Problématiques et objectifs

D'une façon générale, nous posons la question de l'impact des anomalies moléculaires et en particulier des gènes des immunoglobulines dans la lymphomagénèse B et plus particulièrement dans la LLC.

D'une part, afin de rechercher de nouveaux oncogènes, nous nous sommes intéressés aux remaniements chromosomiques (translocations) impliquant les gènes des immunoglobulines dans les lymphomes B, et en particulier dans la LLC et le lymphome lymphocytique. Nous avons sélectionné les cas pour lesquels le partenaire n'est pas connu, dans la mesure où ce type de translocation entraîne la dérégulation du gène partenaire via le rapprochement d'un élément régulateur du gène des immunoglobulines. L'objectif a donc été de mettre en évidence les points de cassures dans trois cas de LLC/lymphome lymphocytique pour lesquels ce type de translocation avait été identifié. Nous avons voulu caractériser le gène partenaire dans l'optique de découvrir de nouveaux oncogènes. Enfin, nous avons tenté de comprendre le mécanisme de survenue de ces translocations et le mode de dérégulation du gène partenaire sous l'effet de la translocation.

D'autre part, dans la suite d'un travail sur les LLC avec pic monoclonal d'immunoglobuline, nous avons étudié l'impact pronostique d'une électrophorèse des protéines sériques normale (immunoglobulines polyclonales sans hypogammaglobulinémie) dans l'ère du NGS. Pour cela, nous avons mis au point un panel Ampliseq appliqué aux lymphomes, permettant de détecter les mutations d'intérêt à visée diagnostique et/ou pronostique, ainsi que les principales variations du nombre de copies de la LLC, à savoir la del(17p), la del(11q), la trisomie 12 et la del(13q). Dans le cadre de la détection des variations du nombre de copies, nous avons collaboré avec l'équipe EA6309 de la faculté de médecine de Limoges afin d'améliorer un outil dédié, le logiciel Cov'Cop, en particulier pour son application dans les pathologies somatiques. Nous avons ensuite établi les performances de cette technique dans la détection des principales anomalies de nombre de la LLC. Enfin, nous avons évalué sa place dans la stratégie diagnostique au laboratoire, ainsi que son apport pour la détection de certaines autres anomalies de nombre et de disomies uniparentales (ou UPD pour UniParental Disomy).

Chapitre 4 : Matériels et méthodes

A- Patients et échantillons

1- Etude de translocations impliquant le gène des immunoglobulines

a- Patient n°1

Le premier patient est un homme chez lequel a été diagnostiqué en 2012, à l'âge de 56 ans, un lymphome lymphocytique. Cliniquement, le patient se présentait alors avec de multiples adénopathies cervicales de 1 à 2 cm. Après biopsie ganglionnaire cervicale gauche en Septembre 2012, le diagnostic a été établi par examen anatomo-pathologique (lymphome lymphocytique de cytologie atypique). Des prélèvements de cette biopsie ganglionnaire avaient également été transmis en cytologie, cytométrie en flux et cytogénétique. En cytologie, l'aspect était plutôt celui de grandes cellules d'aspect immunoblastique. Un immunophénotypage lymphocytaire avait alors été réalisé et montrait la présence d'un clone B représentant environ 30 % des cellules de l'échantillon et de profil CD5+, CD20+ d'expression modérée et de monotypie Kappa d'intensité faible. Une phase leucémique *a minima* était également retrouvée (lymphocytose à environ 2,5 G/L avec 1/3 de lymphocytes B monotypiques en cytométrie en flux, avec score de Matutes à 3, de profil CD5+, CD23+, FMC7+ fort, CD22+fort, monotypie Kappa d'intensité faible, CD11c+, compatible avec la cytologie atypique mise en évidence en anatomo-pathologie).

La présence en cytogénétique d'une translocation impliquant le gène *IGH* et un locus situé en 17q24 inconnu en font un cas intéressant. Ces résultats obtenus sur ganglion ont été confirmés sur sang.

