are required to stabilize large cofactor complexes like Mediator and Cohesin.

The identification of FOXA1 or FOXA2 as essential transcription factors in cancer cells (**Figures 2.3 and 2.4**) proposes the existence of a conserved role of the FOXA family in the core transcriptional regulatory circuitry. As pioneer transcription factors, FOXA1 and FOXA2 recognize their DNA motif even when the DNA is wrapped around nucleosomes⁵¹. This property could lead the way for master transcription factors to activate gene transcription. In hormone-dependent cancers, FOXA1 is essential to reprogram the genome-wide occupancy of the estrogen and androgen receptors and rewire the transcriptional program^{117,168}. These observations suggest that FOXA transcription factors have the ability to direct binding of master transcription factors and to control the core transcriptional regulatory circuitry. Considering this central role of FOXA transcription factors, the successful development of specific inhibitors could lead to a new class of molecules to target cancer cells.

While key transcriptional regulators in breast adenocarcinoma (MCF7) and hepatocarcinoma (HEPG2) are known^{47,117,160}, transcription factors maintaining lung adenocarcinoma (A549) are not

well established. Integration of the genome-wide positioning of Mediator and Cohesin provide a valuable tool to predict key transcription factors in cell state maintenance. Indeed, our results suggest that the core transcriptional regulatory circuitry of A549 lung cancer cells is centered on FOSL2 and FOXA2 (**Figures 2.3 and 2.4**). Interestingly, both FOXA2 and FOSL2 are suspected to play major role in lung cancer progression. Indeed, FOSL2 is important to integrate the signaling through the TGF-beta/SMAD pathway and promote cellular migration of non-small cell lung adenocarcinomas¹⁶⁹. Furthermore, decreased expression of FOXA2 following TGF-beta treatment leads to epithelial-to-mesenchymal transition¹⁷⁰. Master transcription factors have been shown to determine the response to TGF-beta in various tissues through SMAD recruitment at regulatory regions¹⁷¹. One may speculate that decreased levels of FOXA2 allows the rewiring of the core transcriptional regulatory circuitry and free up FOSL2 for a role during the epithelial-to-mesenchymal transition. Therefore, the core transcriptional regulatory circuitry containing FOSL2 and FOXA2 could control the balance between maintenance of a lung epithelial *versus* mesenchymal phenotype.

The possibility of a general transcriptional mechanism centered on Mediator and Cohesin controlling the transcriptional program of normal and cancer cells raises questions regarding many human diseases. In addition to various types of cancers, genetic alterations in Mediator and Cohesin have been observed in developmental syndromes, neurological disorders and metabolic diseases^{157,172}. For example, mutations of the cohesin subunit *STAG2* are frequent in various cancers¹⁷³. Interestingly, evidence of aneuploidy are questioned and a transcription-dependent function is favored to explain the contribution of the Cohesin mutations toward cancer^{156,174}. Similarly, *MED12* is mutated at high frequencies in uterine leiomyomas and prostate cancers^{175,176}. However, in uterine leiomyomas, all *MED12* mutations are found in exon 2 emphasizing alteration of a specific function of the Mediator subunit¹⁷⁵. Additionally, during embryonic development, complete loss of Mediator and Cohesin are severely detrimental while missense mutations are less deleterious, but with complex tissue-specific phenotypes^{157,172}. Taken together, these observations suggest that Mediator and Cohesin are potentially involved in a tissue-specific rather than a general transcriptional mechanism. Furthermore, the absence of genomic abnormalities in patients with Mediator and Cohesin mutations minimizes the possibility of cell cycle defects as the primary cause of these diseases. Therefore, we propose that modulation of the activity of the core transcriptional program is the central theme of many human diseases.

If transcriptional regulation is the primary cellular function of Mediator and Cohesin, their importance for other diseases is likely underestimated. Indeed, modulating the activity of regulatory

regions is suspected to be an important mechanism leading to human diseases as SNPs are common in enhancer regions ^{96,154}. Variations in the sequence of an enhancer or promoter regions affect the binding affinity of the residing transcription factors ¹⁷⁷ which will result in differential recruitment of their associated cofactors including Mediator and Cohesin. Further experiments will demonstrate the importance of Mediator and Cohesin recruitment through FOXA and master transcription factors for many other human diseases.

4. Figures

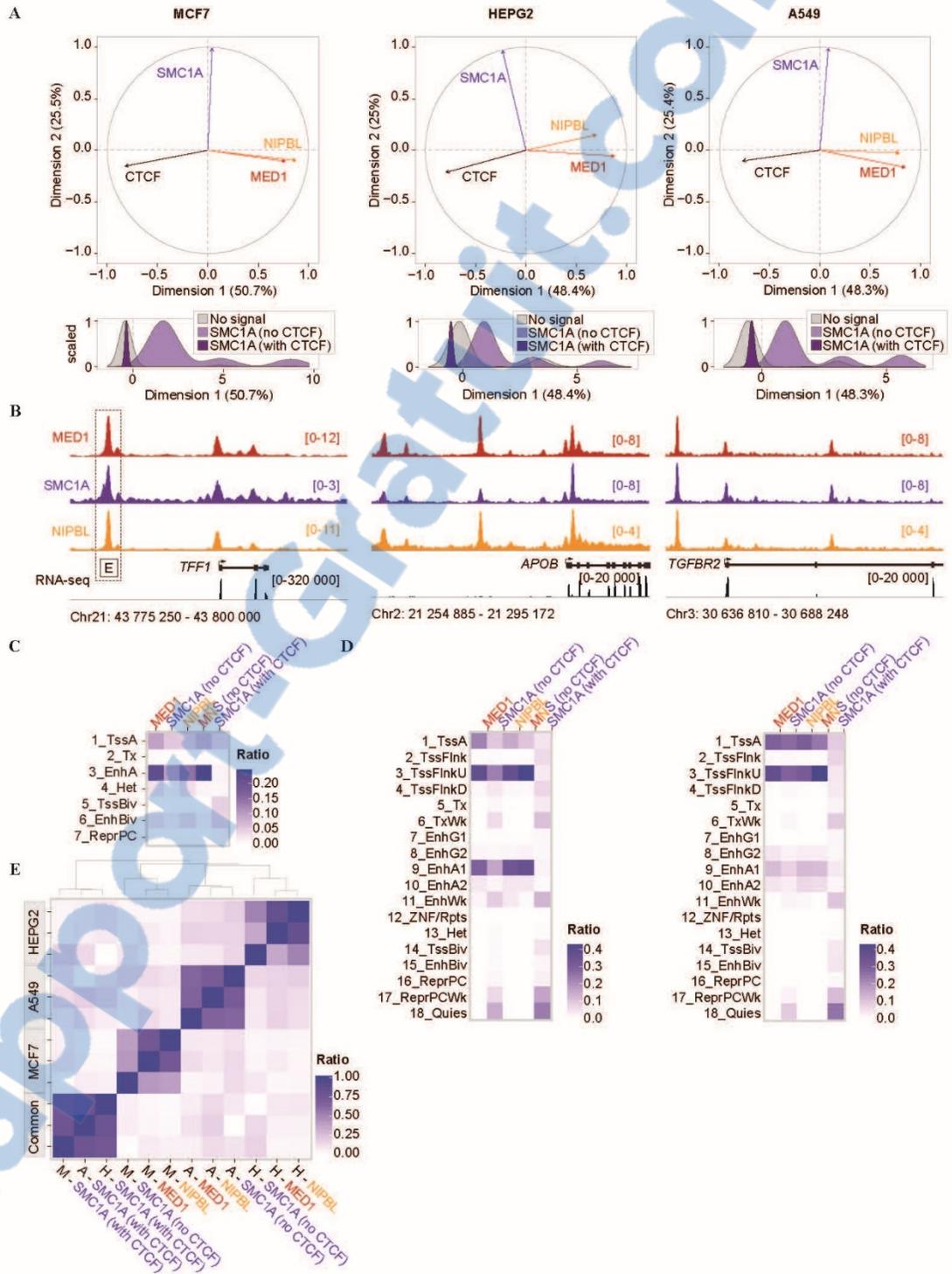


Figure 2.1: Mediator, Cohesin and NIPBL co-localize at regulatory regions of actively transcribed genes in cancer cells.

(A) Principal component analysis (PCA) of the distribution of MED1-, SMC1A-, NIPBL- and CTCF-occupied regions. Top – PCA variable factor map of the MED1, SMC1A, NIPBL and CTCF variables on the first two components showing a positive correlation between NIPBL and MED1. In addition, dimension 1 captures the negative correlation between the transcription factor CTCF and MED1/NIPBL-occupied regions. Bottom – Density plot of the genomic region coordinates on the dimension 1 showing a split distribution of SMC1A. SMC1A is associated with CTCF (negative values) or MED1 and NIPBL (positive values). The y axes are normalized to 1 while the x axes represent arbitrary units.

(B) Occupancy profiles of Mediator (MED1), Cohesin (SMC1A) and NIPBL showing colocalization at regulatory regions. The *TFF1*, *APOP* and *TGFBR2* loci are displayed for MCF7, HEPG2 and A549 respectively. The characterized enhancer region of *TFF1* (box labeled E) is indicated ¹¹⁹. The scales of ChIP-Seq and RNA-Seq profiles are displayed in reads per million.

(C) Overlaps between MED1, SMC1A (no CTCF), SMC1A (with CTCF), the regions co-occupied by MED1-SMC1A-NIPBL and the functional annotation of the MCF7 genome (see **Supplementary Information**). MED1, SMC1A (no CTCF) and NIPBL occupy active TSS (1_TssA) and active enhancer (3_EnhA), but not inactive regions. The color scale indicates the ratio of overlap.

(D) Overlaps between MED1, SMC1A (no CTCF), SMC1A (with CTCF), the regions co-occupied by MED1-SMC1A-NIPBL and the functional annotation of the HEPG2 (left) and A549 (right) genomes using the ChromHMM 18-state model ^{46,158}. MED1, SMC1A (no CTCF) and NIPBL occupy active promoters (1_TssA and 3_TssFlnkU) and active enhancer regions (8_EnhG2, 9_EnhA1 and 10_Enh_A2), but not inactive regions in both HEPG2 and A549 cells (see **Supplementary Information** for abbreviations). The color scales indicate the ratio of overlap.

(E) Heat map showing the colocalization frequencies of MED1, SMC1A and NIPBL enriched regions in MCF7, HEPG2 and A549 cells. The color scale reflects the colocalization of each pair of regulators. Regulators were clustered along both axes based on the similarity in their correlation with each other.

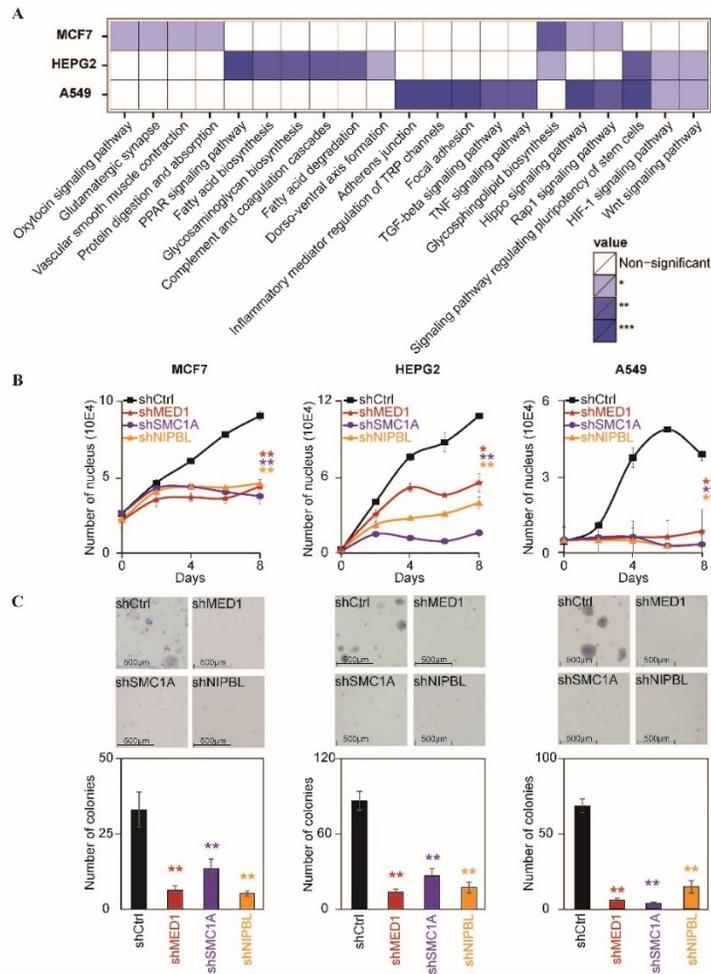


Figure 2.2: Mediator, Cohesin and NIPBL are essential for maintenance of cancer cells properties.

(A) Heat map representation of KEGG pathways enriched for genes occupied by MED1, SMC1A and NIPBL in MCF7, HEPG2 and A549 cells (Fisher's exact test, *p < 0.05, **p < 0.01, ***p < 0.001).

(B) Growth curve of MCF7, HEPG2 and A549 following decreased expression of MED1, SMC1A and NIPBL showing a global negative effect on cell growth. Each line represents the average of a minimum of two biological replicates for two independent shRNA constructs. The p-values were calculated by comparing with the shRNA controls at day 8 (paired t-test, *p < 0.05, **p < 0.01, ***p < 0.001).

(C) Soft agar colony formation assays with decreased levels of MED1, SMC1A and NIPBL showing a loss of colony formation potential. Colonies were quantified with the ImageJ software using 4 images per well. The bar graph represents the average of a minimum of two biological replicates for two independent shRNA constructs. The p-values were calculated by comparing with the shRNA controls (paired t-test, *p < 0.05, **p < 0.01, ***p < 0.001).

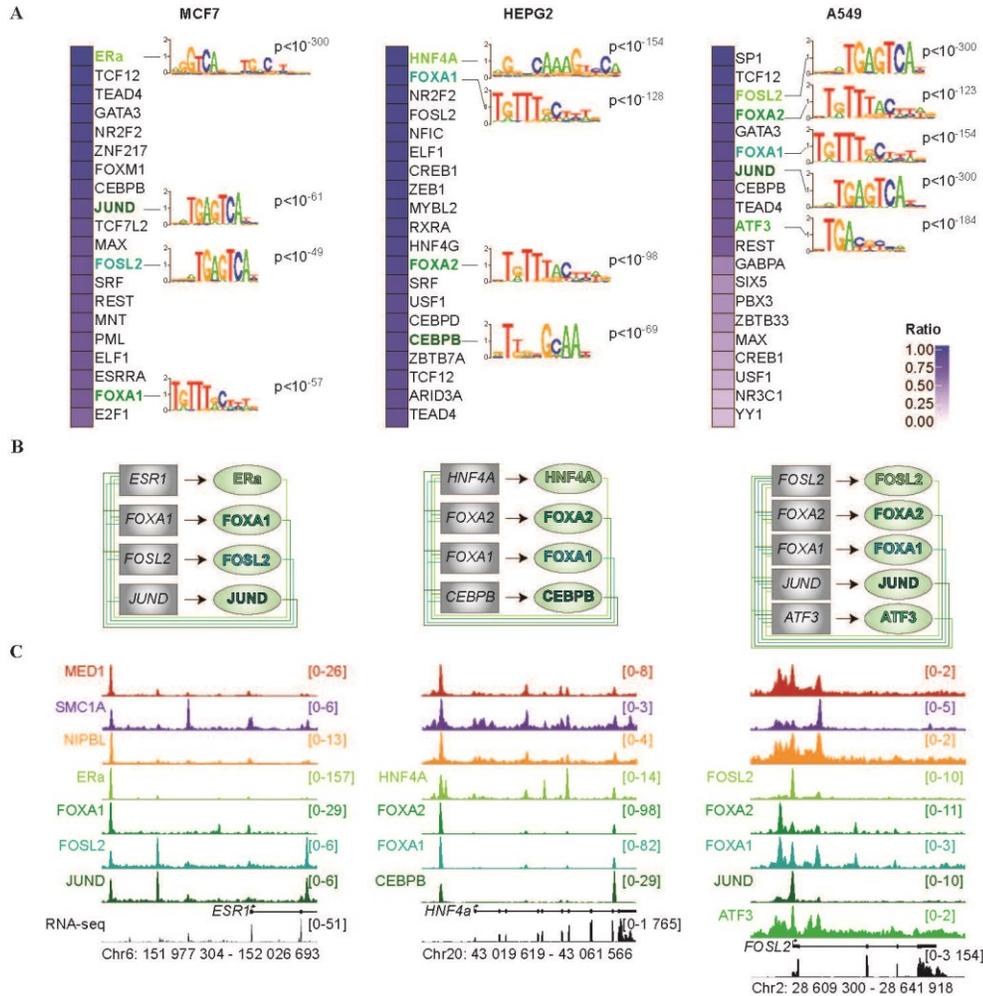


Figure 2.3: The core circuitry of cancer cells connect with Mediator and Cohesin.

(A) Percentage of overlapping regions occupied by MED1, SMC1A and NIPBL with transcription factors. The best 20 overlaps with transcription factors for each cell type are presented. The color scale indicates the percentage of overlap. The DNA binding motif logos found in the regions occupied by MED1, SMC1A and NIPBL are indicated next to their associated transcription factor. The motifs were found in the top 20 of the most enriched position-weight matrices. The associated p-values are indicated.

(B) Core transcriptional regulatory circuitry of MCF7, HEPG2 and MCF7 cells. The transcription factors identified in (A) are interconnected. The box represents the transcription factor gene loci while the oval represents the protein identified in (A). Each line represents the interaction of a transcription factor with the indicated gene loci (-10kb to the end of the gene).

(C) Occupancy profiles of Mediator (MED1), Cohesin (SMC1A) and NIPBL in addition to the core transcription factors identified in (a) showing colocalization. The *ESR1*, *HNF4A* and *FOSL2* loci are displayed for MCF7, HEPG2 and A549 respectively. The scales of ChIP-Seq profiles are displayed in reads per million.

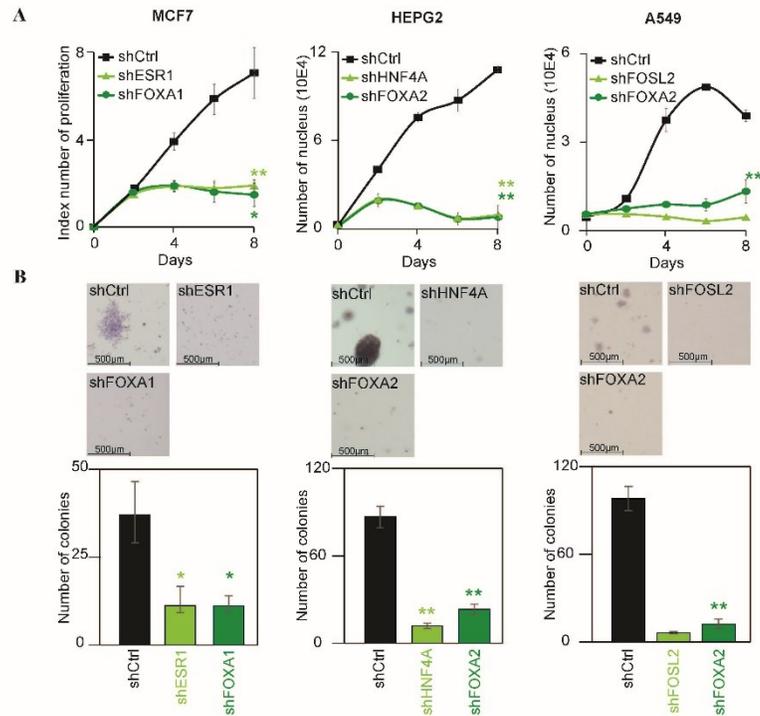


Figure 2.4: Depletion of the core transcription factors phenocopies the loss of Mediator and Cohesin.

(A) Growth curve of MCF7, HEPG2 and A549 following decreased expression of ER α and FOXA1 in MCF7, HNF4A and FOXA2 in HEPG2 and FOSL2 and FOXA2 in 549. The results show a global negative effect on cell growth. Each line represents the average of a minimum of two biological replicates for two independent shRNA constructs with the exception of FOSL2 where one shRNA construct was used. The p-values were calculated by comparing with the shRNA controls at day 8 (paired t-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

(B) Soft agar colony formation assays showing a loss of colony formation potential with decreased expression of core transcription factors. Colonies were quantified with the ImageJ software using 4 images per well. The bar graph represents the average of a minimum of two biological replicates for two independent shRNA constructs with the exception of FOSL2 where one shRNA construct was used. The p-values were calculated by comparing with the shRNA controls (paired t-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

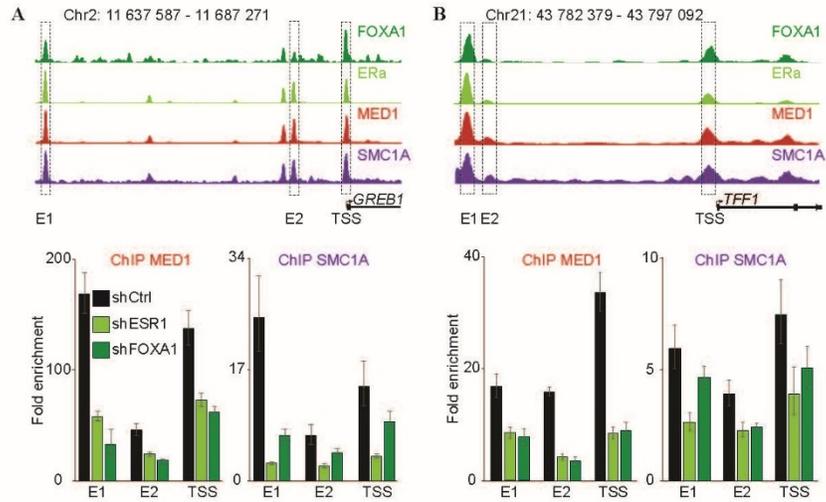


Figure 2.5: FOXA and master transcription factors are essential to recruit Mediator and Cohesin.

Depletion of either the master transcription factor ERa or the pioneer transcription factor FOXA1 in MCF7 cells decreases MED1 and SMC1 recruitment at regulatory elements of the *TFF1* (A) and *GREB1* (B) genes in MCF7. Amplified enhancer (E1 and E2) and TSS regions are indicated by boxes. Enrichment fold were calculated relative to control regions without MED1 or SMC1 signal. The bar graphs display the enrichment fold of a representative experiment. Error bars show the standard error of the mean for technical qPCR triplicates.

5. Methods

Full materials and methods are available in Supplementary Information.

5.1 Cell Culture

MCF7 (ATCC, HTB-22) and HEPG2 (ATCC, HB-8065) were grown in DMEM (Gibco, 11965-092). A549 (ATCC, CCL-185) were grown in F12K medium (Gibco, 21127022). Culture medium were supplemented with 10% fetal bovine serum (Invitrogen, qualified 12483020), 100 μ M MEM nonessential amino acids (Cellgro, 25-0250), 2 mM L-glutamine (Gibco, 25030-081), 100 U/ml penicillin, 100 μ g/ml streptomycin (Gibco, 15170-063).

For ChIP experiments, MCF7 were kept the last 3 days into DMEM w/o phenol-red (Gibco, 31053-028) supplemented with 5% of Charcoal/Dextran treated FBS (Hyclone, AVH78911), 100 μ M MEM nonessential amino acids (Cellgro, 25-0250), 2 mM L-glutamine (Gibco, 25030-081) and treated with 100nM of Beta-Estradiol (SIGMA, E8875) for two hours.

5.2 Chromatin Immunoprecipitation

ChIP-Seq were performed in biological duplicates as previously described¹⁷⁸. Briefly, cells were crosslinked with 1% formaldehyde. Sonicated DNA fragments were immunoprecipitated with antibodies directed against MED1 (Bethyl Laboratories, A300-793A), SMC1A (Bethyl Laboratories, A300-055A), NIPBL, (Bethyl Laboratories, A301-779A) and ERa (Santa Cruz, sc-543x). ChIP-seq libraries and DNA sequencing were performed by the McGill University and Génome Québec Innovation Centre.

For single gene analysis, primers were designed in the predicted enhancers and promoters regions of *TFF1* and *GREB1* loci. For the *TFF1* locus, primers GAGATGATGCCACCGTACAC & CCCTCACTCACTTTGAAGCA were used for enhancer 1 (E1), ATCCAGTCCAGGAAGGAGGT & GTCAAGAGAGGAGGCTGTGG for enhancer 2 (E2) and CCGAGTCAGGGATGAGAGG & GGCCTCCTTAGGCAAATGTT for the TSS. For the *GREB1* locus, primers GAGCTGACCTTGTGGTAGGC & CAGGGGCTGACAACTGAAAT were used for enhancer 1 (E1), GGGATATGGCTTGTCCATTGT & CATGACACCAGGACCGTAAAG for enhancer 2 (E2) and ACCCAGCAAAACACTTCAGG & ATTCAGCAGTAGCCCTTCCA for the TSS. Two negative control regions (ATGTCAGGCCCATGAACGAT & GCATTCATGGAGTCCAGGCTTT

and AGGACCTGCAGCAAACAGAA & TGTCTACATGGGCTAGTGTGCT) were used to calculate fold enrichments.