Par ailleurs, l'étude du statut mutationnel IGHV retrouve un réarrangement IGHV5-51/IGHD6-6/IGHJ6 de profil muté (taux de mutations de 3,5 %). L'électrophorèse des protéines sériques réalisée en Octobre 2012 retrouvait la présence de micropics multiples de composantes IgG Kappa et IgG Lambda.

Le cas de ce patient avait été discuté en réunion de concertation pluridisciplinaire, décidant d'une surveillance simple avec la réalisation de bilans sanguins tous les 4 mois et consultation médicale annuelle.

Nous avons sélectionné ce cas pour étude du remaniement entre *IGH* et le locus 17q24. Nous disposons de matériel en cytogénétique et biologie moléculaire sur ganglion et sur sang.

b- Patient n°2

Le deuxième patient est une femme chez laquelle le diagnostic de LLC a été réalisé en 2007 à l'âge de 75 ans. Elle est alors classée en stade Binet A et est restée en abstention thérapeutique. Un immunophénotypage lymphocytaire avait alors été réalisé : le score de Matutes était à 5 (monotypie Lambda), avec un CD38 positif. La cytologie ne retrouve pas d'atypie notable. L'électrophorèse des protéines sériques retrouvait une hypogammaglobulinémie à 4,1 g/L, et l'immunofixation ne mettait pas en évidence de composant monoclonal.

Début 2012, devant la survenue d'infections à répétition, une substitution par immunoglobulines polyvalentes est initiée.

Un bilan biologique complet est alors réalisé. Le statut IGHV est muté (réarrangement IGHV4-61/IGHJ4, taux de mutations à 4 %). En cytogénétique, une translocation impliquant les chromosomes 17 et 22 est mise en évidence dans 19 mitoses sur 20. Le locus au niveau du chromosome 17 est situé en 17q25. Il n'y a pas de del(17p) mise en évidence.

En Décembre 2012, devant des signes de progression de la maladie (augmentation de la lymphocytose et des adénopathies, et splénomégalie volumineuse), un traitement par Chloraminophène est initié. A partir du mois de Mars 2013, du Rituximab est également instauré pour 6 cures. Devant une bonne réponse, le traitement est arrêté en Août 2013 avec surveillance.

Début 2016, une reprise de la LLC est objectivée avec une lymphocytose à 111 G/L (14,2 G/L en Juin 2015), et une volumineuse splénomégalie. La cytogénétique réalisée alors ne met pas en évidence d'anomalie supplémentaire à la t(17;22). Un traitement par Rituximab-Bendamustine est alors mis en place en Avril 2016. Après 6 cures et l'obtention d'une bonne réponse (correction de la lymphocytose et absence de splénomégalie palpable), la patiente est en abstention thérapeutique. Les dernières nouvelles de Janvier 2018 rapportent une maladie stable.

Devant la présence d'une translocation impliquant le locus 22q11 (soit potentiellement *IGL*) et le locus 17q25 (soit proche de celui du patient n°1), ce dossier a été sélectionné. Nous disposons de matériel disponible en cytogénétique et biologie moléculaire.

c- Patient n°3

Le troisième patient est une femme chez laquelle a été diagnostiqué une LLC en 2009 au Royaume-Uni, à l'âge de 50 ans. Devant la présence de volumineuses adénopathies progressives, elle a bénéficié en 2010 d'un traitement par FCR (1 seule cure devant la survenue d'effets indésirables digestifs majeurs ainsi que de complications infectieuses).

En Janvier 2013, la patiente est vue en consultation au CHU de Limoges. Devant un stade Binet B progressif, un traitement par Rituximab-Bendamustine a été mis en place (6 cures). Une rémission complète avait alors été obtenue et une surveillance simple a alors été débutée.

Fin 2016, une reprise de la maladie avec un temps de doublement de la lymphocytose inférieur à 6 mois (Lymphocytes à 0,42 G/L en Septembre, 74 G/L en Décembre 2016), a motivé l'initiation d'un nouveau traitement. Depuis Janvier 2017 la patiente est sous Ibrutinib. En Mai 2017, devant la survenue d'infections de la sphère ORL et la présence d'une hypogammaglobulinémie majeure (1,7 g/L ; valeurs normales locales de 6,0 à 11 g/L), une substitution à base d'immunoglobulines polyvalentes a été mise en place.