5.3 Bioinformatic analyses

All ChIP-Seq reads (**Supplementary Table S2**) were aligned to the hg19 build of the human genome using the ChIP-Seq pipeline developed at McGill University and Génome Québec Innovation Centre (https://bitbucket.org/mugqic/mugqic_pipeline). BigWig files were generated using Samtools 1.2¹⁷⁹, bedtools 2.17.0¹⁸⁰ and wigToBigWig (developed by the ENCODE team) with normalized count in millions of reads and a smoothing parameter of 225bp. Enriched regions were annotated using the UCSC database R package TxDb.Hsapiens.UCSC.hg19. KEGG pathways enrichments were assessed by hypergeometric tests. *In silico* motif searches were performed with the HOCOMOCO database using the PWMEnrich package. For overlaps with chromatin states and between ChIP-Seq datasets, the GenomicRanges R package was used.

5.4 Lentiviral shRNAs

Plasmid catalog numbers and shRNA sequences are provided in **Supplementary Table S3**. Lentiviruses were produced as previously described¹⁸¹. Cells were transduced with the lentiviruses for 24 h before puromycin selection for the indicated time. Validation of efficiency were performed by Western blot (**Supplementary Figures S1 and S3**) using MED1 (Bethyl A300-793A, 1:5000), SMC1A (Bethyl, A300-055A, 1:5000), NIPBL, (Bethyl, A301-779A, 1:1000), ERa (Santa Cruz, sc-543x, 1:5000), FOXA1 (Santa Cruz, sc-101058 1:1000), FOXA2 (Abnova, 89-019-034, 1:1000), HNF4a (Santa Cruz, sc-8987x, 1:1000) and FOSL2 (Santa Cruz, sc-604x, 1:1000).

5.5 Proliferation and colony formation assays

For proliferation and soft agar colony formation assays, cells were transduced with two shRNA per target with the exception of FOSL2 for which only one shRNA was efficiently knocking down all the isoforms (**Supplementary Figure S3**). Briefly, cells infected with lentiviral particles were selected for 24 hours prior to seeding. For proliferation assays, cell counts were measured using the Cytation 5 cell Imaging Multi-Mode Reader (BioTek Instruments) or the xCELLigence Real-Time Cell Analyzer (RTCA) instrument (ESBE Scientific and ACEA Biosciences) up to 8 days following seeding. Soft agar colony formation assays were performed as previously described¹⁸². Cells were maintained for 13 days (A549 and HEPG2) or 28 days (MCF7). Colonies were quantified using the ImageJ 1.46r software (NIH).

Acknowledgments

We thank Samer Hussein and Anne-Marie Pulichino for critical review of the manuscript and the laboratories of Dimcho Bachvarov and Jean-Yves Masson for easing access to the xCELLigence Real-Time Cell Analyzer (RTCA) and the Cytation 5. In addition, we would like to thank Éric Paquet, Astride-Louise Deschênes and Audrey Lemaçon for insightful discussions regarding the statistical and bioinformatic approaches. This work was supported by funds to S.B from the Canada Research Chair in Transcriptional Genomics, the Louisa Gale Scholar of the Canadian Cancer Society (grant #2013-701963) and from the Natural Sciences and Engineering Research Council of Canada (grant #436266-2013).

Author contributions statement

M.F, G.B, M.C.C, V.C and S.B conceived and performed the experiments. F.C.L, E.F, C.J.B, A.D and S.B conceived and performed the biocomputational analysis. S.G provided shRNA plasmids. S.B wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Accession codes

The data sets supporting the results of this article are available in the NCBI's Gene Expression Omnibus repository and are accessible through GEO Series accession number GSE76893 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=qrspsmoohdytnct&acc=GSE76893>)

Competing financial interests

The authors declare that they have no competing financial interests.

6. Supplementary Information

6.1 Supplementary Tables

Supplementary Table S1 – KEGG pathways enriched with genes occupied by MED1, SMC1A and NIPBL in MCF7, HEPG2 and A549 cells.

	-log10 (FDR corrected p-value)		
	HEPG2	MCF7	A549
hsa00970 Aminoacyl-tRNA biosynthesis	0.06448989	0.1352676	0.0173519
hsa02010 ABC transporters	0.88476264	0.45408152	1.80907079
hsa03008 Ribosome biogenesis in eukaryotes	0.0366085	0.08763324	0.00549542
hsa03010 Ribosome	0.0081719	0.03666085	0.00526644
hsa03013 RNA transport	0.0037942	0.01820138	0.01809528
hsa03015 mRNA surveillance pathway	0.13116847	0.08333723	0.03204529
hsa03018 RNA degradation	0.04760122	0.29100893	0.36194062
hsa03020 RNA polymerase	0.18573026	0.26993475	0.0914896
hsa03022 Basal transcription factors	0.31232912	0.45408152	0.04760122
hsa03030 DNA replication	0.16345299	0.24263714	0.23081073
hsa03040 Spliceosome	0.04760122	0.03778088	0.00549542
hsa03050 Proteasome	0.31428034	0.20452893	0.04760122
hsa03060 Protein export	0.24263714	0.31428034	0.65354423
hsa03320 PPAR signaling pathway	3.37498227	0.31428034	0.294212
hsa03410 Base excision repair	0.18047004	0.26152946	0.08690141
hsa03420 Nucleotide excision repair	0.30530343	0.44019414	0.30530343
hsa03430 Mismatch repair	0.24263714	0.31428034	0.35364118
hsa03440 Homologous recombination	0.45408152	0.28263629	0.29100893
hsa03450 Non-homologous end-joining	0.33263641	0.42695714	0.24837149
hsa03460 Fanconi anemia pathway	0.27876652	0.72059919	0.27109174
hsa04010 MAPK signaling pathway	1.03566082	0.70771226	1.13728408
hsa04012 ErbB signaling pathway	0.43715382	0.45121207	1.05343614
hsa04014 Ras signaling pathway	0.84592124	0.31428034	1.08501418
hsa04015 Rap1 signaling pathway	0.44925419	1.34726651	2.23139097
hsa04020 Calcium signaling pathway	0.79585701	0.45121207	1.07781407
hsa04022 cGMP-PKG signaling pathway	1.08501418	1.20476898	0.43052127
hsa04024 cAMP signaling pathway	0.36194062	0.98704607	0.74617128
hsa04060 Cytokine-cytokine receptor interaction	0.79898036	0.24837149	1.80907079
hsa04062 Chemokine signaling pathway	1.3041869	0.8281362	0.32561133
hsa04064 NF-kappa B signaling pathway	0.85891916	0.08333723	0.78294821
hsa04066 HIF-1 signaling pathway	1.90543498	0.61199009	1.80907079
hsa04068 FoxO signaling pathway	1.383675	0.44053573	1.20699442
hsa04070 Phosphatidylinositol signaling system	0.80134283	0.9119735	0.51172707
hsa04071 Sphingolipid signaling pathway	0.42747802	0.5101745	0.16615426
hsa04080 Neuroactive ligand-receptor interaction	0.05611979	0.62106254	0.38551948
hsa04110 Cell cycle	0.40622927	0.30530343	0.32608645
hsa04114 Oocyte meiosis	0.19172148	0.55697608	0.29100893
hsa04115 p53 signaling pathway	0.36194062	1.20699442	0.87162763
hsa04120 Ubiquitin mediated proteolysis	0.23360708	0.27899562	0.28263629
hsa04122 Sulfur relay system	0.88476264	0.46800068	0.29100893
hsa04130 SNARE interactions in vesicular transport	0.17428611	0.25265786	0.70791776
hsa04140 Regulation of autophagy	0.14722851	0.22553797	0.0628508
hsa04141 Protein processing in endoplasmic reticulum	0.35364118	0.32091736	0.31428034
hsa04142 Lysosome	0.28263629	0.16615426	0.03301846
hsa04144 Endocytosis	0.53353389	0.36194062	0.83877231
hsa04145 Phagosome	0.42414457	0.23500199	0.21447197
hsa04146 Peroxisome	1.20699442	0.09783438	0.11910916
hsa04150 mTOR signaling pathway	0.67680873	0.35714788	0.54287035
hsa04151 PI3K-Akt signaling pathway	0.36194062	0.56704944	1.64368251
hsa04152 AMPK signaling pathway	0.59038668	0.30530343	0.96434529
hsa04210 Apoptosis	1.20537507	0.27109174	0.84629036
hsa04260 Cardiac muscle contraction	0.16779307	0.80197349	0.36194062
hsa04261 Adrenergic signaling in cardiomyocytes	0.30727917	0.82766809	0.31428034
hsa04270 Vascular smooth muscle contraction	0.83098243	1.30206728	0.82766809
hsa04310 Wnt signaling pathway	1.30206728	0.27109174	1.81815184
hsa04320 Dorsal-ventral axis formation	1.80907079	0.65928509	1.20699442
hsa04330 Notch signaling pathway	0.2970733	0.43420076	0.72059919
hsa04340 Hedgehog signaling pathway	0.28428077	0.18261675	0.44019414
hsa04350 TGF-beta signaling pathway	0.69026265	1.05913914	2.93344883
hsa04360 Axon guidance	0.7819282	0.70924373	1.30206728
hsa04370 VEGF signaling pathway	0.4160026	0.35364118	0.75842459
hsa04380 Osteoclast differentiation	1.20476898	0.93889294	1.07781407
hsa04390 Hippo signaling pathway	1.20537507	1.62576397	3.25432318
hsa04510 Focal adhesion	0.8281362	0.75705689	3.06722551
hsa04512 ECM-receptor interaction	0.15462351	1.07781407	1.09981997
hsa04514 Cell adhesion molecules (CAMs)	0.32374736	0.4160026	0.45408152
hsa04520 Adherens junction	1.08501418	0.30530343	3.53315075

hsa04530	Tight junction	0.12782002	0.27876652	0.98704607
hsa04540	Gap junction	0.65354423	0.44376705	0.30530343
hsa04550	Signaling pathways regulating pluripotency c	2.40080884	0.85725599	3.25432318
hsa04610	Complement and coagulation cascades	2.42901562	0.13116847	0.64253303
hsa04611	Platelet activation	1.05343614	0.75193439	0.33844188
hsa04612	Antigen processing and presentation	0.04760122	0.11215822	0.00927162
hsa04614	Renin-angiotensin system	0.95348518	0.31428034	0.65354423
hsa04620	Toll-like receptor signaling pathway	0.20675133	0.36194062	0.21993112
hsa04621	NOD-like receptor signaling pathway	0.44053573	0.16118378	0.24009169
hsa04622	RIG-I-like receptor signaling pathway	0.35338791	0.31279641	0.06634004
hsa04623	Cytosolic DNA-sensing pathway	0.21993112	0.13839477	0.19890669
hsa04630	Jak-STAT signaling pathway	0.75193439	0.35516932	1.30206728
hsa04640	Hematopoietic cell lineage	0.43715382	0.26349006	0.20093494
hsa04650	Natural killer cell mediated cytotoxicity	0.35714788	0.28428077	0.02137295
hsa04660	T cell receptor signaling pathway	0.75193439	0.86653541	0.1352316
hsa04662	B cell receptor signaling pathway	0.82432591	0.84880426	0.28263629
hsa04664	Fc epsilon RI signaling pathway	0.60457038	0.59038668	0.66590546
hsa04666	Fc gamma R-mediated phagocytosis	1.09912109	0.98308861	0.97054854
hsa04668	TNF signaling pathway	0.9119735	0.56704944	2.38790137
hsa04670	Leukocyte transendothelial migration	0.43613898	0.31596355	1.20699442
hsa04672	Intestinal immune network for IgA productio	0.30530343	0.19464349	0.04385493
hsa04710	Circadian rhythm	0.78294821	0.27356757	0.46967574
hsa04713	Circadian entrainment	0.38551948	0.95348518	0.37807177
hsa04720	Long-term potentiation	0.61540791	0.32561133	0.45582565
hsa04721	Synaptic vesicle cycle	0.06959479	0.34168589	0.32608645
hsa04722	Neurotrophin signaling pathway	0.61894828	0.31428034	0.16615426
hsa04723	Retrograde endocannabinoid signaling	0.22541751	0.62681662	0.34331287
hsa04724	Glutamatergic synapse	0.87162763	1.33573973	0.53683756
hsa04725	Cholinergic synapse	0.46505437	0.82766809	0.11477478
hsa04726	Serotonergic synapse	0.19172148	0.19172148	0.40123762
hsa04727	GABAergic synapse	0.43420076	1.02696491	0.43676602
hsa04728	Dopaminergic synapse	0.3741105	0.29235464	0.13534378
hsa04730	Long-term depression	0.42414457	0.35714788	0.77539653
hsa04740	Olfactory transduction	0.00064792	0.00927162	0.00193814
hsa04742	Taste transduction	0.10117202	0.40667886	0.27876652
hsa04744	Phototransduction	0.20452893	0.28489626	0.29841977
hsa04750	Inflammatory mediator regulation of TRP ch	1.05343614	1.20699442	3.2145199
hsa04810	Regulation of actin cytoskeleton	0.59038668	1.09981997	0.76503832
hsa04910	Insulin signaling pathway	1.09404656	0.1352316	0.18573026
hsa04911	Insulin secretion	0.28489626	1.05343614	0.65354423
hsa04912	GnRH signaling pathway	0.4160026	0.69026265	0.78294821
hsa04913	Ovarian steroidogenesis	0.28666618	0.75842459	1.55025317
hsa04914	Progesterone-mediated oocyte maturation	0.28263629	1.03566082	0.30727917
hsa04915	Estrogen signaling pathway	0.56307329	1.20537507	0.06773798
hsa04916	Melanogenesis	0.36194062	0.3848011	0.85725599
hsa04917	Prolactin signaling pathway	0.82432591	1.20476898	0.60457038
hsa04918	Thyroid hormone synthesis	0.56704944	0.5598669	0.28489626
hsa04919	Thyroid hormone signaling pathway	0.84592124	1.05343614	1.04486087
hsa04920	Adipocytokine signaling pathway	1.43121608	0.56704944	1.08501418
hsa04921	Oxytocin signaling pathway	0.28428077	1.56271724	0.78294821
hsa04922	Glucagon signaling pathway	0.54570427	0.37894248	0.34331287
hsa04960	Aldosterone-regulated sodium reabsorption	1.31857664	0.88476264	0.88476264
hsa04961	Endocrine and other factor-regulated calciu	0.30530343	0.78606864	0.30530343
hsa04962	Vasopressin-regulated water reabsorption	0.13116847	0.45408152	0.04760122
hsa04964	Proximal tubule bicarbonate reclamation	0.54570427	0.31428034	0.35364118
hsa04966	Collecting duct acid secretion	0.211747	0.65928509	0.12177361
hsa04970	Salivary secretion	0.42747802	0.44019414	0.43052127
hsa04971	Gastric acid secretion	0.5394339	0.83485935	0.58187188
hsa04972	Pancreatic secretion	0.59038668	0.6552596	0.53485508
hsa04973	Carbohydrate digestion and absorption	0.87162763	0.45408152	0.04760122
hsa04974	Protein digestion and absorption	0.64253303	1.30206728	0.99575064
hsa04975	Fat digestion and absorption	1.80907079	0.21775761	0.35364118
hsa04976	Bile secretion	0.56704944	1.20537507	0.83098243
hsa04977	Vitamin digestion and absorption	1.31857664	0.30610017	0.96758561
hsa04978	Mineral absorption	0.28428077	0.4160026	0.28263629
hsa00010	Glycolysis / Gluconeogenesis	0.36194062	0.1346108	0.45408152
hsa00020	Citrate cycle (TCA cycle)	0.19464349	0.27876652	0.28428077
hsa00030	Pentose phosphate pathway	0.45408152	1.07781407	0.82766809
hsa00040	Pentose and glucuronate interconversions	0.37387207	0.54287035	1.68440823
hsa00051	Fructose and mannose metabolism	0.42414457	0.26993475	0.27094648
hsa00052	Galactose metabolism	0.44053573	0.27876652	0.49755095
hsa00053	Ascorbate and aldarate metabolism	0.46924688	0.29100893	1.64706179

hsa00061	Fatty acid biosynthesis	2.964536	0.42695714	0.24837149
hsa00062	Fatty acid elongation	0.22553797	0.30530343	0.1346108
hsa00071	Fatty acid degradation	2.26999488	0.20452893	0.18047004
hsa00072	Synthesis and degradation of ketone bodies	0.37988252	0.46800068	0.29100893
hsa00100	Steroid biosynthesis	1.04486087	1.28155332	0.40123762
hsa00120	Primary bile acid biosynthesis	1.10808662	0.36194062	0.19890669
hsa00130	Ubiquinone and other terpenoid-quinone biosynthesis	0.84696662	0.45408152	0.64253303
hsa00140	Steroid hormone biosynthesis	0.43585987	0.15738486	1.30206728
hsa00190	Oxidative phosphorylation	0.13534378	0.28489626	0.00549542
hsa00230	Purine metabolism	0.13551339	0.30727917	0.63611076
hsa00232	Caffeine metabolism	0.53800258	0.64253303	0.98704607
hsa00240	Pyrimidine metabolism	0.33836256	0.20452893	0.22337659
hsa00250	Alanine, aspartate and glutamate metabolism	0.70771226	0.24837149	0.69026265
hsa00260	Glycine, serine and threonine metabolism	0.14308613	0.22288598	0.20093494
hsa00270	Cysteine and methionine metabolism	0.88476264	0.20452893	0.32160523
hsa00280	Valine, leucine and isoleucine degradation	0.30530343	0.19464349	0.04385493
hsa00290	Valine, leucine and isoleucine biosynthesis	0.59038668	0.69026265	0.45408152
hsa00300	Lysine biosynthesis	0.75184517	0.83877231	0.62315874
hsa00310	Lysine degradation	0.28263629	0.1798849	0.27876652
hsa00330	Arginine and proline metabolism	0.28666618	0.18573026	0.68827865
hsa00340	Histidine metabolism	0.95348518	0.31428034	0.65354423
hsa00350	Tyrosine metabolism	0.70771226	0.24837149	0.69026265
hsa00360	Phenylalanine metabolism	0.67219723	0.36194062	1.20699442
hsa00380	Tryptophan metabolism	0.96868685	0.22288598	0.36194062
hsa00400	Phenylalanine, tyrosine and tryptophan biosynthesis	1.20537507	0.64253303	0.98704607
hsa00410	beta-Alanine metabolism	0.43420076	0.27356757	0.78294821
hsa00430	Taurine and hypotaurine metabolism	0.36194062	0.45408152	0.28056609
hsa00450	Selenocompound metabolism	0.29100893	0.36194062	0.45408152
hsa00460	Cyanocinnamic acid metabolism	0.45408152	0.56294305	0.34957002
hsa00471	D-Glutamine and D-glutamate metabolism	0.59038668	0.69026265	1.08501418
hsa00472	D-Arginine and D-ornithine metabolism	0.88476264	0.98894236	0.78294821
hsa00480	Glutathione metabolism	0.28428077	0.18261675	0.03666085
hsa00500	Starch and sucrose metabolism	0.72726789	0.16436535	1.08501418
hsa00510	N-Glycan biosynthesis	0.82766809	0.18889409	0.45408152
hsa00511	Other glycan degradation	1.08501418	0.35714788	0.19172148
hsa00512	Mucin type O-Glycan biosynthesis	1.14845944	0.60457038	1.43192959
hsa00514	Other types of O-glycan biosynthesis	0.43420076	0.27356757	0.78294821
hsa00520	Amino sugar and nucleotide sugar metabolism	0.83099427	0.19172148	0.29754233
hsa00524	Butirosin and neomycin biosynthesis	1.20537507	0.64253303	0.42012315
hsa00531	Glycosaminoglycan degradation	0.62770512	0.34719976	0.18261675
hsa00532	Glycosaminoglycan biosynthesis - chondroitin-6-sulfate	0.60457038	0.78512292	0.40123762
hsa00533	Glycosaminoglycan biosynthesis - keratan sulfate	0.72531309	0.8850649	0.50629835
hsa00534	Glycosaminoglycan biosynthesis - heparan sulfate	2.74847876	1.19069333	0.33836256
hsa00561	Glycerolipid metabolism	0.98704607	0.15462351	0.55697608
hsa00562	Inositol phosphate metabolism	0.56704944	0.85891916	0.42608461
hsa00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	0.22553797	0.30530343	0.1346108
hsa00564	Glycerophospholipid metabolism	1.08501418	0.23360708	0.54287035
hsa00565	Ether lipid metabolism	0.87162763	0.20093494	0.31428034
hsa00590	Arachidonic acid metabolism	0.22738515	0.14479506	0.74617128
hsa00591	Linoleic acid metabolism	0.45408152	0.28263629	0.52086844
hsa00592	alpha-Linolenic acid metabolism	0.51172707	0.30530343	0.60405584
hsa00600	Sphingolipid metabolism	0.84592124	0.19464349	0.04385493
hsa00601	Glycosphingolipid biosynthesis - lactosylceramide	1.28155332	0.29539896	0.31428034
hsa00603	Glycosphingolipid biosynthesis - globoseries	1.87654726	2.42901562	0.23360708
hsa00604	Glycosphingolipid biosynthesis - ganglioseries	1.20537507	0.8850649	0.50629835
hsa00620	Pyruvate metabolism	0.34468011	0.22288598	0.36194062
hsa00630	Glyoxylate and dicarboxylate metabolism	0.45408152	0.28489626	0.11580111
hsa00640	Propanoate metabolism	0.18573026	0.26993475	0.45408152
hsa00650	Butanoate metabolism	0.211747	0.29100893	0.12177361
hsa00670	One carbon pool by folate	0.27110033	0.33747584	0.40123762
hsa00730	Thiamine metabolism	0.59038668	0.69026265	0.45408152
hsa00740	Riboflavin metabolism	0.59038668	0.69026265	1.08501418
hsa00750	Vitamin B6 metabolism	0.48561659	0.5958386	1.51492444
hsa00760	Nicotinate and nicotinamide metabolism	0.82404684	1.07781407	0.52086844
hsa00770	Pantothenate and CoA biosynthesis	1.08501418	0.35714788	0.19172148
hsa00780	Biotin metabolism	0.65354423	0.75842459	1.20699442
hsa00785	Lipoic acid metabolism	0.65354423	0.75842459	0.53353389
hsa00790	Folate biosynthesis	0.31596355	0.4106878	0.23360708
hsa00830	Retinol metabolism	0.06634004	0.13551339	1.43192959
hsa00860	Porphyrin and chlorophyll metabolism	0.13534378	0.21206425	1.43121608
hsa00900	Terpenoid backbone biosynthesis	0.24931982	0.31596355	0.15462351
hsa00910	Nitrogen metabolism	0.67219723	0.36194062	0.19890669

hsa00920 Sulfur metabolism	0.88476264	0.46800068	0.67509174
hsa00980 Metabolism of xenobiotics by cytochrome P450	0.32608645	0.118757	1.02696491
hsa00982 Drug metabolism - cytochrome P450	0.20219592	0.1324597	0.87162763
hsa00983 Drug metabolism - other enzymes	0.30565481	0.19776214	1.30206728
hsa01040 Biosynthesis of unsaturated fatty acids	0.24263714	0.31428034	0.14482444
hsa01100 Metabolic pathways	1.80907079	0.04760122	0.64634045
hsa01200 Carbon metabolism	0.30727917	0.33747584	0.54287035
hsa01210 2-Oxocarboxylic acid metabolism	0.67219723	0.36194062	0.45408152
hsa01212 Fatty acid metabolism	1.55025317	0.19172148	0.04185459
hsa01220 Degradation of aromatic compounds	0.65354423	0.75842459	0.53353389
hsa01230 Biosynthesis of amino acids	0.5394339	0.53800258	0.58187188

Supplementary Table S1: KEGG pathways enriched with genes occupied by MED1, SMC1A and NIPBL in MCF7, HEPG2 and A549 cells.