Le score de Matutes est à 4 (chaîne légère de surface Lambda d'intensité medium), profil IGHV non muté (réarrangement IGHV3-21/ IGHJ6, taux de mutations somatiques à 0 %, assignée au subset 2). La cytologie de départ était commune, avec apparition au cours de l'évolution de prolymphocytes. La cytogénétique a mis en évidence un caryotype complexe avec notamment une translocation entre les chromosomes 8 et 14, impliquant *IGH* et n'entraînant pas de remaniement de *MYC* sur le chromosome 8 (vérification effectuée en FISH). Ce dossier a donc été sélectionné. Nous disposons de matériel en cytogénétique et biologie moléculaire sur sang. Les électrophorèses des protéines sériques réalisées depuis le diagnostic en 2012 ne mettent en évidence de pic monoclonal, mais une hypogammaglobulinémie.

2- LLC et électrophorèse des protéines sériques

Les critères de sélection et les caractéristiques des patients pour l'étude de l'impact pronostique de l'électrophorèse des protéines sériques sont détaillés dans l'article 1 (page 194).

3- Evaluation du logiciel Cov'Cop dans la LLC

Dans cette étude, 82 cas de LLC avec annotation cytogénétique complète ont été sélectionnés. Nous avons récupéré les données cliniques concernant le stade Binet et la survie sans

traitement. Les caractéristiques des patients au diagnostic sont résumées dans le Tableau 7. L'échantillon du diagnostic était disponible pour 50 patients.

		Total (n = 82)
Sexe		Hommes : 58,5 % (n = 48) Femmes : 41,5 % (n = 34)
Age (ans)		64,9 +/- 10,7
Stade Binet (n = 78)	A B C	76,9 % (n = 60) 14,1 % (n = 11) 9,0 % (n = 7)
Statut mutationnel IGHV (n = 77) (M-CLL)		42,9 % (n = 33)
Lymphocytose (G/L)		30,6 +/- 39,5
Hémoglobine (g/dL)		13,8 +/- 1,7
Plaquettes (G/L)		204,0 +/- 91,7
Cytogénétique (Caryotype +/- FISH et/ou QMPSF)	- del(17p) - del(11q) - trisomy 12 - del(13q) - del(13q) isolée	11,0 % (n = 9) 20,7 % (n = 17) 11,0 % (n = 9) 53,7 % (n = 44) 25,6 % (n = 21)
Mutations	*NOTCH1* *SF3B1* *TP53* *BIRC3* *ATM*	9,8 % (n = 8) 24,4 % (n = 20) 20,7 % (n = 17) 4,9 % (n = 4) 21.7% (n = 18)
Score de Rossi	Risque très faible Risque faible Risque intermédiaire Risque haut	25,6 % (n = 21) 15,9 % (n = 13) 34,1 % (n = 28) 24,4 % (n = 20)

Tableau 7 : Caractéristiques des patients de l'étude sur l'évaluation du logiciel Cov'Cop dans la LLC La moyenne +/- l'écart-type sont indiquées pour l'âge, la lymphocytose, le taux d'hémoglobine et la numération plaquettaire.

B- Techniques conventionnelles

Les techniques conventionnelles utilisées pour cette thèse (extraction d'ADN, QMPSF (pour Quantitative Multiplex PCR of Short Fluorescent Fragments), Mutations somatiques caryotype et FISH) sont décrites dans l'article 1 (page 194).

Les caractéristiques des sondes de routine utilisée pour la recherche de translocation sont détaillées ci-après.

Concernant le locus *IGH*, deux sondes différentes ont été utilisées. La première est fournie par Kreatech (Leica, Wetzlar, Germany). Il s'agit d'une sonde de séparation, constituée de 2 signaux. Le signal rouge cible les gènes de jonction et de diversité, l'ensemble de parties constantes d'*IGH* ainsi que la région en aval du gène, tandis que le signal vert ne couvre qu'une partie de la région IGHV et la région en amont de celle-ci (Figure 32).