Supplementary Table S2 - List of ChIP-Seq datasets used in the study (MCF7 cells)

Biosample_name	Target	Experiment	Accession
MCF-7	Control	Data will be provided through GEO upon publication	
MCF-7	NIPBL	Data will be provided through GEO upon publication	
MCF-7	MED1	Data will be provided through GEO upon publication	
MCF-7	SMC1A	Data will be provided through GEO upon publication	
MCF-7	Era	Data will be provided through GEO upon publication	
MCF-7	CEBPB	ENCSR000BSR	ENCFF000QLL
MCF-7	CEBPB	ENCSR000BSR	ENCFF000QLO
MCF-7	Control	ENCSR000EWW	ENCFF000VHM
MCF-7	Control	ENCSR000EWW	ENCFF000VHL
MCF-7	Control	ENCSR000DWI	ENCFF001HUM
MCF-7	Control	ENCSR000DMW	ENCFF000SAZ
MCF-7	Control	ENCSR000AHE	ENCFF000QQJ
MCF-7	Control	ENCSR000AHE	ENCFF000QQG
MCF-7	Control	ENCSR000AHE	ENCFF000QQI
MCF-7	Control	ENCSR000AHE	ENCFF000QKQ
MCF-7	CTCF	ENCSR000DWH	ENCFF001HUF
MCF-7	CTCF	ENCSR000DWH	ENCFF001HUG
MCF-7	CTCF	ENCSR000DMV	ENCFF000RZX
MCF-7	CTCF	ENCSR000DMV	ENCFF000RZY
MCF-7	CTCF	ENCSR000DMS	ENCFF000SAB
MCF-7	CTCF	ENCSR000DMS	ENCFF000SAD
MCF-7	CTCF	ENCSR000DMR	ENCFF000SAY
MCF-7	CTCF	ENCSR000DMR	ENCFF000SAX
MCF-7	CTCF	ENCSR000DMO	ENCFF000SAL
MCF-7	CTCF	ENCSR000DMO	ENCFF000SAK
MCF-7	CTCF	ENCSR000DML	ENCFF000SAP
MCF-7	CTCF	ENCSR000DML	ENCFF000SAO
MCF-7	CTCF	ENCSR000AHD	ENCFF000QLW
MCF-7	CTCF	ENCSR000AHD	ENCFF000QLT
MCF-7	EGR1	ENCSR000BUX	ENCFF000QMF
MCF-7	EGR1	ENCSR000BUX	ENCFF000QMC
MCF-7	ELF1	ENCSR000BSS	ENCFF000QMJ
MCF-7	ELF1	ENCSR000BSS	ENCFF000QMK
MCF-7	EP300	ENCSR000BTR	ENCFF000QPK
MCF-7	EP300	ENCSR000BTR	ENCFF000QPP
MCF-7	FOSL2	ENCSR000BUI	ENCFF000QMU
MCF-7	FOSL2	ENCSR000BUI	ENCFF000QMS
MCF-7	FOXM1	ENCSR000BUJ	ENCFF000QMZ
MCF-7	FOXM1	ENCSR000BUJ	ENCFF000QNC
MCF-7	GABPA	ENCSR000BUK	ENCFF000QNK
MCF-7	GABPA	ENCSR000BUK	ENCFF000QNJ
MCF-7	GATA3	ENCSR000EWW	ENCFF000ZKL
MCF-7	GATA3	ENCSR000EWW	ENCFF000ZKJ
MCF-7	GATA3	ENCSR000EWS	ENCFF000ZKV
MCF-7	GATA3	ENCSR000EWS	ENCFF000ZKT
MCF-7	GATA3	ENCSR000EWS	ENCFF000ZLE
MCF-7	GATA3	ENCSR000BST	ENCFF000QNS
MCF-7	GATA3	ENCSR000BST	ENCFF000QNP
MCF-7	H3K27ac	ENCSR000EWR	ENCFF000VHD

MCF-7	H3K27ac	ENCSR000EWR	ENCF000VHC
MCF-7	H3K27me3	ENCSR000EWP	ENCF000VFU
MCF-7	H3K27me3	ENCSR000EWP	ENCF000VFW
MCF-7	H3K27me3	ENCSR000EWP	ENCF000VGE
MCF-7	H3K36me3	ENCSR000EWO	ENCF000VGG
MCF-7	H3K36me3	ENCSR000EWO	ENCF000VGI
MCF-7	H3K4me3	ENCSR000DWJ	ENCF001FYK
MCF-7	H3K4me3	ENCSR000DWJ	ENCF001FYE
MCF-7	H3K9me3	ENCSR000EWQ	ENCF000VFG
MCF-7	H3K9me3	ENCSR000EWQ	ENCF000VFJ
MCF-7	H3K9me3	ENCSR000EWQ	ENCF000VFE
MCF-7	HA-E2F1	ENCSR000EWX	ENCF000ZLD
MCF-7	HA-E2F1	ENCSR000EWX	ENCF000ZLV
MCF-7	HDAC2	ENCSR000BTP	ENCF000QNY
MCF-7	HDAC2	ENCSR000BTP	ENCF000QQA
MCF-7	JUND	ENCSR000BSU	ENCF000QOD
MCF-7	JUND	ENCSR000BSU	ENCF000QOB
MCF-7	MAX	ENCSR000BUL	ENCF000QON
MCF-7	MAX	ENCSR000BUL	ENCF000QOS
MCF-7	MYC	ENCSR000DMQ	ENCF000RZT
MCF-7	MYC	ENCSR000DMQ	ENCF000RZV
MCF-7	MYC	ENCSR000DMP	ENCF000RYO
MCF-7	MYC	ENCSR000DMP	ENCF000RYP
MCF-7	MYC	ENCSR000DMM	ENCF000RZE
MCF-7	MYC	ENCSR000DMM	ENCF000RZC
MCF-7	MYC	ENCSR000DMJ	ENCF000RZL
MCF-7	MYC	ENCSR000DMJ	ENCF000RZJ
MCF-7	NR2F2	ENCSR000BUY	ENCF000QOY
MCF-7	NR2F2	ENCSR000BUY	ENCF000QOV
MCF-7	PML	ENCSR000BUZ	ENCF000QPW
MCF-7	PML	ENCSR000BUZ	ENCF000QPV
MCF-7	POLR2A	ENCSR000DMT	ENCF000SBK
MCF-7	POLR2A	ENCSR000DMT	ENCF000SBI
MCF-7	POLR2A	ENCSR000DMN	ENCF000SBP
MCF-7	POLR2A	ENCSR000DMN	ENCF000SBO
MCF-7	POLR2A	ENCSR000DMK	ENCF000SBV
MCF-7	POLR2A	ENCSR000DMK	ENCF000SBY
MCF-7	RAD21	ENCSR000BTQ	ENCF000QPX
MCF-7	RAD21	ENCSR000BTQ	ENCF000QQA
MCF-7	REST	ENCSR000BSP	ENCF000QPH
MCF-7	REST	ENCSR000BSP	ENCF000QPE
MCF-7	SIN3A	ENCSR000BUM	ENCF000QQX
MCF-7	SIN3A	ENCSR000BUM	ENCF000QQU
MCF-7	SRF	ENCSR000BVA	ENCF000QRE
MCF-7	SRF	ENCSR000BVA	ENCF000QRH
MCF-7	TAF1	ENCSR000AHF	ENCF000QRS
MCF-7	TAF1	ENCSR000AHF	ENCF000QRL
MCF-7	TCF12	ENCSR000BUN	ENCF000QRV
MCF-7	TCF12	ENCSR000BUN	ENCF000QRU
MCF-7	TCF7L2	ENCSR000EWT	ENCF000ZLK
MCF-7	TCF7L2	ENCSR000EWT	ENCF000ZLM

MCF-7	TEAD4	ENCSR000BUO	ENCF000QSI
MCF-7	TEAD4	ENCSR000BUO	ENCF000QSB
MCF-7	ZNF217	ENCSR000EWU	ENCF000ZLW
MCF-7	ZNF217	ENCSR000EWU	ENCF000ZLT
MCF-7	FOXA1	SRP044607/SRX747793	SRR1635461
MCF-7	FOXA1	SRP044607/SRX747794	SRR1635462
MCF-7	Control	SRP044607	SRR1635437
MCF-7	Control	SRP044607	SRR1635438
MCF-7	H3K4me1	SRX153145	SRR507786
MCF-7	H3K4me1	SRX153146	SRR507787
MCF-7	H3K4me1	SRX153147	SRR507788

Supplementary Table S2: List of ChIP-Seq datasets used in the study (MCF7).

Supplementary Table S2 - List of ChIP-Seq datasets used in the study (HEPG2 cells)

Biosample_name	Target	Experiment	Accession
HEPG2	Control	Data will be provided through GEO upon publication	
HEPG2	NIPBL	Data will be provided through GEO upon publication	
HEPG2	MED1	Data will be provided through GEO upon publication	
HEPG2	SMC1A	Data will be provided through GEO upon publication	
HepG2	ARID3A	ENCSR000EDP	ENCFF000XOU
HepG2	ARID3A	ENCSR000EDP	ENCFF000XOS
HepG2	ATF3	ENCSR000BKE	ENCFF000PFR
HepG2	ATF3	ENCSR000BKE	ENCFF000PFQ
HepG2	BHLHE40	ENCSR000EDT	ENCFF000XPC
HepG2	BHLHE40	ENCSR000EDT	ENCFF000XPA
HepG2	BHLHE40	ENCSR000BID	ENCFF000PFY
HepG2	BHLHE40	ENCSR000BID	ENCFF000PGB
HepG2	BRCA1	ENCSR000EDY	ENCFF000XPS
HepG2	BRCA1	ENCSR000EDY	ENCFF000XPL
HepG2	CEBPB	ENCSR000EEX	ENCFF000XPT
HepG2	CEBPB	ENCSR000EEX	ENCFF000XPR
HepG2	CEBPB	ENCSR000EEE	ENCFF000XQN
HepG2	CEBPB	ENCSR000EEE	ENCFF000XQM
HepG2	CEBPB	ENCSR000BQI	ENCFF000PGI
HepG2	CEBPB	ENCSR000BQI	ENCFF000PGG
HepG2	CEBPD	ENCSR000BQJ	ENCFF000PGK
HepG2	CEBPD	ENCSR000BQJ	ENCFF000PGQ
HepG2	CEBPZ	ENCSR000EDO	ENCFF000XPZ
HepG2	CEBPZ	ENCSR000EDO	ENCFF000XQB
HepG2	CHD2	ENCSR000EED	ENCFF000XQQ
HepG2	CHD2	ENCSR000EED	ENCFF000XQP
HepG2	Control	ENCSR000DUC	ENCFF001HOD
HepG2	Control	ENCSR000EES	ENCFF000XQR
HepG2	Control	ENCSR000EES	ENCFF000XQS
HepG2	Control	ENCSR000EER	ENCFF000XRI
HepG2	Control	ENCSR000EER	ENCFF000XRH
HepG2	Control	ENCSR000EVT	ENCFF000XTE
HepG2	Control	ENCSR000AME	ENCFF000BDZ
HepG2	Control	ENCSR000AME	ENCFF000BDT
HepG2	Control	ENCSR000DLT	ENCFF000RUP
HepG2	Control	ENCSR000EZR	ENCFF000XTB
HepG2	Control	ENCSR000EZR	ENCFF000XSV
HepG2	Control	ENCSR000EZR	ENCFF000XSY
HepG2	Control	ENCSR000BLN	ENCFF000POH
HepG2	Control	ENCSR000BLN	ENCFF000POC
HepG2	Control	ENCSR000BPH	ENCFF000PPI
HepG2	Control	ENCSR000BPH	ENCFF000PPF
HepG2	Control	ENCSR000BPH	ENCFF000PPC
HepG2	Control	ENCSR000BLH	ENCFF000POO
HepG2	Control	ENCSR000BLH	ENCFF000POQ
HepG2	Control	ENCSR000BLG	ENCFF000POU
HepG2	Control	ENCSR000BLG	ENCFF000POV
HepG2	CREB1	ENCSR000BVL	ENCFF000PGX
HepG2	CREB1	ENCSR000BVL	ENCFF000PGZ
HepG2	CTCF	ENCSR000DUG	ENCFF001HNU

HepG2	CTCF	ENCSR000DUG	ENCF001HNT
HepG2	CTCF	ENCSR000AMA	ENCF000BEI
HepG2	CTCF	ENCSR000AMA	ENCF000BED
HepG2	CTCF	ENCSR000BIE	ENCF000PHG
HepG2	CTCF	ENCSR000BIE	ENCF000PHE
HepG2	CTCF	ENCSR000DLS	ENCF000RUI
HepG2	CTCF	ENCSR000DLS	ENCF000RUJ
HepG2	ELF1	ENCSR000BMZ	ENCF000PHP
HepG2	ELF1	ENCSR000BMZ	ENCF000PHM
HepG2	EP300	ENCSR000EDV	ENCF000XVS
HepG2	EP300	ENCSR000EDV	ENCF000XVT
HepG2	EP300	ENCSR000BLW	ENCF000PMZ
HepG2	EP300	ENCSR000BLW	ENCF000PNA
HepG2	ESRRA	ENCSR000EEW	ENCF000XRN
HepG2	ESRRA	ENCSR000EEW	ENCF000XRO
HepG2	EZH2	ENCSR000ARI	ENCF000BEL
HepG2	EZH2	ENCSR000ARI	ENCF000BEQ
HepG2	FOSL2	ENCSR000BHP	ENCF000PHU
HepG2	FOSL2	ENCSR000BHP	ENCF000PIE
HepG2	FOXA1	ENCSR000BMO	ENCF000PIB
HepG2	FOXA1	ENCSR000BMO	ENCF000PHY
HepG2	FOXA1	ENCSR000BLE	ENCF000PIM
HepG2	FOXA1	ENCSR000BLE	ENCF000PIJ
HepG2	FOXA2	ENCSR000BNI	ENCF000PIX
HepG2	FOXA2	ENCSR000BNI	ENCF000PIT
HepG2	GABPA	ENCSR000BJK	ENCF000PJF
HepG2	GABPA	ENCSR000BJK	ENCF000PJC
HepG2	oat-IgG-contr	ENCSR000EDR	ENCF000XSE
HepG2	oat-IgG-contr	ENCSR000EDR	ENCF000XSG
HepG2	H2AFZ	ENCSR000AOK	ENCF000BEP
HepG2	H2AFZ	ENCSR000AOK	ENCF000BES
HepG2	H3K27ac	ENCSR000AMO	ENCF000BFD
HepG2	H3K27ac	ENCSR000AMO	ENCF000BFH
HepG2	H3K27me3	ENCSR000DUE	ENCF001FLT
HepG2	H3K27me3	ENCSR000DUE	ENCF001FLQ
HepG2	H3K27me3	ENCSR000AOL	ENCF000BFR
HepG2	H3K27me3	ENCSR000AOL	ENCF000BFU
HepG2	H3K36me3	ENCSR000DUD	ENCF001FMA
HepG2	H3K36me3	ENCSR000DUD	ENCF001FMC
HepG2	H3K36me3	ENCSR000AMB	ENCF000BFZ
HepG2	H3K36me3	ENCSR000AMB	ENCF000BFT
HepG2	H3K4me1	ENCSR000APV	ENCF000BEY
HepG2	H3K4me1	ENCSR000APV	ENCF000BEX
HepG2	H3K4me2	ENCSR000AMC	ENCF000BGD
HepG2	H3K4me2	ENCSR000AMC	ENCF000BGA
HepG2	H3K4me3	ENCSR000DJF	ENCF001FMH
HepG2	H3K4me3	ENCSR000DJF	ENCF001FMG
HepG2	H3K4me3	ENCSR000AMP	ENCF000BGF
HepG2	H3K4me3	ENCSR000AMP	ENCF000BGE
HepG2	H3K79me2	ENCSR000AOM	ENCF000BGR
HepG2	H3K79me2	ENCSR000AOM	ENCF000BGW
HepG2	H3K9ac	ENCSR000AMD	ENCF000BGQ
HepG2	H3K9ac	ENCSR000AMD	ENCF000BGS

HepG2	H3K9me3	ENCSR000ATD	ENCF000BFB
HepG2	H3K9me3	ENCSR000ATD	ENCF000BFI
HepG2	H4K20me1	ENCSR000AMQ	ENCF000BGX
HepG2	H4K20me1	ENCSR000AMQ	ENCF000BHC
HepG2	HDAC2	ENCSR000BMC	ENCF000PJJ
HepG2	HDAC2	ENCSR000BMC	ENCF000PJD
HepG2	HNF4A	ENCSR000EEU	ENCF000XRT
HepG2	HNF4A	ENCSR000EEU	ENCF000XSM
HepG2	HNF4A	ENCSR000BLF	ENCF000PJT
HepG2	HNF4A	ENCSR000BLF	ENCF000PJX
HepG2	HNF4G	ENCSR000BNJ	ENCF000PKH
HepG2	HNF4G	ENCSR000BNJ	ENCF000PKJ
HepG2	HSF1	ENCSR000EET	ENCF000XSL
HepG2	HSF1	ENCSR000EET	ENCF000XSK
HepG2	IRF3	ENCSR000EEJ	ENCF000XTN
HepG2	IRF3	ENCSR000EEJ	ENCF000XTO
HepG2	JUN	ENCSR000EEK	ENCF000XQU
HepG2	JUN	ENCSR000EEK	ENCF000XQT
HepG2	JUND	ENCSR000EEI	ENCF000XTR
HepG2	JUND	ENCSR000EEI	ENCF000XTQ
HepG2	JUND	ENCSR000BGK	ENCF000PKR
HepG2	JUND	ENCSR000BGK	ENCF000PKK
HepG2	MAFF	ENCSR000EEC	ENCF000XTZ
HepG2	MAFF	ENCSR000EEC	ENCF000XUA
HepG2	MAFK	ENCSR000EEB	ENCF000XUJ
HepG2	MAFK	ENCSR000EEB	ENCF000XUL
HepG2	MAFK	ENCSR000EDZ	ENCF000XUU
HepG2	MAFK	ENCSR000EDZ	ENCF000XVK
HepG2	MAX	ENCSR000EDS	ENCF000XUH
HepG2	MAX	ENCSR000EDS	ENCF000XUF
HepG2	MAX	ENCSR000BTM	ENCF000PKZ
HepG2	MAX	ENCSR000BTM	ENCF000PKX
HepG2	MAZ	ENCSR000EDN	ENCF000XUP
HepG2	MAZ	ENCSR000EDN	ENCF000XUN
HepG2	MBD4	ENCSR000BQW	ENCF000PLF
HepG2	MBD4	ENCSR000BQW	ENCF000PLH
HepG2	MXI1	ENCSR000EDU	ENCF000XVA
HepG2	MXI1	ENCSR000EDU	ENCF000XUX
HepG2	MXI1	ENCSR000EDU	ENCF000XVC
HepG2	MYBL2	ENCSR000BRO	ENCF000PLP
HepG2	MYBL2	ENCSR000BRO	ENCF000PLM
HepG2	MYC	ENCSR000DLR	ENCF000RTV
HepG2	MYC	ENCSR000DLR	ENCF000RTX
HepG2	MYC	ENCSR000DLR	ENCF000RTY
HepG2	NFIC	ENCSR000BQX	ENCF000PMB
HepG2	NFIC	ENCSR000BQX	ENCF000PLW
HepG2	NR2C2	ENCSR000EVS	ENCF000XZK
HepG2	NR2C2	ENCSR000EVS	ENCF000XZJ
HepG2	NR2F2	ENCSR000BVM	ENCF000PMC
HepG2	NR2F2	ENCSR000BVM	ENCF000PMF
HepG2	NR3C1	ENCSR000EEV	ENCF000XRV
HepG2	NR3C1	ENCSR000EEV	ENCF000XRU
HepG2	NRF1	ENCSR000EEH	ENCF000XVP

HepG2	NRF1	ENCSR000EEH	ENCF000XVR
HepG2	POLR2A	ENCSR000EEM	ENCF000XWJ
HepG2	POLR2A	ENCSR000EEM	ENCF000XWR
HepG2	POLR2A	ENCSR000EEP	ENCF000XWF
HepG2	POLR2A	ENCSR000EEP	ENCF000XWI
HepG2	POLR2A	ENCSR000EZQ	ENCF000XWW
HepG2	POLR2A	ENCSR000EZQ	ENCF000XWT
HepG2	POLR2A	ENCSR000EZQ	ENCF000XWV
HepG2	POLR2A	ENCSR000DLQ	ENCF000RUT
HepG2	POLR2A	ENCSR000DLQ	ENCF000RUR
HepG2	POLR2A	ENCSR000DLQ	ENCF000RLUY
HepG2	POLR2A	ENCSR000BJM	ENCF000PNT
HepG2	POLR2A	ENCSR000BJM	ENCF000PNM
HepG2	LR2Aphosphc	ENCSR000EDX	ENCF000XXB
HepG2	LR2Aphosphc	ENCSR000EDX	ENCF000XXC
HepG2	LR2Aphosphc	ENCSR000BPI	ENCF000PNI
HepG2	LR2Aphosphc	ENCSR000BPI	ENCF000PNL
HepG2	PPARGC1A	ENCSR000EEQ	ENCF000XVQ
HepG2	PPARGC1A	ENCSR000EEQ	ENCF000XWG
HepG2	bbbit-IgG-contr	ENCSR000EEN	ENCF000XTF
HepG2	RAD21	ENCSR000EEG	ENCF000XXK
HepG2	RAD21	ENCSR000EEG	ENCF000XXL
HepG2	RAD21	ENCSR000BLS	ENCF000PNY
HepG2	RAD21	ENCSR000BLS	ENCF000PNW
HepG2	RCOR1	ENCSR000EDQ	ENCF000XQY
HepG2	RCOR1	ENCSR000EDQ	ENCF000XQW
HepG2	REST	ENCSR000BJL	ENCF000PMN
HepG2	REST	ENCSR000BJL	ENCF000PMK
HepG2	REST	ENCSR000BOT	ENCF000PMU
HepG2	REST	ENCSR000BOT	ENCF000PMV
HepG2	RFX5	ENCSR000EEA	ENCF000XXZ
HepG2	RFX5	ENCSR000EEA	ENCF000XXX
HepG2	RXRA	ENCSR000BHU	ENCF000PPQ
HepG2	RXRA	ENCSR000BHU	ENCF000PPM
HepG2	SIN3A	ENCSR000BGL	ENCF000PPW
HepG2	SIN3A	ENCSR000BGL	ENCF000PQC
HepG2	SMC3	ENCSR000EDW	ENCF000XXY
HepG2	SMC3	ENCSR000EDW	ENCF000XYC
HepG2	SP1	ENCSR000BJX	ENCF000PQJ
HepG2	SP1	ENCSR000BJX	ENCF000PQE
HepG2	SP2	ENCSR000BOU	ENCF000PQL
HepG2	SP2	ENCSR000BOU	ENCF000PQP
HepG2	SREBF1	ENCSR000EEO	ENCF000XYE
HepG2	SREBF1	ENCSR000EEO	ENCF000XYA
HepG2	SREBF1	ENCSR000EEO	ENCF000XYB
HepG2	SREBF1	ENCSR000EZP	ENCF000XYK
HepG2	SREBF1	ENCSR000EZP	ENCF000XYD
HepG2	SREBF1	ENCSR000EZP	ENCF000XYL
HepG2	SREBF2	ENCSR000EZO	ENCF000XYM
HepG2	SREBF2	ENCSR000EZO	ENCF000XYN
HepG2	SREBF2	ENCSR000EZO	ENCF000XYS
HepG2	SRF	ENCSR000BLV	ENCF000PQW
HepG2	SRF	ENCSR000BLV	ENCF000PQU

HepG2	TAF1	ENCSR000BJN	ENCFF000PRD
HepG2	TAF1	ENCSR000BJN	ENCFF000PRB
HepG2	TBP	ENCSR000EEL	ENCFF000XZI
HepG2	TBP	ENCSR000EEL	ENCFF000XZL
HepG2	TCF12	ENCSR000BJG	ENCFF000PRJ
HepG2	TCF12	ENCSR000BJG	ENCFF000PRL
HepG2	TCF7L2	ENCSR000EVQ	ENCFF000XYX
HepG2	TCF7L2	ENCSR000EVQ	ENCFF000XYV
HepG2	TEAD4	ENCSR000BRP	ENCFF000PRQ
HepG2	TEAD4	ENCSR000BRP	ENCFF000PRU
HepG2	USF1	ENCSR000BGM	ENCFF000PSH
HepG2	USF1	ENCSR000BGM	ENCFF000PSA
HepG2	USF2	ENCSR000EEF	ENCFF000XZT
HepG2	USF2	ENCSR000EEF	ENCFF000XZV
HepG2	YY1	ENCSR000BNT	ENCFF000PSE
HepG2	YY1	ENCSR000BNT	ENCFF000PSD
HepG2	ZBTB33	ENCSR000BNA	ENCFF000PSY
HepG2	ZBTB33	ENCSR000BNA	ENCFF000PST
HepG2	ZBTB33	ENCSR000BHR	ENCFF000PSP
HepG2	ZBTB33	ENCSR000BHR	ENCFF000PSW
HepG2	ZBTB7A	ENCSR000BQA	ENCFF000PTI
HepG2	ZBTB7A	ENCSR000BQA	ENCFF000PTF
HepG2	ZEB1	ENCSR000BVN	ENCFF000PTQ
HepG2	ZEB1	ENCSR000BVN	ENCFF000PTO
HepG2	ZNF274	ENCSR000EVR	ENCFF000XZW
HepG2	ZNF274	ENCSR000EVR	ENCFF000XZX

Supplementary Table S2: List of ChIP-Seq datasets used in the study (HEPG2).

Supplementary Table S2 - List of ChIP-Seq datasets used in the study (A549 cells)

Biosample_name	Target	Experiment	Accession
A549	Control	Data will be provided through GEO upon publication	
A549	NIPBL	Data will be provided through GEO upon publication	
A549	MED1	Data will be provided through GEO upon publication	
A549	SMC1A	Data will be provided through GEO upon publication	
A549	ATF3	ENCSR000BPS	ENCFF000MWR
A549	ATF3	ENCSR000BPS	ENCFF000MWV
A549	BCL3	ENCSR000BQH	ENCFF000MXC
A549	BCL3	ENCSR000BQH	ENCFF000MXD
A549	BHLHE40	ENCSR000DYJ	ENCFF000VOD
A549	CEBPB	ENCSR000DYI	ENCFF000VOL
A549	CEBPB	ENCSR000DYI	ENCFF000VON
A549	CEBPB	ENCSR000BUB	ENCFF000MXK
A549	CEBPB	ENCSR000BUB	ENCFF000MXH
A549	Control	ENCSR000DPE	ENCFF001GWE
A549	Control	ENCSR000DNB	ENCFF000RMI
A549	Control	ENCSR000BVQ	ENCFF000NGN
A549	Control	ENCSR000BVQ	ENCFF000NGK
A549	Control	ENCSR000BVQ	ENCFF000NGM
A549	Control	ENCSR000BPN	ENCFF000NGB
A549	Control	ENCSR000BPD	ENCFF000NGH
A549	Control	ENCSR000BNC	ENCFF000NFR
A549	Control	ENCSR000BKZ	ENCFF000NGC
A549	Control	ENCSR000ASS	ENCFF000AHV
A549	Control	ENCSR000ASS	ENCFF000AHM
A549	CREB1	ENCSR000BRC	ENCFF000MXT
A549	CREB1	ENCSR000BRA	ENCFF000MXO
A549	CTCF	ENCSR000DYD	ENCFF000VPF
A549	CTCF	ENCSR000DYD	ENCFF000VPD
A549	CTCF	ENCSR000DPF	ENCFF001GVU
A549	CTCF	ENCSR000DPF	ENCFF001GWB
A549	CTCF	ENCSR000DNA	ENCFF000RMB
A549	CTCF	ENCSR000DNA	ENCFF000RMC
A549	CTCF	ENCSR000BHW	ENCFF000MYJ
A549	CTCF	ENCSR000BHW	ENCFF000MYN
A549	CTCF	ENCSR000AUF	ENCFF000AHW
A549	CTCF	ENCSR000AUF	ENCFF000AHX
A549	E2F6	ENCSR000BTC	ENCFF000MYY
A549	E2F6	ENCSR000BTC	ENCFF000MYX
A549	ELF1	ENCSR000BPT	ENCFF000MZF
A549	ELF1	ENCSR000BPT	ENCFF000MZG
A549	EP300	ENCSR000BPW	ENCFF000NEG
A549	EP300	ENCSR000BPW	ENCFF000NEI
A549	ETS1	ENCSR000BPU	ENCFF000MZO
A549	ETS1	ENCSR000BPU	ENCFF000MZH
A549	FOSL2	ENCSR000BQO	ENCFF000MZW
A549	FOSL2	ENCSR000BQO	ENCFF000MZT
A549	FOXA1	ENCSR000BRD	ENCFF000NAI
A549	FOXA1	ENCSR000BRD	ENCFF000NAH
A549	FOXA2	ENCSR000BRE	ENCFF000NAQ
A549	FOXA2	ENCSR000BRE	ENCFF000NAU

A549	GABPA	ENCSR000BPY	ENCF000NAY
A549	GABPA	ENCSR000BPY	ENCF000NBA
A549	GATA3	ENCSR000BTI	ENCF000NBH
A549	GATA3	ENCSR000BTI	ENCF000NBM
A549	H2AFZ	ENCSR000AUH	ENCF000AIK
A549	H2AFZ	ENCSR000AUH	ENCF000AIJ
A549	H3K27ac	ENCSR000AUI	ENCF000AKS
A549	H3K27ac	ENCSR000AUI	ENCF000AKR
A549	H3K27me3	ENCSR000AUK	ENCF000ALD
A549	H3K27me3	ENCSR000AUK	ENCF000ALK
A549	H3K36me3	ENCSR000AUL	ENCF000ALU
A549	H3K36me3	ENCSR000AUL	ENCF000ALQ
A549	H3K4me1	ENCSR000AUM	ENCF000AJC
A549	H3K4me1	ENCSR000AUM	ENCF000AIW
A549	H3K4me2	ENCSR000AVI	ENCF000AJI
A549	H3K4me2	ENCSR000AVI	ENCF000AJH
A549	H3K4me3	ENCSR000DPD	ENCF001EMB
A549	H3K4me3	ENCSR000DPD	ENCF001EMC
A549	H3K4me3	ENCSR000ASH	ENCF000AJT
A549	H3K4me3	ENCSR000ASH	ENCF000AJX
A549	H3K79me2	ENCSR000ATP	ENCF000AMB
A549	H3K79me2	ENCSR000ATP	ENCF000AMD
A549	H3K9ac	ENCSR000ASV	ENCF000AJY
A549	H3K9ac	ENCSR000ASV	ENCF000AJZ
A549	H3K9me3	ENCSR000AUN	ENCF000AKK
A549	H3K9me3	ENCSR000AUN	ENCF000AKG
A549	H4K20me1	ENCSR000AUO	ENCF000AMF
A549	H4K20me1	ENCSR000AUO	ENCF000AME
A549	JUND	ENCSR000BRF	ENCF000NDF
A549	JUND	ENCSR000BRF	ENCF000NDH
A549	MAX	ENCSR000DYG	ENCF000VPS
A549	MAX	ENCSR000DYG	ENCF000VPU
A549	MAX	ENCSR000BTJ	ENCF000NDP
A549	MAX	ENCSR000BTJ	ENCF000NDT
A549	MYC	ENCSR000DYC	ENCF000VOW
A549	MYC	ENCSR000DYC	ENCF000VOU
A549	NR3C1	ENCSR000BJT	ENCF000NDD
A549	NR3C1	ENCSR000BJT	ENCF000NDA
A549	NR3C1	ENCSR000BIO	ENCF000NCL
A549	NR3C1	ENCSR000BIO	ENCF000NCO
A549	PBX3	ENCSR000BTN	ENCF000NEY
A549	PBX3	ENCSR000BTN	ENCF000NEP
A549	POLR2A	ENCSR000DMZ	ENCF000RMM
A549	POLR2A	ENCSR000DMZ	ENCF000RMO
A549	POLR2A	ENCSR000BHH	ENCF000NEZ
A549	POLR2A	ENCSR000BHH	ENCF000NFG
A549	_R2Aphosph	ENCSR000DYF	ENCF000VPZ
A549	_R2Aphosph	ENCSR000DYF	ENCF000VQB
A549	obit-IgG-cont	ENCSR496AXR	ENCF000VPI
A549	obit-IgG-cont	ENCSR496AXR	ENCF000VPK
A549	RAD21	ENCSR000DYE	ENCF000VQI
A549	RAD21	ENCSR000DYE	ENCF000VQK
A549	RAD21	ENCSR000BUC	ENCF000NFN

A549	RAD21	ENCSR000BUC	ENCF000NFL
A549	REST	ENCSR000BQP	ENCF000NDY
A549	REST	ENCSR000BQP	ENCF000NDV
A549	SIN3A	ENCSR000BRM	ENCF000NGY
A549	SIN3A	ENCSR000BRM	ENCF000NHA
A549	SIX5	ENCSR000BRL	ENCF000NHG
A549	SIX5	ENCSR000BRL	ENCF000NHI
A549	SP1	ENCSR000BPE	ENCF000NHM
A549	SP1	ENCSR000BPE	ENCF000NHP
A549	TAF1	ENCSR000BPF	ENCF000NHV
A549	TAF1	ENCSR000BPF	ENCF000NHY
A549	TCF12	ENCSR000BQQ	ENCF000NIG
A549	TCF12	ENCSR000BQQ	ENCF000NID
A549	TEAD4	ENCSR000BUD	ENCF000NIP
A549	TEAD4	ENCSR000BUD	ENCF000NIM
A549	USF1	ENCSR000BPV	ENCF000NJL
A549	USF1	ENCSR000BPV	ENCF000NJM
A549	USF1	ENCSR000BJB	ENCF000NJA
A549	USF1	ENCSR000BJB	ENCF000NJE
A549	YY1	ENCSR000BPM	ENCF000NJU
A549	YY1	ENCSR000BPM	ENCF000NJS
A549	ZBTB33	ENCSR000BPZ	ENCF000NKG
A549	ZBTB33	ENCSR000BPZ	ENCF000NKB

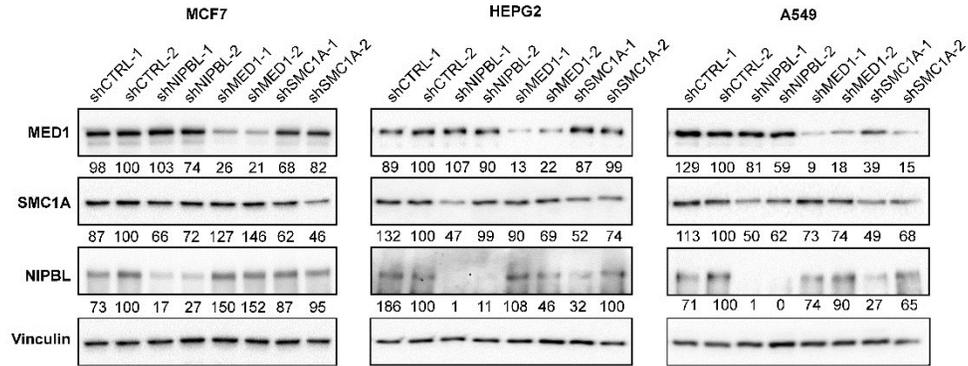
Supplementary Table S2: List of ChIP-Seq datasets used in the study (A549).

Supplementary Table S3 - List of shRNA plasmids and sequences used in the study

Targets	Reference Numbers	shRNA sequences
shCTRL-1	Addgene (#1864)	CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACTTAGG
shCTRL-2	SHC002 (Sigma)	CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCCTTTCATCTTGTGTTTT
shMED1-1	RCN0000019799 (TRC1.5 library - Sigma)	CCGGCAGTAAATCAGAAGTTTCATCTCGAGTGAACCTTCTGATTTACTGCTTTTT
shMED1-2	TRCN0000019800 (TRC1.5 library - Sigma)	CCGGCCGAGTTCCTCTTATCCTAACTCGAGTTAGGATAAGAGGAACTCGGCTTTTT
shNIPBL-1	RCN0000146423 (TRC1.5 library - Sigma)	CCGGCGGATGCTTGTCTTACAACACTCGAGTAGTTGTAAGACAAGCATCCGTTTTTTG
shNIPBL-2	TRCN0000148381 (TRC1.5 library - Sigma)	CCGGCTCAGGATTCAGACTCCATAACTCGAGTTATGGAGTCTGAATCCTGAGTTTTTG
shSMC1A-1	TRCN0000062554 (TRC1.5 library - Sigma)	CCGGCGACAGATTATCGGACCATTTCTCGAGAAATGGTCCGATAATCTGCTTTTTTG
shSMC1A-2	TRCN0000062555 (TRC1.5 library - Sigma)	CCGGCCCGGACTGATTCAGTATACTCGAGTATAATACAGTCCCGCTTTTTTG
shESR1-1	TRCN0000003299 (TRC1.5 library - Sigma)	CCGGAGCACCTGAAGTCTCTGGAACCTCGAGTCCAGAGACTTCAGGGTGCTTTTT
shESR1-2	TRCN0000003301 (TRC1.5 library - Sigma)	CCGGCTCTACTTCACTGCATTCTTCTCGAGAAGGAATGCATGAAGTAGAGTTTTT
shHNF4A-1	TRCN0000019189 (TRC1.5 library - Sigma)	CCGGCCACATGTACTCCTGCAAGTTCGCGAATCTCGAGAGTACATGTGTTTTT
shHNF4A-2	TRCN0000019190 (TRC1.5 library - Sigma)	CCGGCCAGGAACATATGGGAACCAACTCGAGTTGGTCCCATATGTTCTGTTTTT
shFOSL2-1	TRCN0000329767 (TRC1.5 library - Sigma)	CCGGCACGGCCAGTGTGCAAGATTCTCGAGAATCTTGCACACTGGGCCGTGTTTTT
shFOSL2-2	TRCN0000329710 (TRC1.5 library - Sigma)	CCGGCAGTGTGATTTGGAAAGACTTCTCGAGAAGTCTTCCAATCTCACTGCTTTTTG
shFOXA1-1	TRCN0000014878 (TRC1.5 library - Sigma)	CCGGCCGAAAGTTAATGATCCACAACCTCGAGTTGTGGATCATTAAACTTCGCTTTTT
shFOXA1-2	TRCN0000014879 (TRC1.5 library - Sigma)	CCGGGCCTACTACCAAGGTGTGTATCTCGAGATACACACCTTGGTAGTACGCTTTTT
shFOXA2-1	TRCN0000014915 (TRC1.5 library - Sigma)	CCGGGAACGGCATGAACACGTACATCTCGAGATGACGTGTTCTATGCGCTTTTTT
shFOXA2-2	TRCN0000014916 (TRC1.5 library - Sigma)	CCGGCTACGCCAACATGAACCTCATCTCGAGATGGAGTTCATGTTGGCGTAGTTTTT

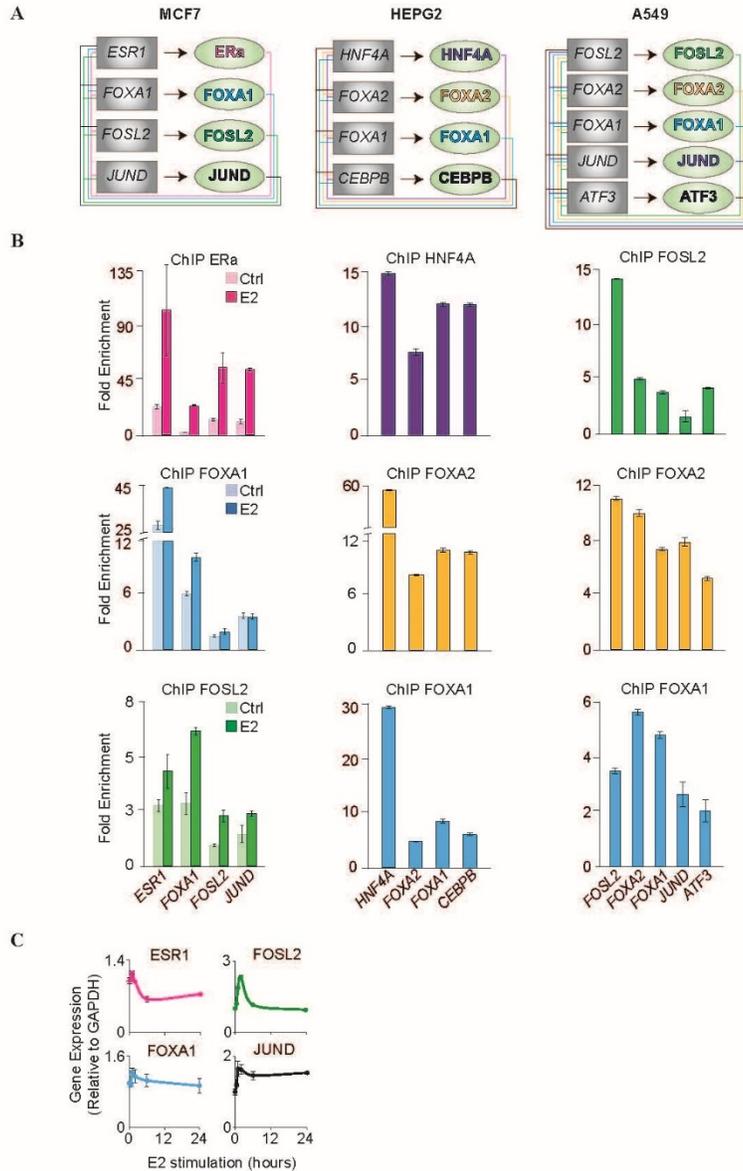
Supplementary Table S3: List of shRNA plasmids and sequences used in the study.

6.2 Supplementary Figures



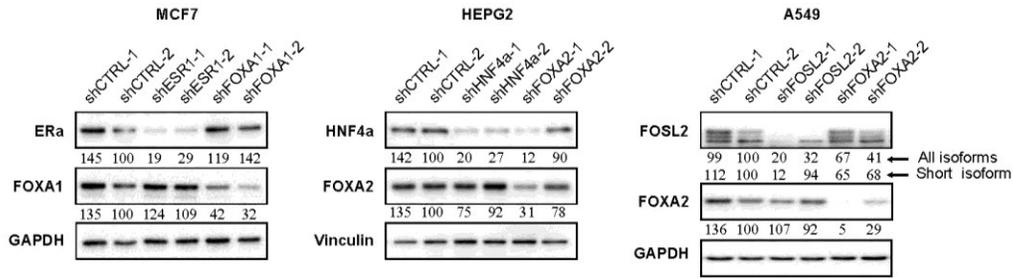
Supplementary Figure S1: shRNA targeting of MED1, SMC1A and NIPBL leads to decreased protein levels.

Western blot validations of MED1, SMC1A and NIPBL depletions in MCF7, HEPG2 and A549 cells for the proliferation and colony formation assays used in **Figure 2.2**. Vinculin was used as loading control to quantify the relative protein levels indicated below each panel. Two shRNAs were used per target and both were efficient at reducing protein levels compared to control shRNAs.



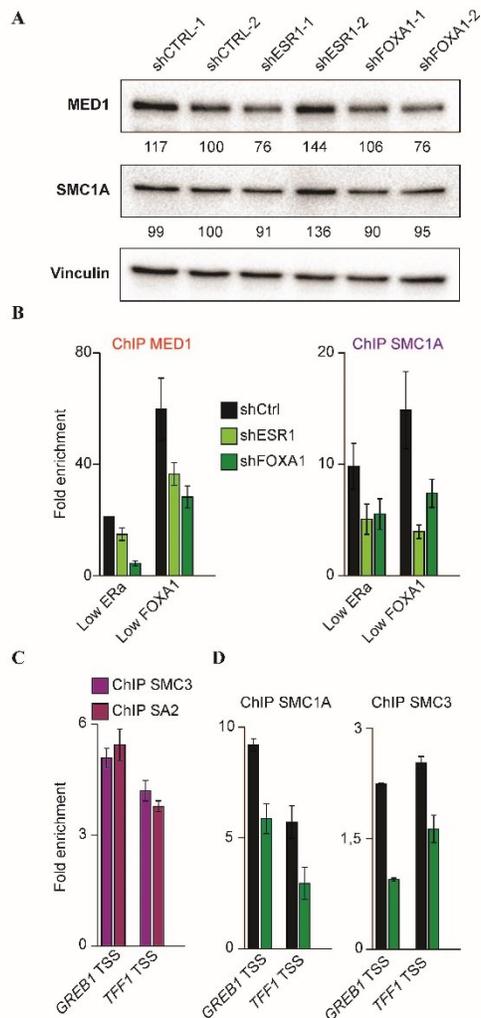
Supplementary Figure S2: Validation of the core transcriptional regulatory circuitry in cancer cells.

(A) Core transcriptional regulatory circuitry of MCF7, HEPG2 and MCF7 cells. The box represents the transcription factor gene loci while the oval represents the protein. Each line represents the interaction of a transcription factor with the indicated gene loci (-10kb to the end of the gene). (B) ChIP-qPCR to validate the transcription factors occupancy at the core transcriptional regulatory circuitry genes. Regions with the most ChIP-Seq signal for each transcription factor was used for quantification. The bar graphs display enrichment fold calculated relative to control regions without ChIP-Seq signal and the error bars show the standard error of the mean for technical qPCR triplicates. (C) Gene expression changes following estrogen (E2, 100 nM) stimulation in MCF7. Fold changes were calculated relative to GAPDH at different time points. Error bars show the standard error of the mean for technical qPCR triplicates.



Supplementary Figure S3: shRNA targeting of the identified transcription factors leads to decreased protein levels.

Western blot validations of the depletions of ERa and FOXA1 in MCF7 cells, HNF4a and FOXA2 in HEPG2 cells and FOSL2 and FOXA2 in A549 cells for the proliferation and colony formation assays used in **Figure 2.4**. GAPDH or vinculin were used as loading control to quantify the relative protein levels indicated below each panel. Two shRNAs were used per target and both were efficient at reducing protein levels compared to control shRNAs. It is to be noted that the shFOSL2-2 did not reduced the level of the all FOSL2 isoforms and was not used in **Figure 2.4**.



Supplementary Figure S4: Validation of the recruitment of Cohesin and Mediator by FOXA and master transcription factors.

(A) Levels of MED1 and SMC1A are not affected by decreased expression of ERa and FOXA1 in MCF7 cells. Protein levels evaluated by Western blot for MED1 and SMC1A following shRNA targeting ERa and FOXA1 in MCF7 cells. Vinculin was used as loading control to quantify the relative protein levels indicated below each panel. Loss of ERa and FOXA1 do not significantly affect expression of MED1 and SMC1A. For **Figure 2.5**, shESR1-1 and shFOXA1-2 were used since they were the most efficient. (B) Depletion of either the master transcription factor ERa or the pioneer transcription factor FOXA1 in MCF7 cells decreases MED1 and SMC1 recruitment at regions with low levels of ERa and FOXA1. Enrichment fold were calculated relative to control regions without MED1 or SMC1 signal. The bar graphs display the enrichment fold and error bars show the standard error of the mean for technical qPCR triplicates. (C) Cohesin subunits SMC3 and SA2 are recruited to the TSS of *GREB1* and *TFF1*. Enrichment fold measured by ChIP-qPCR were calculated relative to control regions without ChIP-Seq signal for Cohesin subunits. (D) Depletion of FOXA1 in MCF7 cells decreases SMC1 and SMC3 recruitment at the TSS of *GREB1* and *TFF1* suggesting the presence of the Cohesin complex.

6.3 Supplementary Data Files

Supplementary Data File 1 is a file containing all the ChIP-Seq enriched regions used in the study. The ZIP archive is accessible using the following URL: https://datahub-kazb7g4u.udes.genap.ca/Scientific_Reports/Supplementary_Data_Files/. (ZIP archive containing BED files; Size: 25.8 MB)

Supplementary Data File 2 is a file containing all the browser tracks used in the study. The ZIP file is accessible using the following URL: https://datahub-kazb7g4u.udes.genap.ca/Scientific_Reports/Supplementary_Data_Files/. (ZIP archive containing bigWig files; Size: 7.74 GB)

6.4 Supplementary Experimental Procedures

6.4.1 Cell Culture Conditions

MCF7 (ATCC, HTB-22), as well as HEPG2 (ATCC, HB-8065) were grown in DMEM (Gibco, 11965-092), while A549 (ATCC, CCL-185) were grown in F12K medium (Gibco, 21127022) as recommended by the supplier. All cell mediums were supplemented with 10% fetal bovine serum (Invitrogen, qualified 12483020), 100 μ M MEM nonessential amino acids (Cellgro, 25-0250), 2 mM L-glutamine (Gibco, 25030-081), 100 U/ml penicillin, 100 μ g/ml streptomycin (Gibco, 15170-063). For ChIP experiments, MCF7 were kept in DMEM w/o phenol-red (Gibco, 31053-028) supplemented with 5% of Charcoal/Dextran treated FBS (Hyclone, AVH78911), 100 μ M MEM nonessential amino acids (Cellgro, 25-0250) and 2 mM L-glutamine (Gibco, 25030-081) for 3 days prior to 100nM of Beta-Estradiol (SIGMA, E8875) for two hours.

6.4.2 Chromatin Immunoprecipitation

Antibodies

For MED1 (CRSP1/TRAP220) occupied genomic regions, we performed the ChIP-Seq experiments using the Bethyl Laboratories (A300-793A) antibody. The affinity purified antibody was raised in rabbit against an epitope corresponding to amino acids 1523-1281 mapping at the C-terminus of human MED1. Specificity of the antibody for ChIP-Seq experiments was validated previously⁶⁹.

For SMC1A occupied genomic regions, we performed the ChIP-Seq experiments using the Bethyl Laboratories (A300-055A) affinity purified rabbit polyclonal antibody. The epitope

recognized by A300-055A maps to a region between residue 1175 and the C-terminus of human SMC1A. Specificity of the antibody for ChIP-Seq experiments was validated previously⁶⁹.

For NIPBL occupied genomic regions, we performed the ChIP-Seq experiments using the Bethyl Laboratories (A301-779A) affinity purified rabbit polyclonal antibody. The affinity purified antibody was raised in rabbit to a region between amino acid residues 1025 and 1075 of human NIPBL. Specificity of the antibody for ChIP-Seq experiments was validated previously⁶⁹.

For ERa occupied genomic regions, we performed the ChIP-Seq experiments using the Santa Cruz (sc-543x) affinity purified rabbit polyclonal antibody. The affinity purified antibody was raised in rabbit to a region between amino acid residues 1025 and 1075 of human ERa. Specificity of the antibody for ChIP-Seq experiments was validated previously by ENCODE project consortium (https://genome.ucsc.edu/ENCODE/validation/antibodies/human_ERalpha_a_validation_Myers.pdf).

Immunoprecipitation

ChIP (ChIP-qPCR) and ChIP coupled with massively parallel sequencing (ChIP-seq) were performed as previously described^{69,178}. Briefly, cells were crosslinked with 1% formaldehyde in culture medium for 10 minutes at room temperature. Cells were then lysed in 10 ml lysis buffer (0.1%SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl ph 8.0 and 150 mM NaCl) and chromatin was sheared to 200-600 bp fragments using a Bioruptor Sonicator (Diagenode). Chromatin extract from 5×10^7 cells was incubated overnight at 4°C with 100 ul of Dynabeads Protein G magnetic beads (Life Technologies, 10004D) previously incubated with 10 ug of selected antibodies (NIPBL (Bethyl, A301-779A), MED1 (Bethyl, A300-793A), SMC1A (Bethyl, A300-055A) and ERa (Santa Cruz, sc543X), SMC3 (Bethyl, A300-158A) and SA2 (Bethyl, A300-060A)).

ChIP-qPCR (Figure 2.5 and Supplementary Figures S2 and S4)

For single gene analysis, primers were designed in the predicted enhancers and promoters regions of *TFF1* and *GREB1* genes. For the *TFF1* locus, primers GAGATGATGCCACCGTACAC and CCCTCACTCACTTTGAAGCA were used for E1, ATCCAGTCCAGGAAGGAGGT and GTCAAGAGAGGAGGCTGTGG for E2 and CCGAGTCAGGGATGAGAGG and GGCCTCCTTAGGCAAATGTT for the TSS. For the *GREB1* locus, primers GAGCTGACCTTGTGGTAGGC and CAGGGGCTGACAACTGAAAT were used for E1, GGGATATGGCTTGTCCATTGT and CATGACACCAGGACCGTAAAG for E2 and

ACCCAGCAAACACTTCAGG and ATTCAGCAGTAGCCCTTCCA for the TSS. Two negative control regions (ATGTCAGGCCCATGAACGAT & GCATTCATGGAGTCCAGGCTTT and AGGACCTGCAGCAAACAGAA & TGTCTACATGGGCTAGTGTGCT) were used to calculate fold enrichments.