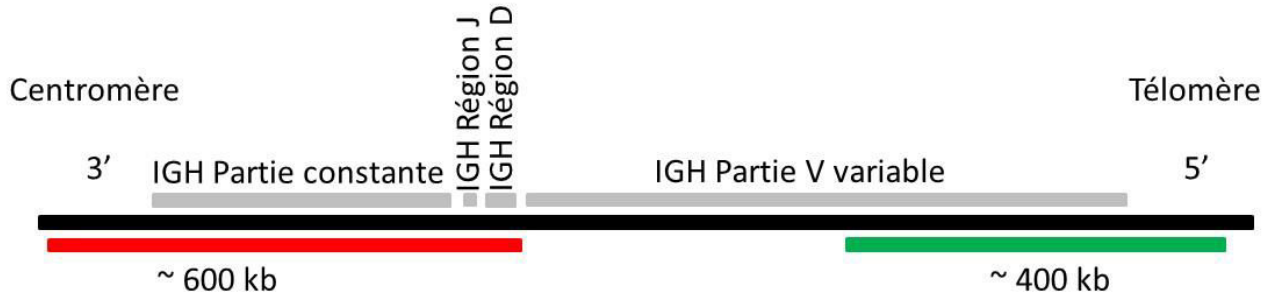

Figure 32 : Schéma de la sonde IGH Kreatech adapté selon les indications du fournisseur.

La seconde sonde utilisée ciblant *IGH* est la sonde Vysis LSI IGH Dual Color Break Appart Probe, fournie par Abbott (North Chicago, IL). Il s'agit également d'une sonde de séparation constituée de deux signaux : un signal vert (fluorochrome Spectrum Green) couvre l'ensemble de la région IGHV et un signal rouge-orange (fluorochrome Spectrum Orange) couvre la région en aval des régions constantes d'*IGH* (voir Figure 33).

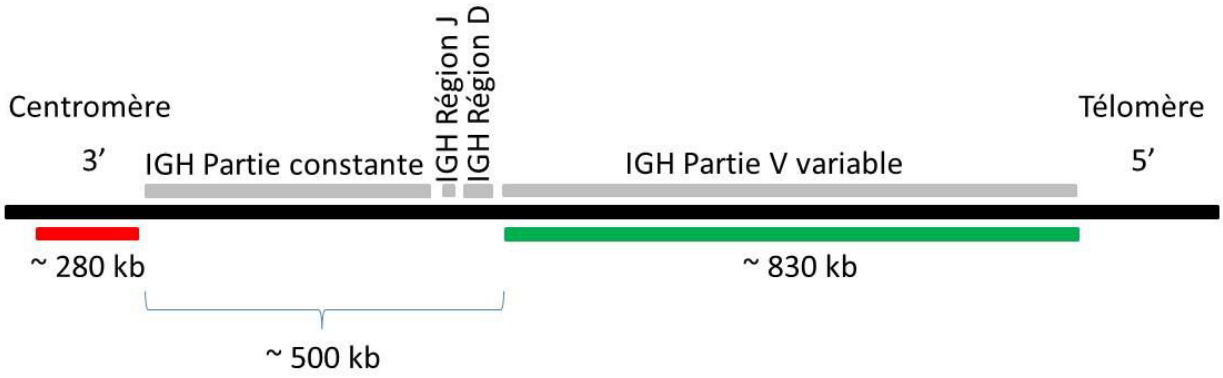

Figure 33 : Schéma de la sonde IGH LSI Vysis, adapté selon les indications du fournisseur

Une sonde ciblant le locus *IGL* a été utilisé. Il s'agit d'une sonde fournie par Dako (Agilent technologies, Santa Clara, CA). Il s'agit d'une sonde de séparation constituée d'un signal rouge ciblant la région télomérique d'IGL (segments de jonction et de régions constantes) ainsi que la région en aval du gène (Figure 34). Deux sondes marquées par de la fluorescéine (signal vert) ciblent IGLV.

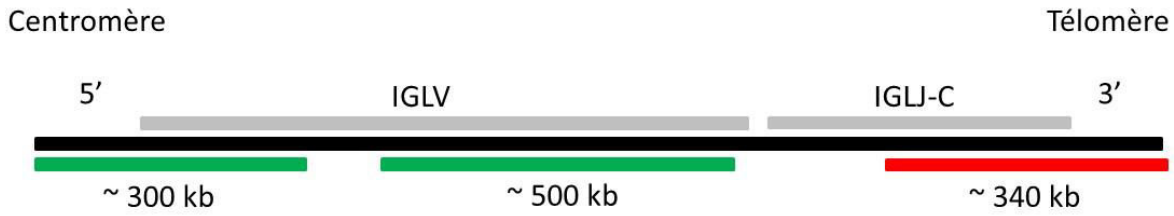

Figure 34 : Schéma de la sonde IGL Dako, adapté selon les indications du fournisseur.

Une sonde ciblant *MYC* a également été utilisée. Il s'agit de la sonde MYC FISH DNA Probe, Split signal de chez Dako (Agilent technologies, Santa Clara, CA). Cette sonde est constituée de deux signaux (vert et rouge), ciblant les régions de part et d'autres de la principale région de cassure (breakpoint cluster region) retrouvée dans le gène *MYC*, entre les exons 1 et 2 et concernant les cas de lymphome de Burkitt sporadique (Figure 35). Les deux sondes couvrent également les régions de cassure en 5' (retrouvées dans le lymphome de Burkitt endémique) et 3' (retrouvées dans certaines translocations impliquant les gènes des chaînes légères des immunoglobulines) de la breakpoint cluster region[57]. Cela permet théoriquement de détecter l'ensemble des réarrangements impliquant *MYC*.

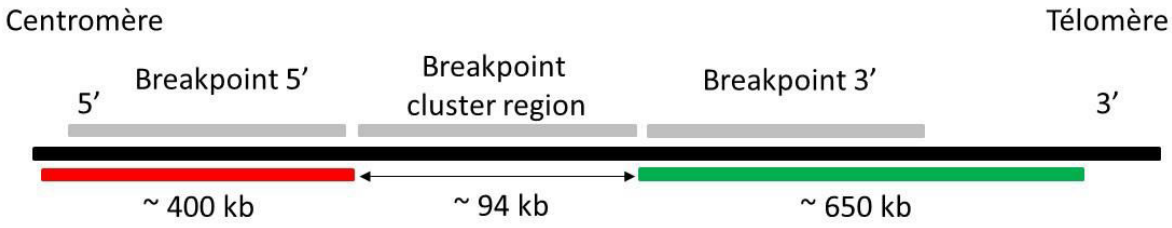

Figure 35 : Schéma de la sonde MYC Dako, adapté selon les indications du fournisseur.

Une seconde sonde MYC, fournie par Metasystems (Altlußheim, Germany) a été utilisée (Figure 36). Elle est constituée de deux signaux, un rouge et un vert situés de part et d'autre du gène *MYC*, espacées de 1,48 Mb.

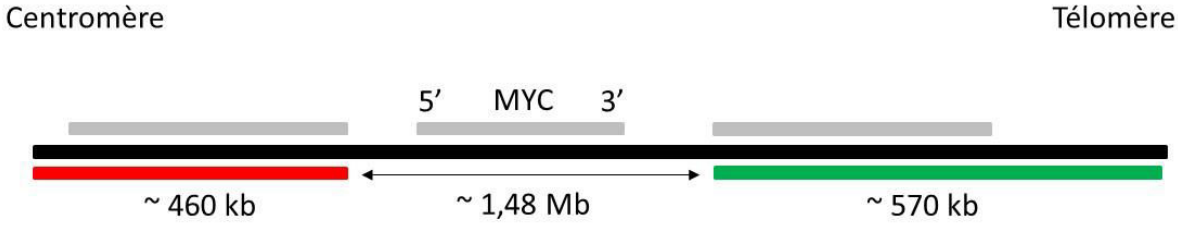

Figure 36 : Schéma de la sonde MYC Metasystem, adapté selon les indications du fournisseur.

C- Techniques spécialisées

1- FISH à façon

Afin de préciser les points de cassure retrouvés chez nos patients, nous avons dessiné des sondes de FISH à façon.