For Supplementary Figures S2 and S4, the following additional primers were used:

ChIP-qPCR		
Target	Forward	Reverse
MCF7		
ESR1	CTGCAGTAGGCACTCAGTAAAT	TCAAACCTAACCTGAAACTCGGT
FOSL2	GCAGGCACTGGGAACAG	CACAGAGGGTCTCCCTATCTC
FOXA1	CACAGCACCGGTTTATATCTTTATG	CTGCTCTCAGTTTCTTCTCTT
JUND	CGGTAAACGCCACAACAG	ATGACGTCAACCCACAAGG
Low_Era	GCAACACCAAACACTGCTATC	TGTTTCTCGAGGAGTATATCTGAAC
Low_FOXA1	GGTAGGATACCACAGGCTGATA	AATGCTGCGGAACGAGTC
HEPG2		
CEBPB (ChIP FOXA)	CTCTCTGCTTCTCCCTCTGC	TAAGCGATTACTCAGGGCCC
CEBPB (ChIP HNF4A)	CTGGGCCAGTGTTTTATCCC	GTTATCTGCTGAGCTGGGGT
FOXA1 (ChIP FOXA)	GGTTTCCGAGGAAGGGATTA	CCGGGACCTAAAAGTCAACA
FOXA2 (ChIP FOXA)	CCCCTCCCTGTTACAGTTCA	GGTGTCTGAGGAGTCGGAGA
FOXA2 (ChIP HNF4A)	CTATTTGCTCCTGGGTGGA	AAAGTCCGCTCCTTGGAAGT
HNF4A	TTCTCCTGGCTCTGACACTG	CAACAGTCTGCTTGTCCCC
A549		
ATF3	TAGCATTACGTCAGCCTGGG	GCGTTGCATCACCCCTTTTA
FOSL2 (ChIP FOSL2)	CAAGGTGGAGTCTGGGGTTG	TCTGCCGAGATGAGTCACT
FOSL2 (ChIP FOXA)	AGTTTTCTGAGTTTGTGCTG	TTTACGCGCTCTCCAG
FOXA1 (ChIP FOSL2)	AGAGCTTCAGACTAGAAAAGGGA	GCCTAACGTCACCTGGCTC
FOXA1 (ChIP FOXA)	CAAGCCCACGCCTTTATATT	GACCTCGGAAGGACATCAAA
FOXA2 (ChIP FOSL2)	ACACCACCTTTTCTTTGCAA	TTGGGGATAGGGAAGCAGAG
FOXA2 (ChIP FOXA)	CTTTTCAGTCAACAACCGAGGT	TCTTCTCTCTCACCCCTCC
JUND	AGTTGTTGGGAGAGAGGAC	TTTACTGCGACTTCCCTCCC
RtqPCR		
ESR1	AGCTCCTCCTCATCCTCC	ATAGAGGGGCACCACGTTCT
FOXA1	GTGAAGATGGAAGGCATGA	CCAGGCCCTGAGTTTATGTT
JUND	ACGAGCTCACAGTTCTCTA	GCTGGTTCTGCTTGTGTAATC
FOSL2	CCTGGCGTGATCAAGACCAT	CGACGCTTCTCCTCTCTTC

Library Preparation and Sequencing

ChIP-seq libraries and DNA sequencing were performed by the McGill University and Génome Québec Innovation Centre (MUGQIC). Briefly, ChIP libraries were prepared with the TruSeq DNA sample prep kit using indexing adapters according to the manufacturer's recommendations (Illumina). The starting amount of ChIP DNA used for library preparation ranged from 2.5 ng to 10.5 ng per sample. Library quality and average fragment size was confirmed with Bioanalyzer DNA analysis chips (25-1000 bp, Agilent). TruSeq libraries were multiplexed on the Illumina HiSeq2000. We used single-end reads of 50 bp. Two biological replicates were performed for each target.

6.4.3 Bioinformatics and Analytical Methods

All the ChIP-Seq datasets used in the study (**Supplementary Table S2**) were analyzed with ChIP-Seq pipeline developed at the MUGQIC (version 2.1.0) which is briefly summarized below:

Short-Read pipeline processes

All samples were processed using Trimmomatic 0.33¹⁸³ with the following parameters: LEADING=30, TRAILING=30, SLIDINGWINDOW=4:30, MINLEN=30. ChIP-seq reads were mapped against the hg19 build of the human reference genome using BWA 0.7.12¹⁷⁹ with default parameters. MACS2 2.1.0¹⁸⁴ with default parameters was used for peak calling. All replicates (two replicates for our ChIP-Seq experiments and up to eight for ENCODE datasets) were processed separately until this step.

Data mining and parsing of the metadata

ENCODEExplorer

(<https://www.bioconductor.org/packages/release/bioc/html/ENCODEExplorer.html>) was used to parse the metadata and retrieve the Fastq files from the ENCODE consortium. ChIPseeker¹⁸⁵ was used for data mining of data deposited in the GEO database.

Gene annotation

All peak regions defined by MACS2 were annotated using the UCSC database R package TxDb.Hsapiens.UCSC.hg19.knownGene (<http://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html>) and ChIPseeker.

Merging of ChIP-seq biological replicates

ConsensusSeekerR (<https://www.bioconductor.org/packages/devel/bioc/html/consensusSeekerR.html>) was used to merge the narrowPeak files generated by MACS2 for all replicates to find the best consensus at a given genomic regions. The following parameters were used: extendingSize = 200, expandToFitPeakRegion = TRUE, shrinkToFitPeakRegion = TRUE. The extendingSize parameter allows to determine the size of padding on both sides of the position of the peaks median to create the consensus region. The expandToFitPeakRegion parameter allows to extend the consensus regions to include the entire narrow peak regions of all peaks included in the unextended consensus region. shrinkToFitPeakRegion allow to shrink the consensus regions size to fit the narrow peak regions of the peaks when all those regions are smaller than the consensus region. The minimum number of

replicates with a significant enriched region (minNbrExp) was set to 2 out of 3 for targets with 3 and more datasets. minNbrExp was set to 2 for targets with 2 biological replicates.

PCA analysis (Figure 2.1A)

A binary matrix encoding the presence or absence of NIPBL, MED1, SMC1A or CTCF at all genomic loci where at least one factor was present was constructed (Rows: loci; Columns: ChIP-Seq targets). Principal component analysis (PCA) was performed using the R package FactoMineR (<http://CRAN.R-project.org/package=FactoMineR>) considering the presence or absence of NIPBL, MED1, SMC1A or CTCF as the variables. Visualization of the variables on the first two components was performed using a “correlation circle plot” with the factoextra R package (<https://github.com/kassambara/factoextra>). For the representation, the correlation/loading values are on the 0 to 1 scale, with a value of 1 meaning 100% correlation/loading with the dimension. The density of loci was calculated as the projection of the loci coordinates onto the first dimension (which discriminated the most the loci with CTCF and without CTCF). The y-axis is scaled to one unit and the x-axis is an arbitrary unit representing the coordinates of the loci on the first dimension. Each group (No signal, SMC1A (no CTCF), and SMC1A (with CTCF)) was assigned a color.

Browser tracks (Figures 2.1B and 2.3C)

BigWig files were generated by using Samtools 1.2¹⁷⁹, bedtools 2.17.0¹⁸⁰ and wigToBigWig (developed by the ENCODE team) with normalized count in millions of reads and a smoothing parameter of 225 bp.

Overlaps with chromatin states (Figure 2.1C, D)

The 18 chromatin states for HEPG2 and A549 were retrieved from the consortium Epigenomic Roadmap. The chromatin states were generated with ChromHMM^{46,158} with a core set of 6 chromatin marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K27ac, H3K9me3). Abbreviations: 1_TssA (Active TSS), 2_TssFlnk (Flanking TSS), 3_TssFlnkU (Flanking TSS Upstream), 4_TssFlnkD (Flanking TSS Downstream), 5_Tx (Strong transcription), 6_TxWk (Weak transcription), 7_EnhG1 (Genic enhancer 1), 8_EnhG2 (Genic enhancer 2), 9_EnhA1 (Active Enhancer 1), 10_EnhA2 (Active Enhancer 2), 11_EnhWk (Weak Enhancer), 12_ZNF/Rpts (ZNF genes & repeats), 13_Het (Heterochromatin), 14_TssBiv (Bivalent/Poised TSS), 15_EnhBiv (Bivalent Enhancer), 16_ReprPC (Repressed PolyComb), 17_ReprPC (Weak Repressed PolyComb), 18_Quies (Quiescent/Low).

For MCF7, the same 6 chromatin marks were used (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K27ac, H3K9me3) as follow. Regions enriched for H3K27ac and H3K4me3 were considered Active TSS (TssA). Regions enriched with H3K36me3 were considered transcriptionally active regions (Tx). Regions enriched with both H3K27ac and H3K4me1 were considered active enhancers (EnhA). Regions enriched with H3K9me3 were considered as heterochromatin (Het). Regions enriched for H3K4me3 and H3K27me3, but not H3K27ac were considered Bivalent/Poised TSS (TssBiv). Regions enriched for H3K4me1 and H3K27me3, but not H3K27ac were considered Bivalent Enhancer (EnhBiv). Regions enriched for H3K27me3, but not H3K27ac, TssBiv or EnhBiv were considered Repressed PolyComb (ReprPC). Abbreviations: 1_TssA (Active TSS), 2_Tx (Strong transcription), 3_EnhA (Active Enhancer), 4_Het (Heterochromatin), 5_TssBiv (Bivalent/Poised TSS), 6_EnhBiv (Bivalent Enhancer), 7_ReprPC (Repressed PolyComb).

Once the chromatin states were defined, the ratio of nucleotides from each chromatin state overlapping the consensus peak regions of NIPBL, MED1, and SMC1A (with and without CTCF) was calculated. Additionally, that same ratio was calculated for regions where the consensus peaks of NIPBL, MED1 and SMC1A intersected.

Clustering of MED1, SMC1A (no CTCF), SMC1A (with CTCF) and NIPBL data (Figure 2.1E)

All MED1, NIPBL, SMC1A (with CTCF) and SMC1A (no CTCF) regions from all cell lines were converted into a binary presence matrix, as in **Figure 2.1A**. The overlap in-between all region categories was calculated using the category with the lowest number of regions as the denominator. The resulting matrix was subjected to hierarchical clustering using euclidian distance and Unweighted Pair Group Method with Arithmetic Mean (UPGMA).

KEGG Pathways (Figure 2.2A)

For each cell line, the intersections of the consensus regions for MED1, NIPBL and SMC1A were computed. The selected regions were then annotated using the TxDb.Hsapiens.UCSC.hg19.knownGene package (<http://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html>). Comparison to the hypergeometric distribution was used to evaluate enrichment of all KEGG pathways within the set of identified genes. All generated p-values were then subjected to a unified FDR correction.

Overlaps with transcription factors (Figure 2.3A)

The overlap between the intersection of the MED1, NIPBL, and SMC1A consensus regions and the consensus regions of all available ENCODE factors were computed. The reported pairwise overlap is:

$$\text{Overlap}_{\text{TF}} = \frac{N_{\text{NMS,TF}}}{\min(N_{\text{NMS}}, N_{\text{TF}})}$$

Where $N_{\text{NMS,TF}}$ is the number of regions where NIPBL, MED1, SMC1A and the ENCODE TF are all present, N_{NMS} is the number of regions where NIPBL, MED1, SMC1A are present, and N_{TF} is the number of regions where the ENCODE TF is present.

DNA motif searches (Figure 2.3A)

The regions bearing NIPBL, MED1 and SMC1A were scanned for known human transcription factor binding site models from the HOCOMOCO v9 ¹⁸⁶ database using the Bioconductor package
(<https://www.bioconductor.org/packages/release/bioc/html/PWMEnrich.html>).

6.4.4 Lentiviral shRNAs

Lentiviruses were produced according to the TRC1.5 system from Sigma and plasmids and shRNA sequences used are shown in the **Supplementary Table S3**. The packaging cell line HEK293T was used to produce lentiviral particles containing the pLKO.1 vectors. Cells were transfected with 1:3:6 μg DNA respectively composed of packaging vectors pMD2.G (Addgene, #12259) and psPAX2 (Addgene, #12260), as well as the specific shRNA constructs. Transfections were performed with X-tremGENE 9 transfection reagent (Roche), according to manufacturer's instructions. Supernatants were harvested 48h following transfection. Supernatants were filtered and added to recipient cell lines with 8 $\mu\text{g}/\text{mL}$ polybrene (SIGMA, H9268), as described on Addgene website. Cells infected with the lentivirus were incubated 24h before selection with puromycin for 48 h. To assess the efficiency of each shRNA, Western blot analyses were performed using antibodies against NIPBL, (Bethyl, A301-779A, 1:1000), MED1 (Bethyl A300-793A, 1:5000), SMC1A (Bethyl, A300-055A, 1:5000), ERa (Santa Cruz, sc-543x, 1:5000), FOXA1 (Santa Cruz, sc-101058,

1:1000), FOXA2 (Abnova, 89-019-034, 1:1000), HNF4a (Santa Cruz, sc-8987x, 1:1000) and FOSL2 (Santa Cruz, sc-604x, 1:1000).

6.4.5 Cell proliferation assays (Figures 2.2B and 2.4A)

Cell proliferation assays were performed on A549, HEPG2 and MCF7 with cells infected with lentiviruses containing the validated shRNA constructs. For the Cytation 5 cell Imaging Multi-Mode Reader (BioTek Instruments), A549 (5×10^3 cells/well) were seeded into black-sided clear bottom 96-well plates (Corning, 3603), while HEPG2 (1.2×10^5 cells/well) and MCF7 (1.5×10^4 cells/well) were seeded into clear bottom 12-well plates (Corning, 3513), following a 24h selection. Cells were stained with Hoechst 33342 at $10 \mu\text{g/ml}$ in media for 30 min, at 37°C , prior to imaging. Cell proliferation was measured up to 8 days after seeding. All cell nucleus counting experiments were performed in biological duplicates or triplicates (using 16 fields/well with a 4X microscope objective. Threshold was set to 8,000 with a minimum object size of $5 \mu\text{m}$ and a maximum size set to $100 \mu\text{m}$.

For the xCELLigence Real-Time Cell Analyzer (RTCA) instrument (ESBE Scientific and ACEA Biosciences), MCF7 (1.5×10^4 cells/well) were seeded into E-Plate VIEW 16 PET, a specialized 16-well plate used with the RTCA instrument, into selection media. The cell proliferation index was measured using impedance up to 8 days after seeding.

6.4.6 Soft agar colony formation assays (Figures 2.2C and 2.4B)

Soft agar colony formation assays were performed as previously described¹⁸². MCF7, HEPG2 and A549 cells infected with lentiviruses containing the validated shRNA constructs and selected for 24 hours were used. Two days after their infection, cells were seeded in complete growth medium with 0.3% agar layered onto 0.5% agar beds in 6-well plates (A549: 2×10^4 cells/well, HEPG2: 2×10^4 cells/well and MCF7: 6×10^4 cells/well). Complete medium was added on top twice a week for 13 days (A549 and HEPG2) or 28 days (MCF7). Colonies larger than $800 \mu\text{m}^2$ were counted as positive for growth with ImageJ 1.46r software (NIH).

Chapitre 3 : La modulation du programme transcriptionnel et la résistance endocrinienne dans le cancer du sein hormono-dépendant

Avant-propos

Ce chapitre fait l'objet d'un projet de maîtrise présentement non-soumis à un journal d'édition, mais que j'ai présenté lors de mon examen BMO-8001 à l'automne 2014. J'ai planifié et effectué les expériences et les analyses. J'ai bénéficié du support des Drs Fabien Claude-Lamaze et Eric Fournier pour les analyses bioinformatiques.

Résumé

La plupart des cancers du sein dépendent du récepteur aux estrogènes (ER), un facteur de transcription qui contrôle la croissance tumorale. Les thérapies endocriniennes utilisent cette propriété en ciblant les modes d'action de ER pour retarder la progression tumorale. Malheureusement, la grande majorité des patientes ainsi traitées acquiert des résistances. Nous proposons que les cellules devenues résistantes partagent le même programme d'expression génique que les cellules sensibles, mais qu'elles ont réussi à s'affranchir de la dépendance vis-à-vis de ER au profit d'un facteur de transcription alternatif. Nous avons comparé les profils transcriptionnels des cellules sensibles et résistantes, et observé que Mediator, Cohésin et NIPBL occupent de nouvelles régions régulatrices dans les cellules résistantes. Nous avons utilisé et caractérisé ces nouvelles régions pour réaliser une recherche de candidats pour le remplacement de ER. Nous proposons qu'AP-1 soit impliqué dans l'acquisition de la résistance endocrinienne par modulation de la régulation du programme transcriptionnel des cellules.

1. Introduction

La dérégulation des programmes d'expression géniques est une caractéristique majeure de l'initiation mais aussi de la progression et de l'évolution d'un cancer suite à certains traitements. La plupart des modèles proposés par la littérature dans l'acquisition de la résistance aux traitements anti-estrogéniques pour traiter le cancer du sein blâment les voies de signalisation mitogéniques en aval de l'activation du facteur de transcription ER, induisant ou bien sa transactivation oncogénique ou bien une reprogrammation de son *cistrome*.

Des approches transcriptomiques et protéomiques ont naturellement été employées avec l'apparition de la résistance au Tamoxifen, pour identifier les gènes associés à ce nouveau phénotype ¹⁸⁷. Des analyses de ChIP-seq semblent être en faveur d'une reprogrammation de la chromatine, soulignant des changements de réseaux de régulation impliqués dans l'acquisition de la résistance ¹⁸⁸. La régulation différentielle de gènes candidats a été suggérée en explorant le transcriptome de modèles cellulaires résistants, avec un rôle accru pour la signalisation NFκB dans les cellules ayant subi une exposition prolongée au Fulvestrant ¹⁸⁹. Des inhibiteurs de NFκB sont actuellement en phase clinique ^{190,191}. Une autre approche, un profilage protéomique et une analyse des niveaux de phosphorylation dans les cellules résistantes, a montré une signalisation prononcée de STAT1, lui attribuant un rôle dans la résistance endocrinienne ¹⁹². De façon intéressante, son co-activateur MED1 a, à plusieurs reprises, été associée à la résistance endocrinienne : MED1 a été décrite comme spontanément sur-activée dans des cellules résistantes au Tamoxifen, par sur-expression et/ou phosphorylation. Sa déplétion reverse la résistance et sensibilise les cellules devenues résistantes au Tamoxifen ainsi qu'au Fulvestrant ¹⁹³⁻¹⁹⁵.

Ces études semblent être en faveur d'un modèle transcriptionnel dans lequel un ou des facteurs de transcription aurait pris le dessus dans la cellule pour contrôler la survie cellulaire. Dans l'article de Bianco et al. ¹⁹⁶, les auteurs attribuent le maintien des cellules résistantes au récepteur nucléaire LRH-1. Une étude a également permis l'induction de lignées stables résistantes par incorporation d'une librairie artificielle de facteurs de transcription à doigt de zinc ¹⁹⁷, démontrant que le programme transcriptionnel des cellules résistantes pouvait reposer sur des facteurs alternatifs.

En effet, le détournement transcriptionnel au profit d'un facteur alternatif, remplaçant ER dans ses fonctions, est un modèle pertinent. Deux possibilités s'offrent alors pour la cellule : ou bien le programme transcriptionnel est complètement modifié par l'intervention d'un facteur alternatif, ou bien il est conservé et pris en charge par ce nouveau facteur disposé à remplir les fonctions de ER.

Nous avons démontré dans le **Chapitre 2** que la distribution génomique des co-facteurs NIPBL Mediator et Cohesin pouvait servir de *proxy* à l'identification de facteurs de transcription au cœur de la biologie d'une cellule. Nous avons donc choisi d'appliquer cette découverte à la problématique de l'acquisition de la résistance endocrinienne, émettant l'hypothèse d'un *switch* moléculaire et d'une re-hiérarchisation des facteurs transcriptionnels responsables du maintien de la survie tumorale.

Rapport-Gratuit.com

2. Résultats

2.1 Validation du modèle d'étude : les traitements anti-estrogéniques n'altèrent pas la prolifération des cellules résistantes

L'acquisition de la résistance endocrinienne est complexe, multi-factorielle, et ses phénotypes sont hétérogènes. Le choix d'un modèle cellulaire s'est imposé comme d'une importance capitale au début de ce projet. Les cellules MCF7 constituent le modèle cellulaire le plus utilisé pour l'étude du cancer du sein hormono-dépendant. De nombreuses lignées ont été dérivées de ces MCF7 par les laboratoires de recherche spécialisés, présentant chacune des caractéristiques d'acquisition et de manifestation de résistance uniques^{198,199}. Le laboratoire du Dr Robert Clarke (Université de Georgetown) nous a gracieusement fait don de sa lignée LCC9. Cette dernière a été acquise progressivement par pressions de sélection à partir des MCF7, d'abord par privation estrogénique²⁰⁰, puis par induction de doses croissantes de Fulvestrant²⁰¹. Elle présente une résistance croisée à tous les traitements endocriniens (Tamoxifen, AIs et Fulvestrant) et surtout un niveau d'expression protéique de ER semblable aux cellules sensibles d'origine. Nous avons jugé que cette lignée serait optimale pour cette étude, car elle correspond selon la littérature à l'acquisition classique et progressive de la résistance, avec une expression maintenue de ER. De plus, de nombreux articles ont été publiés sur la résistance au Tamoxifen, alors que peu de mécanismes expliquent encore la résistance aux SERDs tels que le Fulvestrant.

Dans un premier temps, nous avons validé la prolifération des cellules résistantes reçues, en réponse aux traitements anti-estrogéniques (Tamoxifen et Fulvestrant). Nous avons donc mesuré la croissance cellulaire en temps réel des LCC9 et des MCF7, comparativement après ces stimulations. L'inhibition de ER, que ce soit par l'action du SERM Tamoxifen ou du SERD Fulvestrant, abolit la croissance dans MCF7 sans avoir aucun effet sur celle des LCC9 (**Figure 3.1A**). La lignée LCC9 semble donc constituer un modèle d'étude idéal pour ce projet.

2.2 Les cellules sensibles et résistantes expriment les mêmes gènes

ER régule de nombreux gènes, par des voies complexes et souvent redondantes. Des études ont pendant longtemps analysé et décrit la signature génétique induite par l'action de ER^{202,203}. Parmi ses cibles transcriptionnelles directes, le gène *GREB1* est fortement induit et son expression a été associée à la progression tumorale^{204,205}. De plus, ses régions régulatrices distales et spécifiques ont été caractérisées fonctionnellement²⁰⁶. Nous avons cherché à savoir si la répression de ER dans les

LCC9 par les traitements endocriniens était capable d'inhiber l'expression de *GREB1* autant que dans les MCF7. Nous avons extrait l'ARN total de ces deux lignées en conditions de privation estrogénique (Contrôle), et de stimulations positives et négatives de ER par traitements aux estrogènes (E2) puis aux anti-estrogènes (TAM et FUL) décrits dans la **Figure 3.1A**. Nous avons effectué une RTqPCR pour quantifier le niveau de transcrits *GREB1* dans chacune de ces conditions hormonales. *GREB1* montre une expression maintenue dans les LCC9 malgré l'inhibition de ER (**Figure 3.1B**). De plus, tel que prédit, les cellules résistantes semblent avoir conservé une fonction transcriptionnelle pour ER puisque *GREB1* est induit par une stimulation estrogénique dans les LCC9. Cependant, il n'est plus réprimé comme dans les MCF7 suite aux traitements anti-estrogéniques, suggérant que certains mécanismes alternatifs ont été acquis pour maintenir son expression dans les cellules résistantes.

La persistance d'expression du gène *GREB1*, malgré la répression du facteur de transcription ER (**Figure 3.1B**), soulève la question de l'expression des autres gènes régulés par E2 dans les cellules résistantes. Nous avons donc ré-analysé les données d'expression génique générées par le laboratoire du Dr Nicolas Gevry dans les MCF7 et les LCC9 ¹⁹⁶, publiquement accessibles, afin de comparer le programme transcriptionnel de ces deux lignées et de déterminer l'état de conservation du programme d'expression génique lors de l'acquisition de la résistance. La **Figure 3.1C** illustre le pourcentage de gènes communément exprimés dans des réplicats MCF7 et entre les MCF7 et les LCC9. Nous avons utilisé les données obtenues dans chaque réplicat biologique de MCF7 comme référence pour établir le degré de variation « normale », expérimentale. Les profils obtenus sont très similaires, et suggèrent fortement que la cellule exprime globalement les mêmes gènes une fois devenue résistante aux traitements endocriniens.

2.3 Un rôle plus secondaire pour ER dans la régulation du programme d'expression des cellules résistantes

Les cellules résistantes partagent le même programme transcriptionnel que les cellules sensibles, mais sa régulation semble d'être affranchie de sa dépendance vis-à-vis de ER tant au niveau phénotypique que transcriptionnel (**Figure 3.1A & 3.1B**). Cependant, ER est toujours présent dans ces cellules, et semble avoir conservé un mode d'action transcriptionnel (**Figure 3.1B**). Afin de mieux comprendre son rôle dans la régulation transcriptionnelle des cellules résistantes, nous avons réalisé des ChIP-seq dirigés contre ER et ARNPII dans les MCF7 et dans les LCC9, avant et après activation du récepteur par les estrogènes.