Nous avons pour cela utilisé des BACs (pour Bacterial Artificial Chromosome) et des fosmides. Ceux-ci ont été choisis à partir des bases de données afin de cibler différentes régions au niveau des chromosomes 14 et 17. La méthode utilisée La méthode utilisée est dérivée du protocole décrit par Bertrand *et al.*[602] et consiste à interroger 2 bases de données différentes afin de limiter les erreurs de localisation possibles. La spécificité des séquences est vérifiée avec NCBI BLAST[603]. Les BACs et fosmides sont ensuite commandés à BACPAC Ressources (Children's Hospital Oakland Research Institute, Oakland, CA). En parallèle, des amorces sont dessinées et commandées pour chaque sonde afin de vérifier par PCR la conformité de la séquence.

Une fois reçus, les BACs et fosmides sont préparés sont mis en culture avec des bactéries (*Escherichia coli*) sur milieu LB (pour Lysogeny Broth) pendant une nuit à 37°C. La suspension bactérienne est ensuite centrifugée et le culot obtenu est utilisé pour extraction d'ADN avec le kit « Wizard Plus SV Minipreps DNA Purification System » (Promega, Fitchburg, WI).

Les BACs ou fosmides sont ensuite marqués à l'aide d'un fluorochrome (Spectrum Orange ou Spectrum Green ou Aqua, Abbott (North Chicago, IL)) par technique de Nick translation (kit Nick Translation, Abbott (North Chicago, IL)). Cette technique utilise une DNAse entraînant des cassures simple brin de l'ADN cible (Figure 37). L'ADN polymerase I répare la brèche obtenue au niveau de l'extrémité 3'OH libre et continue l'élongation dans le 5'→3', le long du brin. Cela permet l'incorporation de dUTP fluorescents par l'ADN polymerase I. La suspension obtenue est ensuite évaporée et reprise dans 10 µL d'eau pour utilisation en FISH.

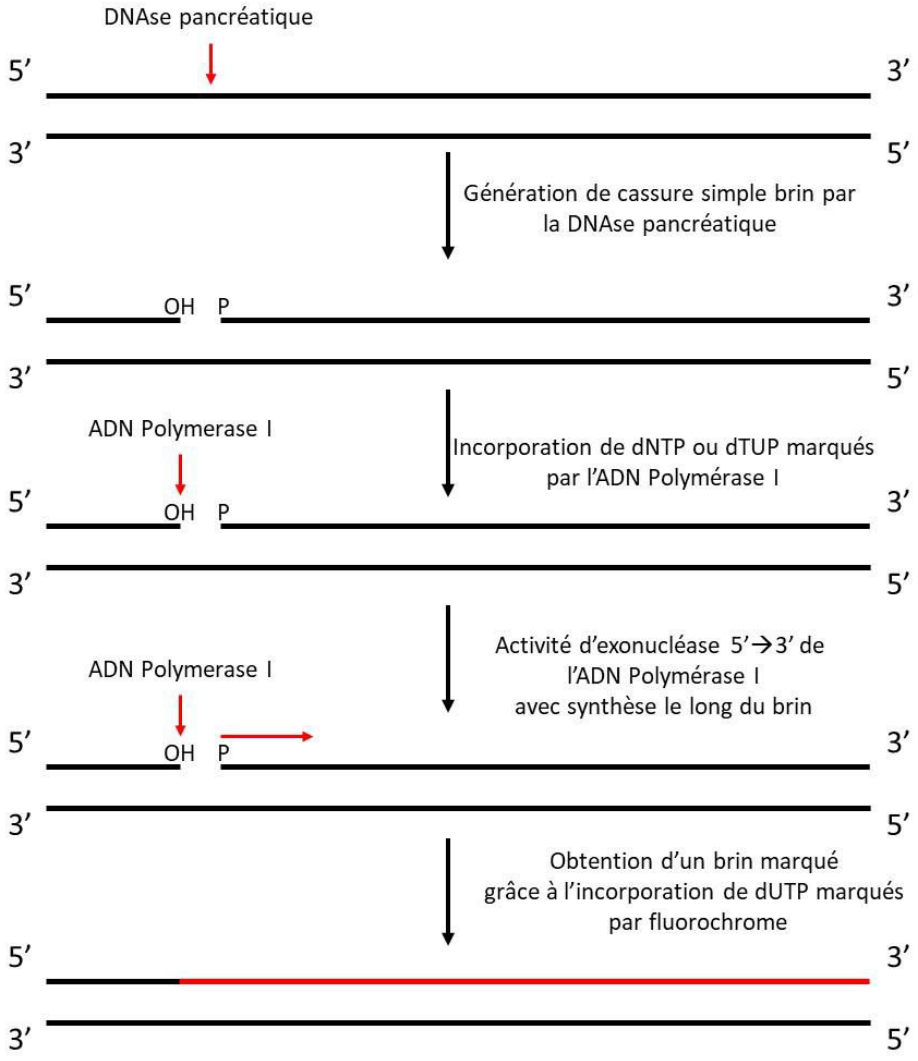

Figure 37 : Principe de la technique de Nick translation

Dans une première approche, des BACs ciblant le chromosome 17 ont été choisis. Ils ciblaient les régions 17q23 (RP11-661B17), 17q24 (RP11-387O17) et 17q25 (RP11-467J3).

Dans un deuxième temps, nous avons sélectionné des BACs ciblant la région IGHV (RP11-47P23, RP11-379F22 et RP11-72N10) (Figure 38).

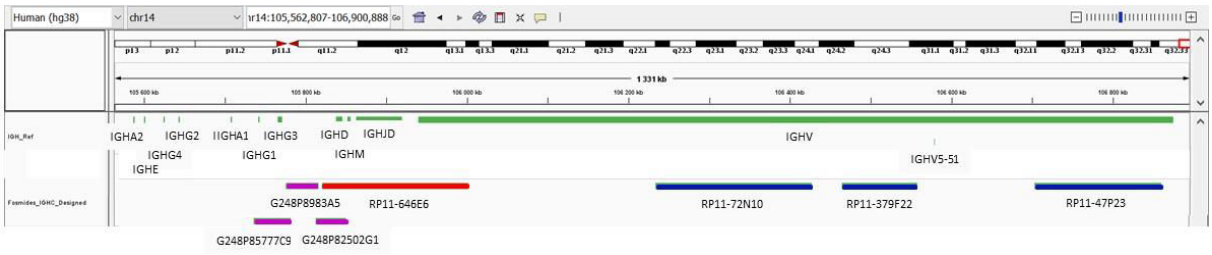

Figure 38 : Schéma de la localisation des BAC et fosmides utilisés au niveau d'IGH.
Avec en bleu les BAC utilisés au niveau d'IGHV, en violet les fosmides ciblant les régions IGHC et enfin en rouge un BAC intermédiaire.
Le segment IGHJD représente les régions de jonction et diversité.

Enfin, nous avons utilisé des fosmides ciblant les régions constantes d'IGH (G248P85777C9, G248P8983A5 et G248P82502G1) et nous avons complété avec un BAC intermédiaire ciblant la région allant d'IGHD à la portion 5' d'IGHV (RP11-645E6).

2- LM-PCR et séquençage

Pour compléter l'approcher cytogénétique, nous avons opté pour une technique de génétique moléculaire capable de détecter les points de cassure des translocations d'intérêt : la LM-PCR (pour Ligation Mediated-PCR). Cette technique a été utilisée pour la recherche de partenaires impliqués dans des translocations du TCR[604].

Cette technique consiste à digérer l'ADN à tester avec des enzymes de restriction générant des bouts francs (Figure 39). Les enzymes utilisées et les sites de restriction correspondant sont indiqués dans le Tableau 8.

Enzyme de restriction	Séquence reconnue
PvuII	CAG-CTG
EcoRV	GAT-ATC
SmaI	CCC-GGG
StuI	AGG-CCT
SspI	AAT-ATT
DraI	TTT-AAA

Tableau 8 : Enzymes de restriction utilisées pour la technique de LM-PCR et séquence reconnue correspondante (sens 5'→3'). Le tiret dans la séquence indique le site de restriction.