Les analyses des données de CHIP-seq ainsi générées révèlent que plus de 70% des régions occupées par ARNPII sont les mêmes dans chaque condition, renforçant l'idée du même programme transcriptionnel. En effet, le profil de CHIP-seq pour ARNPII dans les MCF7 et les LCC9 stimulées est similaire (**Figure 3.1D**) et concordent avec l'idée d'un programme transcriptionnel conservé entre les deux lignées (**Figure 3.1C**). En revanche, bien que les profils pour ER présentent des similitudes qualitatives, le signal moyen semble plus faible dans les LCC9 (**Figure 3.1D**). Nous avons ensuite mesuré et comparé le signal d'occupation moyen de ER entre chaque condition. La **Figure 3.1E** représente le signal d'occupation globale de ER dans chacune des conditions testée (toutes les régions génomiques immuno-précipitées), centré sur le niveau maximal de signal pour chaque région trouvée. Ce résultat confirme la présence basale de ER sur la chromatine des cellules résistantes, ainsi que son activation par les estrogènes. Cependant, le niveau global du signal moyen est bien plus faible dans les cellules résistantes, suggérant un rôle plus secondaire de ER dans ces cellules.

Ce résultat laisse penser qu'un ou des facteur(s) alternatif(s) aurai(en)t pu prendre la place de ER, être devenu plus important dans les cellules résistantes. On pense naturellement à un gain de fonction, et tout aussi naturellement à un partenaire direct ou indirect de ER dans les cellules sensibles, dont l'importance aurait grandi progressivement suite aux attaques répétées ciblant ER. Nous avons donc choisi de mettre en œuvre un criblage de candidats potentiels pour le remplacement de ER dans les cellules résistantes.

2.4 Mediator, NIPBL et Cohesin occupent de nouvelles régions régulatrices dans les cellules résistantes

A l'issue du **Chapitre 2**, nous avons attribué un rôle clef conservé des co-facteurs Mediator, NIPBL et Cohesin dans la biologie des cellules cancéreuses. Nos résultats suggèrent que leur distribution génomique pourrait servir de *proxy* afin de déterminer les facteurs de transcription importants, au cœur du réseau transcriptionnel interconnecté d'une cellule cancéreuse.

Nous avons tout d'abord tenu à vérifier l'importance de ces co-facteurs sur la prolifération des cellules résistantes. Comme décrit dans la **Figure 2.2B** du **Chapitre 2** ainsi que dans le Matériel et Méthodes, nous avons infecté chaque lignée avec des constructions lentivirales d'interférence aux ARN MED1 et NIPBL, et avons mesuré la réponse proliférative des cellules MCF7 et LCC9, respectivement en milieu de culture estrogénique (les MCF7 ayant besoin d'un environnement hormonal favorable à l'activation continue de ER) et privé en hormone (voir **Matériel et Méthodes** section Culture cellulaire). La déplétion de MED1 et de NIPBL affecte la croissance des MCF7 ainsi

que des LCC9 (**Figure 3.2A**). Puisque le rôle des co-facteurs MED1 et NIPBL semble conservé, deux modèles s'offrent alors : ou bien ces co-facteurs occupent les mêmes régions régulatrices dans les cellules sensibles et dans les cellules résistantes, ou bien leur distribution génomique est modifiée.

Nous avons donc effectué les ChIP-seq des sous-unités MED1, SMC1A et NIPBL dans les MCF7 et les LCC9, avant et après traitement aux estrogènes, tels que décrits dans le **Matériel et Méthodes**. Nous avons ensuite identifié les régions conjointement occupées par MED1, SMC1 et NIPBL, en absence de ER, afin de nous limiter aux régions non sous le contrôle du récepteur. Nous avons finalement recoupé ces régions entre les conditions. La **Figure 3.2B** illustre le nombre de régions uniques à chaque condition. Dans les MCF7, une minorité de régions occupées par MED1, NIPBL et SMC1A ne le sont pas par ER. En revanche, il semblerait que les LCC9 aient acquis de nombreuses régions supplémentaires, tout particulièrement en condition de sevrage estrogénique, où les mécanismes de résistance sont vraisemblablement apparus.

2.5 Le rôle d'AP-1 dans la néo-régulation du programme transcriptionnel des cellules devenues résistantes

Comme décrit dans le **Chapitre 1**, Mediator est recruté par les facteurs de transcription, et SMC1A ne possède pas de domaine de liaison à l'ADN, rendant obligatoire leur recrutement par un ou plusieurs facteurs de transcription. Puisque vraisemblablement ER ne contrôle plus ce recrutement, du moins en partie, nous avons choisi de caractériser plus précisément ces régions occupées par MED1, NIPBL et SMC1A dans les LCC9.

Nous avons d'abord effectué une revue de littérature afin de dresser une liste de candidats potentiels pour le remplacement de ER dans les cellules résistantes²⁰⁷. Notre modèle d'acquisition de la résistance est en faveur d'un gain de fonction pour un partenaire déjà établi de ER, et une partie de la réponse transcriptionnelle induite par ER passe par l'action d'AP-1²⁰⁸. Dans ce sens, des résultats d'analyses transcriptomiques, obtenus par le laboratoire du Dr Clarke¹⁸⁷ ont associé l'acquisition de la résistance endocrinienne à une sur-expression du facteur de transcription AP-1. En réalité, l'activité transcriptionnelle de AP-1 a été associée à la résistance endocrinienne depuis les années 90^{209,210}. Plus récemment, des études ont montré que l'inhibition de FOSL1 et que le blocage de AP-1 dans des cellules devenues résistantes étaient capables de renverser ce phénotype^{211,212}.

Des données de ChIP-seq étant disponibles publiquement pour les composants FOSL2 et JUND du complexe AP-1 dans les MCF7 (**Supplementary Table S2**), nous avons choisi de comparer

leur degré de co-occupation génomique avec MED1, NIPBL et SMC1 dans les MCF7 E2 et les LCC9 Ctrl respectivement. La **Figure 3.3A** représente le pourcentage de co-localisation génomique entre les composants d'AP-1, FOSL2 et JUND, et nos co-facteurs. On remarque un gain certain de co-localisation pour les composants AP-1 dans les cellules LCC9 par comparaison aux cellules MCF7. En outre, la recherche de motifs effectuée sur les régions uniques aux LCC9 montre la présence significative de motifs pour les composants AP-1 (**Figure 3.3B**). AP-1 semble donc être un candidat pour la régulation du programme transcriptionnel des cellules résistantes effectuée par MED1, NIPBL et SMC1.

2.6 La sur-expression de FOSL1 modifie la réponse proliférative des cellules sensibles traitées aux anti-estrogènes

Parmi les composants du complexe AP-1, FOSL1 attire l'attention, tout d'abord car son l'inhibition dans des cellules devenues résistantes semble capable de limiter leur prolifération²¹². De plus, FOSL1 a été associé à la résistance au Fulvestrant dans une signature génétique²¹³. Enfin, FOSL1 est un oncogène fréquemment amplifié dans de nombreux cancers, notamment le cancer du sein (**Figure 3.3C**).

Le laboratoire du Dr Laplante (Université Laval) possède une banque lentivirale de surexpression et il se trouve que la construction pour FOSL1 était disponible. Nous avons donc voulu savoir si la sur-expression de FOSL1 dans les MCF7 était capable de rendre ces cellules résistantes au Fulvestrant. Pour cela, nous avons produit le lentivirus contenant la séquence ADNc de FOSL1 et en avons infecté les MCF7 (voir **Matériel et Méthodes**). La **Figure 3.3D** présente les courbes de croissance des cellules MCF7 infectées ou non avec FOSL1, et traitées ou non au Fulvestrant. Ce résultat suggère tout d'abord que la surexpression de FOSL1 dans les MCF7 apporte un gain de prolifération très significatif par comparaison aux cellules non infectées. Lorsque les cellules sont traitées au Fulvestrant, il y a une inhibition marquée de la croissance pour chaque condition, en revanche, on remarque que les cellules infectées se maintiennent en culture contrairement aux cellules non infectées qui meurent rapidement. Les cellules sur-exprimant FOSL1 et traitées au Fulvestrant maintiennent une prolifération quasi-identique aux cellules non infectées et non traitées. AP-1, et plus précisément FOSL1, aurait donc potentiellement un rôle non négligeable à jouer dans le maintien et/ou l'acquisition de la résistance endocrinienne, en contrôlant le programme transcriptionnel des cellules et en recrutant les facteurs nécessaires à la transcription des gènes importants.

3. Discussion

Les oncogènes ont bien longtemps été considérés comme des biomarqueurs, dérégulés dans de nombreux cancers, synonymes de mauvais pronostic et d'agressivité ⁵. Nous proposons ici un modèle plus intégré pour expliquer la résistance endocrinienne, selon lequel les oncogènes de la famille AP-1 seraient en partie responsables de l'acquisition de la résistance endocrinienne dans les cellules de cancer du sein hormono-dépendant, par l'acquisition d'un statut plus important suite à une rehiérarchisation des facteurs transcriptionnels. Nous pensons que la cellule se contente de moduler son programme d'expression génique de manière à modifier les facteurs qui le contrôlent.

Le rôle d'un oncogène est bien souvent d'amplifier le programme transcriptionnel déjà en place dans la cellule ¹⁵¹. Par conséquent, on pourrait naturellement s'attendre à ce que la sur-expression de n'importe quel facteur de transcription oncogénique dans la cellule augmente sa réponse proliférative. Dans les résultats présentés au sein de ce chapitre, nous proposons que les cellules devenues résistantes aient acquis une « addiction transcriptionnelle » aux oncogènes de la famille AP-1. Le rôle discriminatoire de ces facteurs pourrait être testé par des essais de perte de fonction dans les cellules LCC9 sur une gamme d'oncogènes connus impliquant AP-1. Nous pensons que les cellules recouvriraient leur sensibilité aux composés anti-estrogéniques en réponse à une perte d'AP-1 spécifiquement, et ce indépendamment de la réponse proliférative normale observée.

Les cancers du sein triple négatifs (TNBC) comptent pour 20% des cancers du sein et ont un mauvais pronostic car ses déterminants malins sont mal connus ²¹⁴. Ces cellules n'expriment ni ER ni HER2, et ne répondent donc pas aux traitements qui convergent sur les modes d'action de ces deux récepteurs. Nous pensons que la régulation du programme transcriptionnel dans les cellules devenues résistantes pour les traitements endocriniens pourrait s'apparenter à celle qui régit ces cellules du TNBC. De manière intéressante, une étude effectuée dans un modèle de TNBC suggère que l'activité du facteur de transcription AP-1 pourrait contrôler la progression tumorale de ces cellules ²¹⁵. En effet, FOSL1 semble sur-exprimé, et sa déplétion, de même que celle de c-JUN, un autre composant du complexe AP-1, inhibe le phénotype invasif et prolifératif des cellules aussi bien *in vitro* qu'*in vivo*.

Dans la biologie de la tumeur, nous pensons que la pression imposée par les traitements pourrait engendrer une plus forte susceptibilité mutationnelle au niveau du gène FOSL1. En fait, il a été montré que FOSL1 se retrouvait amplifié dans de nombreux cancers. L'acquisition de la résistance s'initie souvent dans un contexte d'instabilité génétique propice à de tels événements. Le déséquilibre

biologique provoqué par les agents endocriniens pourrait forcer la cellule à sur-exprimer des gènes de secours au maintien du programme transcriptionnel et de la survie cellulaire.

Le cancer du sein ainsi que le cancer de la prostate sont tous les deux des cancers hormono-dépendants, le premier sous le contrôle de ER, le second dirigé par l'action du Récepteur aux Androgènes (AR). FOXA1 est connu pour interagir avec chacun de ces récepteurs nucléaires en initiant leur fixation sur de nombreuses régions régulatrices ²¹⁶. La compréhension du fonctionnement de FOXA1 est critique, et sa contribution dans l'acquisition de résistances aux traitements endocriniens est encore mal connue et sujette à controverse ²¹⁷. Un mécanisme proposé pour l'acquisition de la résistance aux anti-androgènes dans le Cancer de la Prostate Résistante à la Castration (CRPCa) passe par l'expression de variants de AR constitutivement actifs, réfractaires aux thérapies endocriniennes de dernière génération ²¹⁸. En maintenant une signature androgénique, ces variants contrôlent la survie et la progression tumorale en condition de castration chimique. La déplétion du facteur pionnier FOXA1 semble capable d'abroger le potentiel oncogénique de ces variants ²¹⁹. En outre, nos résultats obtenus dans le **Chapitre 2** suggèrent que FOXA1 est un déterminant du maintien des programmes d'expression pathologiques dans plusieurs cancers non apparentés. Il serait alors intéressant de déterminer le rôle de FOXA1 dans le maintien des cellules du cancer du sein résistantes, notamment dans le cas d'un gain pour AP-1, dont le statut pionnier a été proposé mais ne fait pas l'unanimité.

Les lignées LCC, produites par le Dr Clarke par pressions de sélection sur les MCF7, offrent un modèle idéal pour étudier la résistance endocrinienne à chaque étape de la progression vers une résistance plus large, de l'acquisition de l'indépendance à E2 jusqu'à la manifestation de résistance croisée à tous les traitements visant ER ²⁰¹. En effet, nous pensons que l'acquisition de la résistance se fait par étapes au niveau moléculaire ¹³¹, et passe par une rehiérarchisation des facteurs de transcription qui contrôlent le programme d'expression, pour acquérir une régulation transcriptionnelle alternative qui ne dépende plus de ER. Nous pourrions par exemple analyser les profils de distribution génomique de FOSL1, ER et ARNPII dans chacune de ces lignées, décrire les réseaux de régulation de ER et FOSL1 et comment ils sont progressivement modulés durant l'acquisition de la résistance pour permettre l'expression des gènes clefs dans la progression tumorale. Un des modèles que nous proposons serait un gain de régions régulatrices occupées par FOSL1, de la même manière que celui observé pour MED1, NIPBL et SMC1 (**Figure 3.2B**) et concomitante avec une perte de signal pour ER (**Figures 3.1D et 3.1E**).

Les conclusions sont basées sur l'utilisation d'un seul modèle cellulaire, alors que l'acquisition de la résistance endocrinienne est un processus par étapes, extrêmement hétérogène ¹³⁰. En effet, dépendamment de la nature des agents, de la durée et même de la posologie des traitements, et enfin de la séquence dans laquelle ils ont été employés durant la progression tumorale, la réponse cellulaire adaptative peut être totalement différente. Il existe de très nombreux modèles cellulaires d'acquisition de la résistance aux traitements anti-estrogéniques, dont les origines varient ^{198,199}. Nous avons choisi un modèle représentatif de la situation clinique, pour lequel les étapes d'acquisition étaient bien établies, en faisant un modèle « contrôlé ». Les chances que les mécanismes proposés soient conservés entre les différents modèles sont plutôt faibles. Cependant, nous pensons que la théorie de l'addiction oncogénique est le point convergent final de l'acquisition d'un phénotype cancéreux agressif.

Les études d'interactions chromatinienne sont en pleine expansion dans le monde de la recherche génomique, et l'application aux cancers offre de nouvelles avenues prometteuses. De plus, il apparaît que les caractéristiques chromatinienne de la cellule d'origine sont des déterminants bien plus solides des profils de mutations cancéreux que la chromatine de cellules cancéreuses apparentées ²²⁰. Nous pourrions alors comparer les structures chromatinienne formées dans les cellules résistantes avec les cellules sensibles, et caractériser les interactions chromatinienne médiées par ER, ARNPIII et FOSL1 dans chaque modèle LCC par ChIA-PET. Un modèle génomique de la résistance impliquant le complexe oncogénique AP-1, non mutuellement exclusif avec un gain de régions régulatrices, pourrait être une dépendance physique et fonctionnelle vis-à-vis de structures déjà en place dans la cellule, comme les usines à transcription décrites dans l'introduction, ou bien la création de nouvelles boucles enhanceurs-promoteurs permettant la co-régulation de gènes importants.

4. Figures

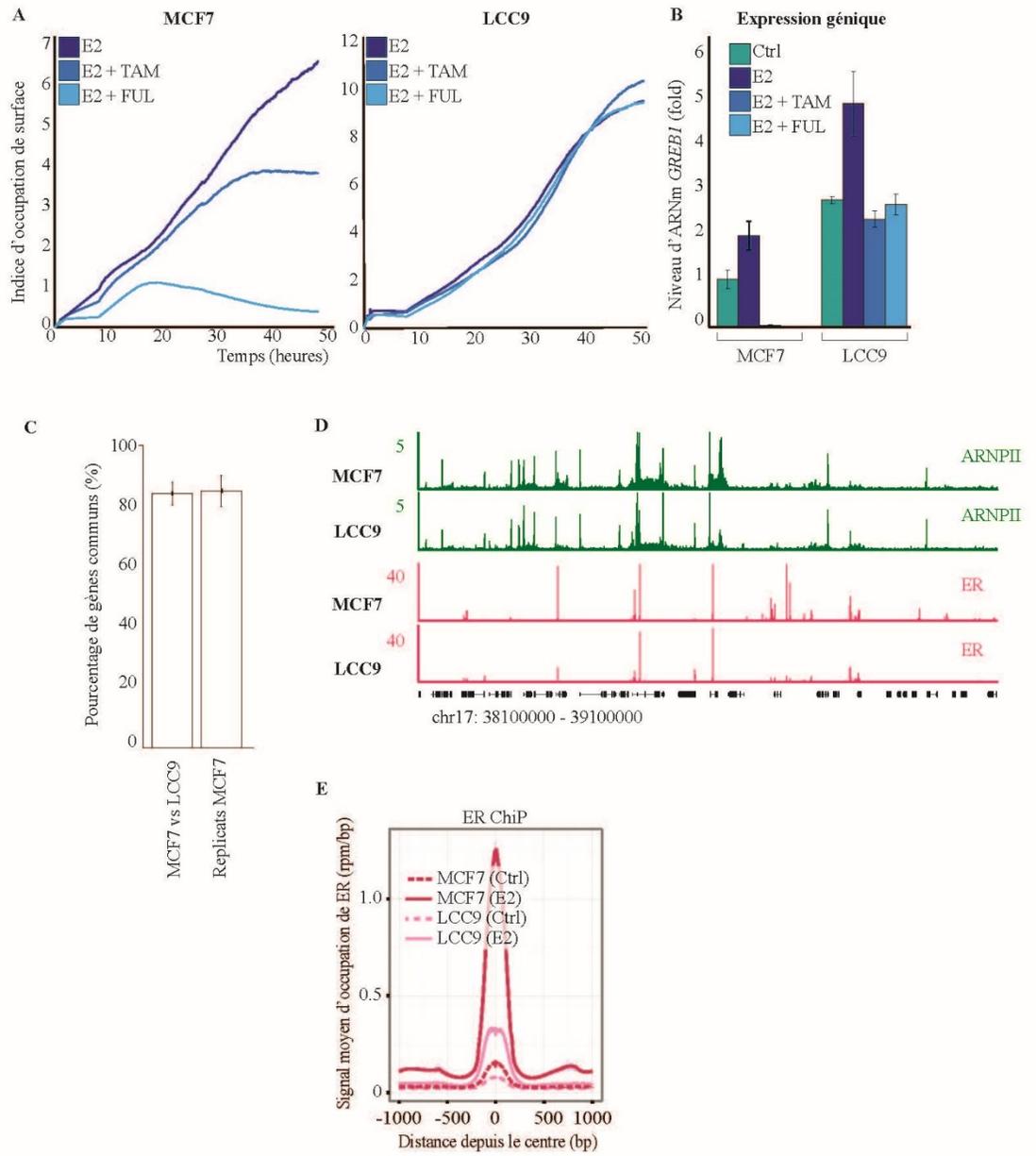


Figure 3.1: Les cellules résistantes LCC9 se sont affranchies de leur dépendance vis-à-vis du récepteur ER.

(A) Mesure en temps réel de la surface occupée par les cellules avec le système xCELLigence RTCA DP. Les MCF7 ont été maintenues dans un milieu de culture naturellement estrogénique (DMEM, 10% FBS) (E2). Les LCC9 ont été maintenues dans un milieu de culture appauvri en hormone et facteurs de croissance (DMEM wo Phenol Red, 5% Charcoal-Dextran Stripped FBS), mais supplémenté avec 1nM d'estrogènes (E2). Les cellules ont ensuite été traitées avec 1uM de Tamoxifen (TAM) ou 100nM de Fulvestrant (FUL). La prolifération des cellules LCC9 n'est pas affectée par les traitements hormonaux ciblant le facteur de transcription ER (TAM & FUL) contrairement aux cellules MCF7.

(B) *GREB1* est une cible transcriptionnelle directe du facteur de transcription ER^{204,205}. Son niveau d'expression a été mesuré dans les cellules MCF7 et LCC9, en condition de privation estrogénique (Ctrl: DMEM wo Phenol Red, 5% Charcoal-Dextran Stripped FBS), et suite à différentes combinaisons de traitements hormonaux (E2, E2+TAM et E2+FUL avec les concentrations mentionnées en (A)). Le niveau de transcrit a ensuite été normalisé avec celui du gène GAPDH. Les barres d'erreur représentent la déviation standard (SD) de la moyenne des triplicats techniques. Les LCC9 maintiennent un fort niveau d'expression de *GREB1* après l'ajout de Tamoxifen ainsi que de Fulvestrant.

(c) Nous avons calculé le pourcentage de gènes communément exprimés entre les deux lignées cellulaires MCF7 et LCC9 parmi les 50% de gènes les plus fortement exprimés¹⁹⁶. Le pourcentage de gènes communs entre des réplicats de MCF7 a été utilisé comme référence. Les profils d'expression génique des cellules MCF7 et des cellules LCC9 sont très similaires.

(D) Les profils de ChIP-seq de l'ARNPII et de ER ont été superposés sur une large région génomique (1000kb) du chromosome 17q12 pour les cellules MCF7 et LCC9. Le signal est représenté en *reads* par million uniquement pour les cellules stimulées aux estrogènes. Les profils d'occupation génique sont similaires entre les deux lignées, malgré un signal global plus faible dans les LCC9.

(E) Nous avons comparé la densité de signal obtenue avec les ChIP-seq dirigés contre ER dans chaque lignée. L'axe des ordonnées reflète la densité de signal obtenue dans chaque condition Ctrl et E2 (100nM) pour chaque lignée. Le signal moyen après soustraction du *whole cell extract* est montré en *reads* par millions par paire de base et est centré pour chaque région enrichie obtenue. Le signal d'occupation de ER est moindre dans les LCC9 par comparaison aux MCF7.

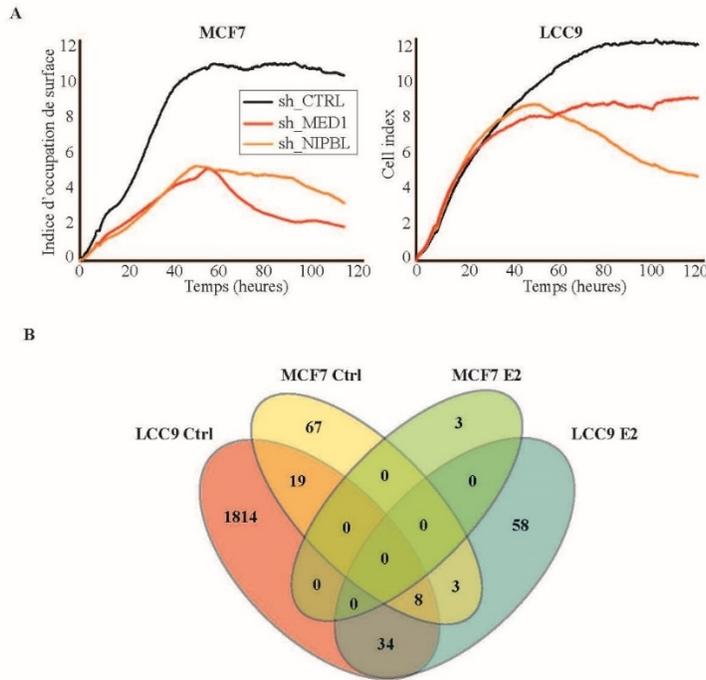


Figure 3.2 : Mediator, NIPBL et Cohesin occupent de nouvelles régions régulatrices dans les cellules résistantes.

(A) Mesure en temps réel de la surface occupée par les cellules avec le système xCELLigence RTCA DP. La prolifération des cellules MCF7 et LCC9, infectées avec un ShRNA contrôle (ShCtrl), un ShRNA dirigé contre NIPBL (ShNIPBL) ou contre MED1 (ShMED1), a été mesurée. Les MCF7 ont été maintenues dans du milieu estrogénique (DMEM, 10% FBS). Les cellules LCC9 ont été maintenues dans du milieu privé d'estrogènes (DMEM wo Phenol Red, 5% Charcoal-Dextran Stripped FBS). Les deux lignées sont sensibles à la déplétion des cofacteurs NIPBL et MED1.

(B) Le diagramme représente les régions génomiques occupées par MED1, NIPBL et SMC1 conjointement, en absence de ER, et leurs intersections pour chaque condition et dans chaque lignée. De nombreuses régions sont uniques à la condition LCC9 Ctrl, suggérant l'acquisition de nouvelles régions régulatrices avec la résistance endocrinienne.

A

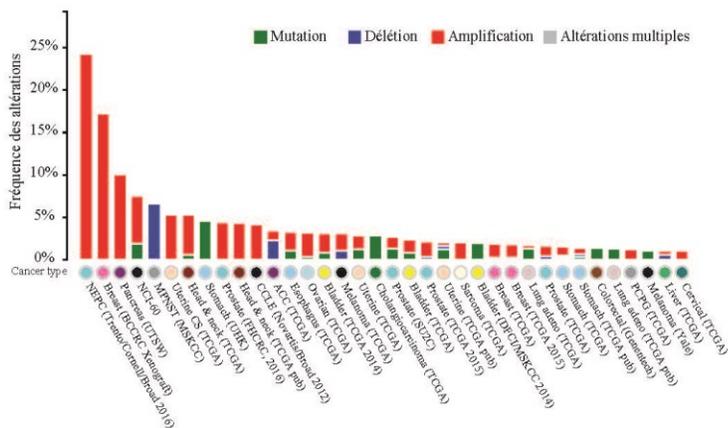
% de co-localisation génomique	FOSL2	JUND
MCF7 E2 MED1	7%	14%
LCC9 Ctrl1 MED1	17%	32%
MCF7 E2 NIPBL	10%	19%
LCC9 Ctrl1 NIPBL	26%	43%
MCF7 E2 SMC1	10%	18%
LCC9 Ctrl1 SMC1	45%	64%

B

Composants AP-1 <i>pvalue</i> motif associé	
cFOS	2,91E-04
cJUN	4,06E-04
FOSL2	1,56E-04
FOSL1	1,86E-04
JUNB	1,86E-04
JUND	1,86E-04
JUN::FOS	5,81E-04



C



D

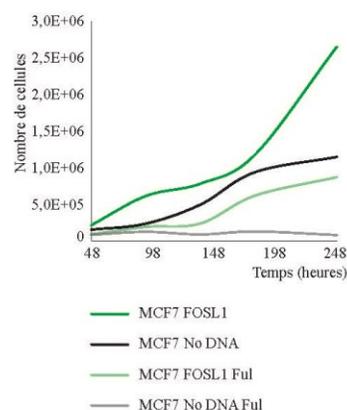


Figure 3.3 : AP-1 est impliqué dans la résistante endocrinienne des cellules du cancer du sein hormono-dépendant.

(A) Pourcentage de co-occupation génomique des facteurs de transcription FOSL2 et JUND (dont les ChIP-seq effectués dans les MCF7 sont disponibles publiquement sur ENCODE, et référencés en **Supplementary Table S2**) avec NIPBL, MED1 et SMC1 respectivement dans les lignées/conditions MCF7 E2 et LCC9 Ctrl. Nous observons un gain de co-localisation pour FOSL2 et JUND pour chaque facteur dans les LCC9.

(B) Représentation statistique génomique des motifs nucléotidiques spécifiques aux composants de AP-1 retrouvés au niveau des régions uniques co-occupées par NIPBL, MED1 et SMC1, en absence de signal ER, dans les cellules LCC9 Ctrl.

(C) cBioPortal (<http://www.cbioportal.org/>) est une interface de regroupement, visualisation, analyse et téléchargement de données génomiques obtenues dans diverses études à large échelle sur le cancer. La figure représente la nature et la fréquence des altérations observées sur le gène FOSL1 dans 36 études différentes. On remarque que la majorité des altérations sont des amplifications, suggérant un gain de fonction associé à ces cancers.

(D) Mesure par points du nombre de cellules MCF7, surexprimant ou non l'oncogène FOSL1 (FOSL1 et No DNA respectivement), soumises ou non au traitement anti-estrogénique Fulvestrant (Ful, 100nM). Les cellules infectées avec FOSL1 ont moins de morbidité que celles non infectées sous pression du Fulvestrant, et leur courbe de croissance ressemble à celle des cellules MCF7 No DNA.

5. Matériel et Méthodes

Les expériences et analyses décrites dans le **Methods** ainsi que dans les **Supplementary Informations** du **Chapitre 2** ont, pour la majorité, été reprises à l'identique dans l'élaboration de ce chapitre. Elles ne seront que brièvement et pertinemment décrites dans cette section.

5.1 Culture cellulaire et traitements hormonaux

Les cellules MCF7 (ATCC, HTB-22) ont été maintenues et amplifiées dans du milieu de culture DMEM (Gibco, 11965-092) supplémenté avec 10% de sérum bovin fœtal (Invitrogen, qualified 12483020), 100 μ M MEM d'acides aminés non-essentiels (Cellgro, 25-0250), 2 mM de L-glutamine (Gibco, 25030-081), 100 U/ml de pénicilline et 100 μ g/ml de streptomycine (Gibco, 15170-063). Les cellules LCC9 (don du Dr R. Clarke) ont été maintenues et amplifiées dans du milieu de culture DMEM w/o phenol-red (Gibco, 31053-028) supplémenté avec 5% de Charcoal/Dextran treated FBS (Hyclone, AVH78911), 100 μ M MEM d'acides aminés non-essentiels (Cellgro, 25-0250), 2 mM de L-glutamine (Gibco, 25030-081), 100 U/ml de pénicilline et 100 μ g/ml de streptomycine (Gibco, 15170-063). Pour les expériences de ChIP-seq, les deux lignées cellulaires ont été amplifiées dans le milieu classique des LCC9 lors de leur dernier passage (trois jours). Les cellules ont été traitées (ou non) avec 100nM of Beta-Estradiol (SIGMA, E8875) deux heures avant leur fixation. Le Tamoxifen (4-HydroTamoxifen, SIGMA, H7904) a été utilisé à une concentration de 1 μ M, et le Fulvestrant (SIGMA, I4409) à 100nM.

5.2 Immunoprécipitation de la chromatine

Les expériences de ChIP-seq ont été réalisées en duplicat biologique pour la lignée MCF7, et en singlicat pour les LCC9. Les cellules ont été fixées avec 1% de formaldéhyde. Les fragments d'ADN soniqués ont ensuite été immuno-précipités avec les anticorps dirigés contre MED1 (Bethyl Laboratories, A300-793A), SMC1A (Bethyl Laboratories, A300-055A), NIPBL (Bethyl Laboratories, A301-779A), ARNPII (Santa Cruz, sc-899x) et ERa (Santa Cruz, sc-543x). Les librairies de ChIP-seq et le séquençage de l'ADN immunoprécipité ont été réalisés par l'Université de McGill et le Centre Innovation Génome Québec.

5.3 Analyses bioinformatiques

Tous les reads de ChIP-seq ont été alignés à la reconstruction hg19 du génome humain en utilisant le pipeline développé par *McGill University and Génome Québec Innovation Centre* (MUGQIC) (https://bitbucket.org/mugqic/mugqic_pipeline).

All ChIP-Seq reads (**Supplementary Table S2**) were aligned to the hg19 build of the human genome using the ChIP-Seq pipeline developed at McGill University and Génome Québec Innovation Centre

Les fichiers bigwig ont été générés avec les outils Samtools 1.2 bedtools 2.17.0¹⁸⁰ et wigToBigWig (développés par l'équipe ENCODE) avec des comptes normalisés en reads par million et un paramètre de smoothing de 225pb. Les régions enrichies ont été annotées avec UCSC database R package TxDb. Pour les associations avec les différents états de la chromatine et les jeux de données de ChIP-seq, l'outil GenomicRanges R package a été utilisé. Les données de ChIP-seq pour FOSL2 et JUND sont référencées en **Supplementary Table S2** de l'article inséré dans le **Chapitre 2**.

5.4 Lentivirus ShRNA

Les numéros de catalogue des plasmides et les séquences des constructions lentivirales sont disponibles dans **Supplementary Table S3**. Les cellules ont été infectées avec les lentivirus durant 24 heures, puis sélectionnées avec de la puromycine (2ug/mL) pendant 48 heures. L'efficacité des déplétions a été jugée et validée par western-blot dans les MCF7 (**Supplementary Figure S1**) avec les anticorps utilisés en ChIP-seq pour MED1 et NIPBL.

5.5 Essais de prolifération

Pour les courbes de croissance en réponse aux drogues ou aux déplétions de MED1 et NIPBL, l'instrument xCELLigence Real-Time Cell Analyzer (RTCA) (ESBE Scientific and ACEA Biosciences) a été utilisé. Les cellules ont été sélectionnées 24 heures avant d'être redistribuées en plaques compatibles avec le système.

Pour les essais de prolifération avec FOSL1, les cellules ont été lysées et comptées en mesure par point tous les 2 jours pendant 10 jours en triplicats avec l'appareil Cellometer K2, Nexcelom Bioscience.

5.6 Extraction d'ARN & RTqPCR

L'ARN total des cellules MCF7 et LCC9 a été extrait selon le protocole TRIZOL Reagent (SIGMA, #15596-018). Les RTPCR ont été effectuées avec 1µg d'ARN purifiés total avec le protocole Superscript VILO Master Mix (Invitrogen, #11754). Les qPCR ont été effectués avec l'intercalant génomique SYBR Select Master Mix (Applied Biosystems, Life Technologies, #4472 908). Les amorces suivantes ont été dessinées et utilisées pour la quantification du gène *GREB1* dans chaque condition (**Figure 3.1B**): Fw CCCATCTTTTCCCAGCTGTA et Rev ATTTGTTTCCAGCCCTCCTT. La normalisation a été réalisée avec GAPDH: Fw CTGCACCACCAACTGCTTAG et Rev GTCTTCTGGGTGGCAGTGAT.

5.7 Lentivirus ORF

La librairie d'expression lentivirale CCSB-Broad est une collection d'*ORFs* (*Open-Reading Frames*) humains intégrés individuellement dans un vecteur lentiviral par clonage Gateway²²¹. Le lentivirus utilisé pour la **Figure 3.3D** est le CCSB-Broad LentiORF – FOSL1 Clone (Accession : BC016648, Clone ID : ccsbBroad304_01858, Gene ID: 8061, Vector Name: pLX304-Blast-V5)

Chapitre 4 : Conclusion/Discussion

Il existe une relation complexe et fonctionnelle à de multiples échelles entre la régulation de la structure chromatinienne et l'exécution et la modulation des programmes d'expression géniques. Les différents travaux relatés dans ce mémoire ont permis de mieux comprendre comment ces programmes sont maintenus dans certains cancers, et suggèrent un modèle simple selon lequel la cellule cancéreuse module son programme et sa régulation pour s'engager dans une réponse proliférative incontrôlée et pour contourner certaines pressions visant à la combattre (**Figure 4**).

Notre compréhension du fonctionnement des programmes transcriptionnels est perpétuellement améliorée mais aussi remise en question par le développement et les progrès de nombreuses techniques d'analyses fonctionnelles et tri-dimensionnelles des interactions dynamiques du génome. Nous avons longtemps cru que les interactions amenant un enhancer à réguler spécifiquement son promoteur cible, à un moment précis dans la vie de la cellule, étaient le mécanisme principal qui régissait la régulation orchestrée et précise des programmes transcriptionnels. Mais les modèles émergents diminuent ce rôle clef prédominant des contacts spécifiques entre les promoteurs et les enhancers, au profit de mécanismes encore plus intégrés à une dimension spatiale d'ordre supérieur.

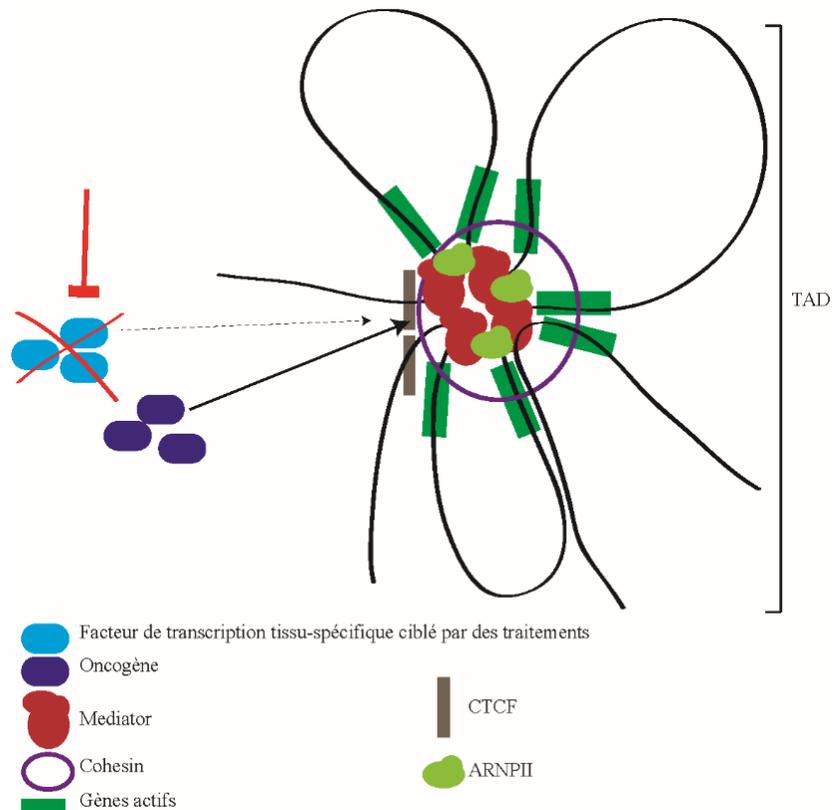


Figure 4 : Modèle de régulation des programmes d'expression génique cancéreux et d'adaptation en réponse à des pressions thérapeutiques d'ordre transcriptionnel. Mediator et Cohesin sont recrutés par les facteurs de transcription tissus-spécifiques au niveau de large *clusters* d'enhancers, au sein de structures tri-dimensionnelles complexes, qui forment des contacts privilégiés pour la co-régulation des gènes clefs dans la progression et la survie tumorale. Lorsque des pressions de sélection d'ordre transcriptionnel surviennent, comme la répression de ces facteurs de transcription qui contrôlent le programme transcriptionnel, la cellule maligne s'adapte afin que son programme d'expression génique soit maintenu, et nous pensons que cela passe par la re-hiérarchisation des facteurs de transcription au profit d'une addiction oncogénique.

1. Les thérapies transcriptionnelles

Les facteurs de transcription sont le point convergent de signalisations oncogéniques, et une expression génique globale aberrante est une caractéristique principale du cancer. Les cellules transformées développent souvent une dépendance vis-à-vis de ces nouveaux programmes d'expression. La transcription est traditionnellement considérée comme non-traitable, mais des agents thérapeutiques se développent progressivement et rapidement pour altérer par exemple les interactions protéiques, la fixation de l'ADN par les facteurs de transcription ou encore pour corriger des modifications épigénétiques ²²². La détermination des facteurs de transcription, qui contrôlent un programme d'expression aberrant, est donc cruciale pour le potentiel thérapeutique de ces agents (**Figure 4**). Ces derniers peuvent faire de la compétition avec les facteurs de transcription pour lier l'ADN à certains *locus* bien précis, ou encore mimer un élément de régulation et fonctionner comme un piège ou une éponge à facteurs de transcription. Les co-régulateurs ont souvent des activités enzymatiques qui peuvent être la cible de molécules chimiques, ayant le pouvoir de moduler ainsi bien les actions d'édition comme d'interprétation de la chromatine, et ainsi de déstabiliser ou bien de stabiliser la structure chromatinienne qui entoure un gène ou un groupe de gènes, pouvant ainsi modifier le comportement d'une cellule cancéreuse. Une autre stratégie thérapeutique est de cibler des acteurs généraux de la régulation transcriptionnelle. Par exemple, un inhibiteur de CDK7 a montré une régression du SCLC (*Small Cells Lung Cancer*) ²²³.

Comme mentionné dans le **Chapitre 1**, les mutations acquises dans de nombreux cancers ont souvent la conséquence de déréguler des enhancers. Par exemple, une étude a intégré les données d'expression de 436 patients de 8 sous-types cancéreux différents avec leur annotations génomiques afin d'identifier et de caractériser les mutations acquises. De nombreuses mutations semblent survenir au sein de régions régulatrices spécifiques et des essais *reporter* ont permis de sélectionner des régions qui montraient des défaut de régulation quand elles contenaient certaines mutations ²²⁴. Une autre étude a utilisé un marquage allélique avec le ChIP-seq DNase1 sur 166 individus et 114 types

cellulaires et ont identifié plus de 60 000 variants communs qui influencent directement la distribution génomique de facteurs de transcription et l'accessibilité de la chromatine *in vivo*¹⁷⁷. Ce genre d'étude à large échelle est de plus en plus fréquent en génomique oncologique, et est complété par le développement de nouvelles technologies qui ont permis une meilleure compréhension et vision globale des mécanismes impliqués dans la régulation des programmes transcriptionnels, et comment leur dérégulation peut entraîner l'apparition de maladies.

2. Mettre à profit les *super-enhancers* dans la lutte contre le cancer

A l'issue du **Chapitre 2**, nous proposons que l'occupation génomique de Mediator, NIPBL et Cohesin puisse servir de *proxy* à l'identification des facteurs de transcription clefs qui contrôlent le programme transcriptionnel d'une cellule cancéreuse. Nous avons utilisé cette propriété dans le **Chapitre 3** pour proposer un rôle pour AP-1 dans la néo-régulation du programme d'expression des cellules cancéreuses ayant progressé de façon adaptative vers une réponse transcriptionnelle plus agressive.

De façon similaire, Mediator a été proposé comme un marqueur de régions régulatrices « d'ordre supérieur », appelées *super-enhancers*²²⁵. Bien que cette appellation soit sujette à controverse²²⁶, les *super-enhancers* consistent néanmoins en des *clusters* d'enhancers, associés à une occupation dense de Mediator et de facteurs de transcription tissus-spécifiques, et associés à une forte activité transcriptionnelle. Basé sur cette définition, un catalogue de *super-enhancers* a été réalisé dans de multiples types cellulaires humains, suggérant une formation tissus-spécifique normale et pathologique de ces structures⁹⁶. Dans le cas du myélome multiple par exemple, Mediator et BRD4 co-occupent des milliers d'enhancers associés à des gènes activement transcrits, mais ils occupent également une proportion minoritaire d'enhancers exceptionnellement larges, des *super-enhancers*, associés à des gènes fortement exprimés dans la maladie, incluant l'oncogène MYC. Le traitement de ces cellules tumorales avec l'inhibiteur du bromodomaine entraîne la perte préférentielle de BRD4 au niveau de ces *super-enhancers* et ainsi un défaut d'élongation de la transcription aux gènes associés, incluant MYC⁹⁵. On pourrait alors penser que ces structures, mais aussi les facteurs transcriptionnels qui les occupent, soient d'excellentes cibles thérapeutiques, et, avec un dosage précis, nous pourrions cibler préférentiellement l'expression des gènes impliqués dans la progression tumorale sans même avoir besoin de les connaître. Ce modèle va à l'encontre de la médecine personnalisée, mais cibler des co-facteurs généraux de la transcription a déjà commencé à faire ses preuves²²³.

Dans le modèle proposé (**Figure 4**), Mediator et Cohésin seraient des composants clés de ces éléments, localisés au sein de structures tri-dimensionnelles complexes comprenant de nombreux contacts chromatinien fonctionnels. Ils seraient recrutés par les facteurs de transcription spécifiques surexprimés ou super-actifs dans la cellule cancéreuse, et recruteraient à leur tour la machinerie transcriptionnelle pour activer et maintenir actifs les gènes de la réponse tumorale. Dans l'exemple discuté de la résistance endocrinienne, le facteur de transcription en question serait réprimé, et ces structures, essentielles pour la cellule, seraient alors compromises. Notre modèle propose que la cellule développe une addiction oncogénique pour maintenir son programme d'expression génique et la formation des structures chromatinien indispensables à cet effet.

Il a également été montré que les *super-enhancers* étaient eux-mêmes activement transcrits, produisant des eRNA dont la fonction dans la régulation des programmes transcriptionnels n'a toujours pas été établie ⁹⁶. Cependant, on pourrait penser que leur identification constitue des biomarqueurs puissants pour le diagnostic de nombreux cancers, bien qu'elle soit encore difficile avec les technologies transcriptomiques classiques actuelles, qui manquent encore de profondeur.

3. Les analyses de contacts tri-dimensionnels à l'échelle génomique offrent une vue d'ensemble des mécanismes qui régulent la transcription normale et aberrante

Les expériences de ChIP, de FISH et de Capture de la Conformation Chromosomique (3C) ont permis une réelle révolution de notre compréhension des mécanismes impliqués dans la régulation génique. Cependant, elles présentent des limites techniques (spécifiques d'une protéine ou d'un locus) et résolutive.

En 2012, Dixon et al. ²²⁷ utilisent la technologie d'analyse des interactions chromatinien *genome-wide*, le Hi-C, pour montrer l'Association de Domaines Topologiques, ou TADs, qui représentent des régions avec un fort taux d'interactions internes. Ces travaux ont été complétés par ceux de Rao et al., ²²⁸ à travers la cartographie de l'architecture tri-dimensionnelle du génome dans une dizaine de lignées cellulaires. Les régions bordantes d'un TAD sont occupées par CTCF, présentent une forte activité transcriptionnelle, et sont des sites d'ancrage de boucles topologiques reliant des régions régulatrices et promotrices. En 2015, Ruan et al. ²²⁹ réalise des ChIA-PET sur les protéines ARNPII et CTCF, et offre une résolution suffisante pour la caractérisation interne de ces TADs, constitués de CCDs (*Chromatin Contact Domains*). Dans le modèle proposé, les frontières des CCDs constituent des usines à transcription pour la co-régulation des gènes *housekeeping* et tissu-spécifiques, suggérant un modèle hautement coopératif. Les auteurs proposent et démontrent

également que l'altération d'un seul nucléotide est capable de modifier ces structures topologiques et d'impacter sur les fonctions transcriptionnelles.

Dans le modèle proposé (**Figure 4**), nous suggérons que la cellule cancéreuse, se dirigeant vers un phénotype plus agressif en réponse à des pressions de sélection, maintienne la structure globale de ses TADs. Dans ce sens, la conservation de la grande majorité des régions occupées par Cohesin – révélatrices des potentiels sites CTCF - dans les cellules MCF7 et LCC9 nous indique que c'est probablement le cas. De plus, des études révèlent que les TADs sont globalement conservés entre les lignées mais aussi entre les espèces ²³⁰. Nos résultats suggèrent également que la cellule devenue résistante acquiert de nouvelles régions régulatrices, modifiant possiblement la structure interne des TADs et certains domaines de contacts. En revanche, les mêmes gènes sont globalement exprimés. En conséquence, l'oncogène répondrait bien à sa définition d'amplificateur transcriptionnel, et viendrait « prendre la place » du facteur réprimé pour maintenir la fonction régulatrice spécifique des contacts chromatinien et permettre au programme de pérenniser. Ce modèle va dans le sens d'une adaptation relativement économique en termes de changements transcriptionnels pour la cellule.

Les analyses transcriptomiques, réalisées dans des cellules cancéreuses ou chez des patients, génèrent souvent des listes exhaustives de gènes dont l'expression est dérégulée, et leur interprétation fonctionnelle et détermination causale est très limitée. Pourtant, la plupart des mutations surviennent au sein de régions régulatrices plutôt que codantes ^{93,94}. Basé sur les récentes découvertes en communication génique, la formation d'un cancer pourrait s'expliquer par l'altération de l'intégrité d'un TAD ou d'un CCD, par exemple par une mutation ponctuelle altérant un site CTCF ou une délétion, résultant en la dérégulation du programme transcriptionnel et en l'expression oncogénique de gènes mitogéniques. On pourrait alors penser que la cartographie des variants génétiques d'un cancer, et leur superposition aux TADs identifiés dans le type cellulaire correspondant, pourrait expliquer de manière globale, précise et causative les dérégulations des programmes transcriptionnels observés. De façon similaire, la comparaison des profils topologiques entre des cellules saines et des cellules cancéreuses correspondantes pourrait proposer de nouveaux modèles génomiques dynamiques et globaux à l'apparition de ces cancers. Une meilleure mise à disposition de matériel biologique pour les laboratoires de recherche va dans le sens des progrès en analyses génomiques tridimensionnelles et leur union constitue une force qui permettra la réalisation de nombreuses études fonctionnelles à large échelle.

4. L'édition du génome: un outil puissant pour combattre le cancer

Le développement de technologies basées sur l'utilisation d'ADN recombinant dans les années 70 a marqué le début d'une nouvelle ère en biologie. Le développement de thérapies géniques par recombinaison homologue a facilité la génération de modèles animaux de *knock-out* et *knock-in* grâce à la manipulation des cellules souches, et l'interrogation de la fonction d'un gène dans son contexte cellulaire et génomique. Malheureusement, cette manipulation est extrêmement inefficace (une cellule sur un million à un milliard effectuerait la recombinaison désirée). De récents progrès ont permis le développement d'outils d'édition du génome efficaces et précis, basés sur l'utilisation d'endonucléases guidées par des ARN, notamment l'application du système bactérien CRISPRcas9, véritable plateforme technologique pour manipuler le génome de façon spécifique et contrôlée²³¹. Le système CRISPRcas9 peut être appliqué à n'importe quel ARN synthétique.

Certaines pathologies sont déjà été associées à des délétions au niveau des régions bordantes de TADs. Par édition du génome, le laboratoire du Dr Lupianez⁹⁷ a généré des souris présentant les mêmes altérations. Chez ces souris, les modifications structurelles causent des interactions ectopiques entre les promoteurs et des éléments régulateurs, un *cluster* d'enhancers est déplacé par rapport à la frontière du TAD associé, altérant l'expression de gènes associés.

Cette étude montre que l'interrogation fonctionnelle d'un variant génétique est maintenant possible *in vivo*. Les perspectives de recherche qu'offre cette technologie sont immenses. Nous pourrions finalement adresser la spécificité de n'importe quel enhancer à n'importe quel moment et dans n'importe quelle cellule. Nous pourrions reproduire n'importe quel système en modèle animal afin d'en étudier la biologie. Les applications ne s'arrêtent pas à la modification génomique covalente, la nucléase pouvant être rendue inactive. Nous pourrions alors créer des protéines de fusion pour altérer spécifiquement l'occupation génomique d'une séquence donnée et la réponse transcriptionnelle qui en résulte.

Outre les possibilités qu'offre ce système pour notre compréhension des mécanismes de régulation normale et pathologique du génome, l'espoir est d'un jour pouvoir appliquer ce système directement en clinique, pour restaurer n'importe quel défaut génomique conduisant à l'apparition ou à la progression de maladies transcriptionnelles, comme le cancer.

Références

1. WATSON, J. D. & CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–8 (1953).
2. Watson, J. D. *Molecular biology of the gene*. (Pearson, 2013).
3. Kornberg, R. D. Structure of chromatin. *Annu. Rev. Biochem.* **46**, 931–54 (1977).
4. Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
5. Lodish, H. *et al.* *Molecular Cell Biology*. (2000).
6. Takata, H. *et al.* Chromatin Compaction Protects Genomic DNA from Radiation Damage. *PLoS One* **8**, 1–11 (2013).
7. Hirano, T. Chromosome Cohesion , Condensation, and Separation. *Annu. Rev. Biochem.* **69**, 115–144 (2000).
8. Koshland, D. & Strunnikov, A. Mitotic chromosome condensation. *Annu. Rev. Cell Dev. Biol.* **12**, 305–333 (1995).
9. Cremer, T. & Cremer, C. Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
10. Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
11. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128**, 707–719 (2007).
12. Workman, J. L. & Kingston, R. E. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579 (1998).
13. Bernstein et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
14. Hahn, S. Structure and mechanism of the RNA Polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).
15. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
16. Vernimmen, D. & Bickmore, W. a. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet.* **31**, 696–708 (2015).
17. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
18. Berger, S. L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).
19. Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720 (1999).
20. Seet, B. T., Dikic, I., Zhou, M.-M. & Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
21. Flaus, A. & Owen-Hughes, T. Mechanisms for ATP-dependent chromatin remodelling: Farewell to the tuna-can octamer? *Curr. Opin. Genet. Dev.* **14**, 165–173 (2004).
22. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **7**, 437–447 (2006).
23. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
24. Noonan, J. P. & McCallion, A. S. Genomics of Long-Range Regulatory Elements. *Annu. Rev. Hum. Genet.* **11**, 1–23 (2010).
25. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a BETA-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
26. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–86 (2014).

27. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr Biol.* **20**, 1–23 (2014).
28. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription. *Cell* **16**, 387–393 (2013).
29. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
30. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483–93 (2016).
31. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, 1377–1388 (2007).
32. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**, 703–713 (2006).
33. Lee, T. I. & Young, R. a. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–51 (2013).
34. Vaquerizas, J., Kummerfeld, S., Teichmann, S. & Luscombe, N. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252–263 (2009).
35. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–26 (2012).
36. Villar, D., Flicek, P. & Odom, D. T. Evolution of transcription factor binding in metazoans mechanisms and functional implications. *Nat Rev Genet* **15**, 221–233 (2014).
37. Biggar, S. R. & Crabtree, G. R. Cell signaling can direct either binary or graded transcriptional responses. *EMBO J.* **20**, 3167–3176 (2001).
38. Sandmann, T. *et al.* A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436–449 (2007).
39. Jakobsen, J. S. *et al.* Temporal CHIP-on-CHIP reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev.* **21**, 2448–2460 (2007).
40. Wilczyński, B. & Furlong, E. E. M. Dynamic CRM occupancy reflects a temporal map of developmental progression. *Mol. Syst. Biol.* **6**, 383 (2010).
41. Halfon, M. S. *et al.* Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* **103**, 63–74 (2000).
42. Giorgetti, L. *et al.* Noncooperative Interactions between Transcription Factors and Clustered DNA Binding Sites Enable Graded Transcriptional Responses to Environmental Inputs. *Mol. Cell* **37**, 418–428 (2010).
43. Merika, M., Williams, A., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFN-beta enhanceosome is required for synergistic activation of transcription. *Mol. Cell* **1**, 277–287 (1998).
44. Voss, T. C. *et al.* Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–554 (2011).
45. Falvo, J. V., Thanos, D. & Maniatis, T. Reversal of intrinsic DNA bends in the IFNbeta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83**, 1101–1111 (1995).
46. Ernst, J. *et al.* Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types. *Nature* **473**, 43–49 (2011).
47. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
48. Zaret, K. S. *et al.* Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol* **73**, 119–126 (2008).
49. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

50. Ghisletti, S. *et al.* Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. *Immunity* **32**, 317–328 (2010).
51. Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161**, 1–14 (2015).
52. Medvedeva, Y. A. *et al.* Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics* **15**, 119 (2014).
53. Magnani, L., Eeckhoutte, J. & Lupien, M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet.* **27**, 465–74 (2011).
54. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
55. Lavin, Y. *et al.* Tissue-Resident Macrophage enhancer landscapes are shaped by the Local Microenvironment. *Cell* **159**, 1312–1326 (2015).
56. D'Alessio, A. C. *et al.* A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports* **5**, 1–13 (2015).
57. Galhardo, M., Berninger, P., Nguyen, T., Sauter, T. & Sinkkonen, L. Cell type-selective disease-association of genes under high regulatory load. *Nucleic Acids Res.* 1–17 (2015). doi:10.1093/nar/gkv863
58. Maston, G. A. *et al.* Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
59. Roeder, R. G. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 201–218 (1998).
60. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–72 (2010).
61. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
62. Poss, Z. C., Ebmeier, C. C. & Taatjes, D. J. The Mediator complex and transcription elongation. *Crit. Rev. Biochem. Mol. Biol.* **1829**, 69–75 (2013).
63. Kang, Y. K., Guermah, M., Yuan, C. & Roeder, R. G. The TRAP /Mediator coactivator complex interacts directly with estrogen receptors α and β through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. *PNAS* **99**, 2642–2647 (2001).
64. Zhang, X. *et al.* MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol. Cell* **19**, 89–100 (2005).
65. Jia, Y., Viswakarma, N. & Reddy, J. K. Med1 subunit of the mediator complex in nuclear receptor-regulated energy metabolism, liver regeneration, and hepatocarcinogenesis. *Gene Expr.* **16**, 63–75 (2014).
66. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
67. Ansari, S. A. *et al.* Mediator, TATA-binding protein, and RNA polymerase II contribute to low histone occupancy at active gene promoters in yeast. *J. Biol. Chem.* **289**, 14981–14995 (2014).
68. Park, S. W. *et al.* Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. *Mol. Cell* **19**, 643–653 (2005).
69. Kagey, M. H. *et al.* Mediator and Cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
70. Nasmyth, K. & Haering, C. H. Cohesin: Its Roles and Mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
71. Kim, J. S. *et al.* Specific recruitment of human cohesin to laser-induced DNA damage. *J. Biol. Chem.* **277**, 45149–45153 (2002).

72. Rudan, M. V & Hadjur, S. Genetic Tailors : CTCF and Cohesin Shape the Genome During Evolution. *Trends Genet.* 1–10 (2015). doi:10.1016/j.tig.2015.09.004
73. Misulovin, Z. *et al.* Association of Cohesin and Nipped-B with transcriptionally active regions of the drosophila melanogaster genome. *Chromosoma* **117**, 89–102 (2008).
74. Liu, J. *et al.* Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.* **7**, e1000119 (2009).
75. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–396 (1999).
76. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
77. Kim, T. H. *et al.* Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell* **128**, 1231–1245 (2007).
78. Xie, X. *et al.* Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *PNAS* **104**, 7145–50 (2007).
79. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
80. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–33 (2008).
81. Levsky, J. M. & Singer, R. H. Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.* **116**, 2833–2838 (2003).
82. Cremer, T. *et al.* Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 777–92 (1993).
83. Bickmore, W. A. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
84. Wit, E. De & Laat, W. De. A decade of 3C technologies-insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
85. Ing-simmons, E. *et al.* Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res.* (2015).
86. Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* **48**, 488–96 (2016).
87. He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in human cells. *PNAS* **111**, 1–9 (2014).
88. Sutherland, H. & Bickmore, W. a. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10**, 457–66 (2009).
89. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306–10 (2013).
90. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S. & Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606–20 (2013).
91. Forbes, S. A. *et al.* COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **39**, 945–950 (2011).
92. Futreal, P. a. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
93. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome TL - 22. *Genome Res.* **22**, 1748–1759 (2012).
94. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science (80-)*. **337**, 1190–1195 (2012).
95. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
96. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–47 (2013).

97. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
98. Zhang, M. H., Man, H. T., Zhao, X. D., Dong, N. & Ma, S. L. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials. *Biomed. reports* **2**, 41–52 (2014).
99. Beatson, G. ON THE TREATMENT OF INOPERABLE CASES OF CARCINOMA OF THE MAMMA: SUGGESTIONS FOR A NEW METHOD OF TREATMENT, WITH ILLUSTRATIVE CASES.1. *Lancet* **148**, 104–107 (1896).
100. Simpson, E. R. & Davis, S. R. Minireview: Aromatase and the Regulation of Estrogen Biosynthesis—Some New Perspectives. *Endocrinology* **142**, 4589–4594 (2001).
101. Ricketts, D. *et al.* Estrogen and progesterone receptors in the normal female breast. *Cancer Res.* **51**, 1817–1822 (1991).
102. Kuiper, G. G., Enmark, E., Peltö-Huikko, M., Nilsson, S. & Gustafsson, J. A. Cloning of a novel receptor expressed in rat prostate and ovary. *PNAS* **93**, 5925–30 (1996).
103. Ali, S. & Coombes, R. C. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat. Rev. Cancer* **2**, 101–12 (2002).
104. Hall, J. M., Couse, J. F. & Korach, K. S. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J. Biol. Chem.* **276**, 36869–72 (2001).
105. Dixon, J. M. Endocrine Resistance in Breast Cancer. *New J. Sci.* **2014**, 1–27 (2014).
106. Tsai, M. & O'Malley, B. W. Molecular mechanisms of action of steroid / thyroid receptor superfamily members. *Annu. Rev. Biochem.* **63**, 451–486 (1994).
107. Smith, C. L. Cross-talk between peptide growth factor and estrogen receptor signaling pathways. *Biol. Reprod.* **58**, 627–632 (1998).
108. Gronemeyer. Transcription: activation by cooperating conformations. *Annu. Rev. Genet.* **25**, 89–123 (1991).
109. Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995).
110. Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* **75**, 567–578 (1993).
111. Hu, M. R., Reid, G., Brand, H., Kos, M. & Gannon, F. Estrogen Receptor- α Directs Ordered, Cyclical, and Combinatorial Recruitment of Cofactors on a Natural Target Promoter. *Cell* **115**, 751–763 (2004).
112. Liu, Z. *et al.* Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. *Cell* **159**, 358–373 (2014).
113. Kocanova, S., Mazaheri, M., Caze-Subra, S. & Bystricky, K. Ligands specify estrogen receptor α nuclear localization and degradation. *BMC Cell Biol.* **11**, 98 (2010).
114. Reid, G. *et al.* Cyclic, Proteasome-Mediated Turnover of Unliganded and Liganded ER α on Responsive Promoters Is an Integral Feature of Estrogen Signaling. *Mol. Cell* **11**, 695–707 (2003).
115. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–97 (2006).
116. Fullwood, M. J. *et al.* An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
117. Hurtado, A., Holmes, K. a., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
118. Magnani, L. & Lupien, M. Chromatin and epigenetic determinants of estrogen receptor α (ESR1) signaling. *Mol. Cell. Endocrinol.* **382**, 633–41 (2014).
119. Pan, Y. F. *et al.* Regulation of estrogen receptor-mediated long range transcription via

- evolutionarily conserved distal response elements. *J. Biol. Chem.* **283**, 32977–88 (2008).
120. Jordan, V. C. Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug Discov* **2**, 205–213 (2003).
 121. Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
 122. Fisher, B. *et al.* Endometrial Cancer in Tamoxifen-Treated Breast Cancer Patients: Findings From the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *JNCI J. Natl. Cancer Inst.* **86**, 527–537 (1994).
 123. Miller, W. R. & Jackson, J. The therapeutic potential of aromatase inhibitors. *Expert Opin. Investig. Drugs* **12**, 337–351 (2003).
 124. Long, X. & Nephew, K. P. Fulvestrant (ICI 182,780)-dependent interacting proteins mediate immobilization and degradation of estrogen receptor- α . *J. Biol. Chem.* **281**, 9607–9615 (2006).
 125. Ciruelos, E. *et al.* The therapeutic role of fulvestrant in the management of patients with hormone receptor-positive breast cancer. *The Breast* **23**, 201–8 (2014).
 126. Dowsett, M. *et al.* Proliferation and apoptosis as markers of benefit in neoadjuvant endocrine therapy of breast cancer. *Clin. Cancer Res.* **12**, 1024–1030 (2006).
 127. Kuukasjärvi T., Kononen J., Helin H., Holli K., I. J. Loss of estrogen receptor in recurrent breast cancer is associated with poor response to endocrine therapy. *J. Clin. Oncol.* **14**, 2584–9 (1996).
 128. Zhao, M. Mechanisms and therapeutic advances in the management of endocrine-resistant breast cancer. *World J. Clin. Oncol.* **5**, 248 (2014).
 129. Musgrove, E. a & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer* **9**, 631–43 (2009).
 130. Hayashi, S. & Kimura, M. Mechanisms of hormonal therapy resistance in breast cancer. *Int. J. Clin. Oncol.* **20**, 262–7 (2015).
 131. Normanno, N. Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer. *Endocr. Relat. Cancer* **12**, 721–747 (2005).
 132. Lupien, M. *et al.* Growth factor stimulation induces a distinct ER α cistrome underlying breast cancer endocrine resistance. *Genes Dev.* **24**, 2219–2227 (2010).
 133. Moerkens, M., Zhang, Y., Wester, L., van de Water, B. & Meerman, J. H. N. Epidermal growth factor receptor signalling in human breast cancer cells operates parallel to estrogen receptor α signalling and results in tamoxifen insensitive proliferation. *BMC Cancer* **14**, 283 (2014).
 134. Miller, T. W. *et al.* Loss of phosphatase and tensin homologue deleted on chromosome 10 engages ErbB3 and IGF-IR signaling to promote antiestrogen resistance in breast cancer. *Cancer Res.* **69**, 4192–4201 (2010).
 135. Shoman, N. *et al.* Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Mod. Pathol.* **18**, 250–259 (2005).
 136. Ciruelos Gil, E. M. Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer. *Cancer Treat. Rev.* **40**, 862–71 (2014).
 137. Milani, A., Geuna, E., Mittica, G. & Valabrega, G. Overcoming endocrine resistance in metastatic breast cancer: Current evidence and future directions. *World J. Clin. Oncol.* **5**, 990–1001 (2014).
 138. Welboren, W.-J. *et al.* ChIP-Seq of ER α and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* **28**, 1418–28 (2009).
 139. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–82 (2008).
 140. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–94 (2009).
 141. Morris, S. A. & Daley, G. Q. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Res.* **23**, 33–48 (2013).

142. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
143. Hochedlinger, K. & Jaenisch, R. Induced Pluripotency and Epigenetic Reprogramming. *Cold Spring Harb. Perspect. Biol.* **7**, 1–25 (2015).
144. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–76 (2006).
145. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–56 (2005).
146. Thayer, M. J. *et al.* Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**, 241–248 (1989).
147. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
148. Wang, A. *et al.* Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
149. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–31 (2012).
150. Zhou, Q., Li, T. & Price, D. H. RNA Polymerase II Elongation Control. *Annu. Rev. Biochem.* **81**, 119–143 (2012).
151. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–45 (2010).
152. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* **21**, 210–219 (2014).
153. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–82 (2016).
154. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
155. Gonda, T. J. & Ramsay, R. G. Directly targeting transcriptional dysregulation in cancer. *Nat. Rev. Cancer* **15**, 686–694 (2015).
156. Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat. Rev. Cancer* **14**, 389–393 (2014).
157. Spaeth, J. M., Kim, N. H. & Boyer, T. G. Mediator and human disease. *Semin. Cell Dev. Biol.* **22**, 776–87 (2011).
158. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
159. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
160. Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science (80-)*. **303**, 1378–81 (2004).
161. Burakov, D., Wong, C. W., Rachez, C., Cheskis, B. J. & Freedman, L. P. Functional interactions between the estrogen receptor and DRIP205, a subunit of the heteromeric DRIP coactivator complex. *J. Biol. Chem.* **275**, 20928–34 (2000).
162. Malik, S., Wallberg, A. E., Kang, Y. K. & Roeder, R. G. TRAP / SMCC / Mediator-Dependent Transcriptional Activation from DNA and Chromatin Templates by Orphan Nuclear Receptor Hepatocyte Nuclear Factor 4. *Mol. Cell. Biol.* **22**, 5626–5637 (2002).
163. Maeda, Y. *et al.* Polyamines modulate the interaction between nuclear receptors and vitamin D receptor-interacting protein 205. *Mol. Endocrinol.* **16**, 1502–10 (2002).
164. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–20 (2013).
165. Alder, O. *et al.* Hippo signaling influences HNF4A and FOXA2 enhancer switching during hepatocyte differentiation. *Cell Rep.* **9**, 261–71 (2014).
166. Wallerman, O. *et al.* Molecular interactions between HNF4a, FOXA2 and GABP identified at

- regulatory DNA elements through ChIP-sequencing. *Nucleic Acids Res.* **37**, 7498–7508 (2009).
167. Shu, J. *et al.* Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963–75 (2013).
 168. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
 169. Wang, J. *et al.* FOSL2 Positively Regulates TGF- β 1 Signalling in Non-Small Cell Lung Cancer. *PLoS One* **9**, e112150 (2014).
 170. Tang, Y., Shu, G., Yuan, X., Jing, N. & Song, J. FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers. *Cell Res.* **21**, 316–26 (2011).
 171. Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell* **147**, 565–76 (2011).
 172. Liu, J. & Krantz, I. D. Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.* **9**, 303–20 (2008).
 173. Solomon, D. a *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science (80-.)*. **333**, 1039–43 (2011).
 174. Xu, H., Tomaszewski, J. M. & McKay, M. J. Can corruption of chromosome cohesion create a conduit to cancer? *Nat. Rev. Cancer* **11**, 199–210 (2011).
 175. Mäkinen, N. *et al.* MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science (80-.)*. **334**, 252–5 (2011).
 176. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–9 (2012).
 177. Maurano, M. T. *et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–401 (2015).
 178. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–9 (2009).
 179. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 180. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
 181. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nat. Protoc.* **1**, 241–245 (2006).
 182. Oster, S. K., Mao, D. Y. L., Kennedy, J. & Penn, L. Z. Functional analysis of the N-terminal domain of the Myc oncoprotein. *Oncogene* **22**, 1998–2010 (2003).
 183. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
 184. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).
 185. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
 186. Kulakovskiy, I. V *et al.* HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* **41**, D195–202 (2013).
 187. Clarke, R. *et al.* Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. *Oncogene* **22**, 7316–39 (2003).
 188. Magnani, L. *et al.* Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *PNAS* 1490–1499 (2013). doi:10.1073/pnas.1219992110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1219992110
 189. Gu, Z. *et al.* Association of Interferon Regulatory Factor-1 , Nucleophosmin , Nuclear Factor-

- κ B , and Cyclic AMP Response Element Binding with Acquired Resistance to Faslodex (ICI 182 , 780). *Cancer Res.* **62**, 3428–3437 (2002).
190. Riggins, R. B., Zwart, A. & Nehra, R. The nuclear factor κ B inhibitor parthenolide restores ICI 182 , 780 (Faslodex ; fulvestrant) – induced apoptosis in antiestrogen-resistant breast cancer cells in antiestrogen-resistant breast cancer cells. *Mol. Cancer Ther.* **780**, 33–41 (2005).
 191. Sas, L. *et al.* The interaction between ER and NF κ B in resistance to endocrine therapy. *Breast Cancer Res.* **14**, 1–14 (2012).
 192. Huang, R. *et al.* Increased STAT1 signaling in endocrine-resistant breast cancer. *PLoS One* **9**, e94226 (2014).
 193. Nagalingam, A. *et al.* Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis* **33**, 918–30 (2012).
 194. Zhang, L. *et al.* Silencing MED1 sensitizes breast cancer cells to pure anti-estrogen fulvestrant in vitro and in vivo. *PLoS One* **8**, e70641 (2013).
 195. Cui, J. *et al.* Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res.* **72**, 5625–34 (2012).
 196. Bianco, S., Brunelle, M., Jangal, M., Magnani, L. & Gévry, N. LRH-1 governs vital transcriptional programs in endocrine-sensitive and -resistant breast cancer cells. *Cancer Res.* **74**, 2015–25 (2014).
 197. Lee, J. *et al.* Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS One* **6**, e21112 (2011).
 198. Berg, D. Characterisation of a tamoxifen-resistant variant of the ZR-75-1 human breast cancer cell line (ZR-75-9al) and stability of the resistant phenotype. *Br. J. Cancer* **59**, 522–526 (1989).
 199. Sweeney, E. E., Mcdaniel, R. E., Maximov, P. Y., Fan, P. & Craig, V. Models and mechanisms of acquired antihormone resistance in breast cancer: significant clinical progress despite limitations. *Horm Mol Biol Clin Investig.* **9**, 143–163 (2012).
 200. Brünner, N., Boulay, V. & Fojo, A. Acquisition of Hormone-independent Growth in MCF-7 Cells Is Accompanied by Increased Expression of Estrogen-regulated Genes but Without Detectable DNA Amplifications. *Cancer Res.* **53**, 283–290 (1993).
 201. Brünner, N. *et al.* MCF7 / LCC9 : An Antiestrogen-resistant MCF-7 Variant in Which Acquired Resistance to the Steroidal Antiestrogen ICI 182 , 780 Confers an Early Cross-Resistance to the Nonsteroidal Antiestrogen Tamoxifen. *Cancer Res.* **57**, 3486–3493 (1997).
 202. Bourdeau, V. *et al.* Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Res.* **36**, 76–93 (2008).
 203. Jagannathan, V. & Robinson-Rechavi, M. Meta-analysis of estrogen response in MCF-7 distinguishes early target genes involved in signaling and cell proliferation from later target genes involved in cell cycle and DNA repair. *BMC Syst. Biol.* **5**, 138 (2011).
 204. Hodgkinson, K. M. & Vanderhyden, B. C. Consideration of GREB1 as a potential therapeutic target for hormone-responsive or endocrine-resistant cancers. *Expert Opin. Ther. drugs* 1–12 (2014).
 205. Rae, J. M. *et al.* GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res. Treat.* **92**, 141–9 (2005).
 206. Deschênes, J., Bourdeau, V., White, J. H. & Mader, S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J. Biol. Chem.* **282**, 17335–9 (2007).
 207. Marcotte, R. *et al.* Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* **164**, 293–309 (2016).
 208. Kushner, P. J. *et al.* Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.* **74**, 311–317 (2000).
 209. Bitonti, A. J., Baumann, R. J., Cashman, A. & Wallace, D. Progression of MCF-7 Breast

- Cancer resistant Phenotype Is Accompanied of AP-1 DNA-binding Activity Cells to Antiestrogen- by Elevated Levels. *Cell Growth Differ.* **7**, 351–359 (1996).
210. Johnston, S. R. D. *et al.* Increased Activator Protein-1 DNA Binding and c-Jun NH2 - Terminal Kinase Activity in Human Breast Tumors with Acquired Tamoxifen Resistance. *Clin. cancer Res.* **5**, 251–256 (1999).
 211. Malorni, L. *et al.* Blockade of AP-1 Potentiates Endocrine Therapy and Overcomes Resistance. *Mol. cancer Res.* **14**, 470–81 (2016).
 212. Pennanen, P. T., Sarvilinna, N. S., Toimela, T. & Ylikomi, T. J. Inhibition of FOSL1 overexpression in antiestrogen-resistant MCF-7 cells decreases cell growth and increases vacuolization and cell death. *Steroids* **76**, 1063–8 (2011).
 213. Knudsen, S. *et al.* Development and validation of a gene expression score that predicts response to fulvestrant in breast cancer patients. *PLoS One* **9**, e87415 (2014).
 214. Hudis, C. A. & Gianni, L. Triple-Negative Breast Cancer: An Unmet Medical Need. *Oncologist* **16**, 1–11 (2011).
 215. Zhao, C. *et al.* Genome-wide profiling of AP-1-regulated transcription provides insights into the invasiveness of triple-negative breast cancer. *Cancer research* **74**, (2014).
 216. Robinson, J. L. L. & Carroll, J. S. FOXA1 is a key mediator of hormonal response in breast and prostate cancer. *Front. Endocrinol. (Lausanne)*. **3**, 1–6 (2012).
 217. Bernardo, G. M. & Keri, R. A. FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci. Rep.* **32**, 113–30 (2012).
 218. Dehm, S. M. & Tindall, D. J. Alternatively spliced androgen receptor variants. *Endocr. Relat. Cancer* **18**, 1–19 (2011).
 219. Jones, D. *et al.* FOXA1 regulates androgen receptor variant activity in models of castrate-resistant prostate cancer. *Oncotarget* **6**, 29782–29794 (2015).
 220. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).
 221. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **8**, 659–661 (2012).
 222. Chunhong Yan, P. J. H. Drugging the undruggable; transcription therapy for cancer. *Biochim. Biophys. Acta* **1835**, 76–85 (2014).
 223. Christensen, C. L. *et al.* Targeting Transcriptional Addictions in Small Cell Lung Cancer with a Covalent CDK7 Inhibitor. *Cancer Cell* **26**, 909–922 (2014).
 224. Melton, C., Reuter, J. a, Spacek, D. V & Snyder, M. Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nat. Genet.* **47**, 710–6 (2015).
 225. Whyte, W. a *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–19 (2013).
 226. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2014).
 227. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485**, 376–380 (2012).
 228. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
 229. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611–1627 (2015).
 230. Rocha, P. P., Raviram, R., Bonneau, R. & Skok, J. A. Breaking TADs: insights into hierarchical genome organization. **7**, 523–526 (2016).
 231. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).