Après digestion enzymatique, des adaptateurs sont ligués. Ensuite deux PCR successives sont réalisées, la deuxième étant une PCR nichée. Cette approche permet d'améliorer la spécificité de la technique. Les amorces utilisées sont, d'une part des amorces spécifiques du locus connu d'intérêt, dessinées par nos soins (dans notre cas, les gènes des immunoglobulines), nommées GSP1 et GSP2 respectivement pour la PCR1 et la PCR2, et d'autre part une amorce ciblant l'adaptateur ligué (nommée AP1 pour la première PCR et AP2 pour la PCR nichée). Les tailles attendues en l'absence de remaniement sont recherchées par alignement sur le génome. La révélation se fait par migration électrophorétique sur gel d'agarose à 1,5 % contenant du Bromure d'Ethidium, avec visualisation sous lampe à Ultra-Violets. Toute bande de taille différente doit être découpée pour purification et séquençage. Par sécurité, j'ai découpé toutes les bandes visibles. Enfin, chaque bande est purifiée à l'aide du kit Nucleospin Gel and PCR clean-up (Macherey Nagel, Düren, Germany) et une réaction de séquence est réalisée en sens et antisens (amorces GSP2 et AP2) avec du BigDye Terminator v1.1 (Applied, Thermo Fisher, Waltham, MA). La purification est alors faite à l'aide de colonnes de Sephadex G50 (VWR, Radnor, PA). Enfin, les séquences sont passées sur séquenceur capillaire ABI-Prism 3130xl (Thermo Fisher, Waltham, MA).

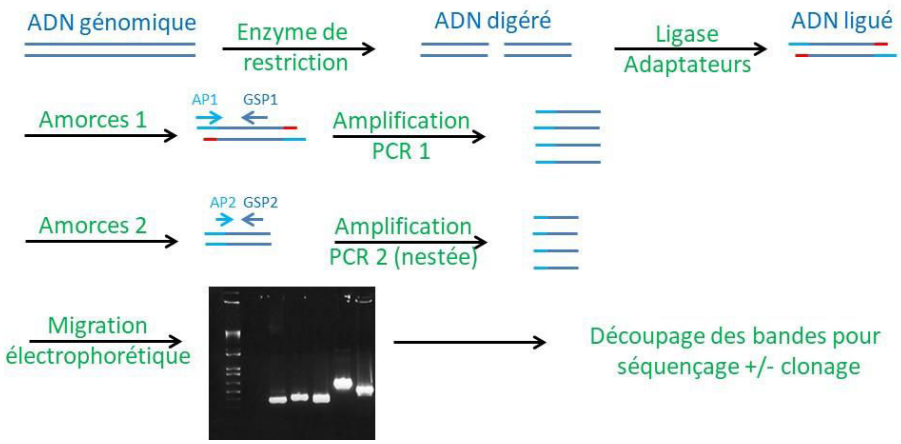

Figure 39 : Principe de la LM-PCR
AP1 et AP2 (pour Adaptator Primer 1 ou 2) sont les amorces ciblant l'adaptateur ligué ; GSP1 et GSP2 (pour Gene Specific Primer 1 ou 2) sont les amorces spécifiques du locus d'intérêt connu.

Le kit utilisé pour la réalisation de la LM-PCR est le Universal Genome Walker 2.0 (Takara Bio, Kyoto, Japan).

Nous avons dessiné des amorces au niveau des points de cassure potentiels tels que décrits dans la littérature. Concernant *IGH*, il s'agit principalement des régions IGHJ et IGHM (au niveau de l'enhancer et du Switch µ ou à proximité), mais aussi des régions switch des autres régions constantes. Concernant *IGL*, les points de cassure concernent les différents gènes de jonction. Nous avons ainsi dessiné plusieurs centaines d'amorces correspondant aux spécifications de la technique, à savoir :

- une longueur de 26 à 30 nucléotides ;

- un pourcentage en GC compris entre 40 et 60 % ;

- l'absence d'hybridation côté 3' de l'amorce avec l'adaptateur ;

- un maximum de 3 G/C dans les 6 dernières positions en 3' de l'amorce.

De plus, dans la mesure du possible, l'amorce nichée ne doit pas être chevauchante avec la première amorce. Si le chevauchement est obligatoire, le côté 3' de l'amorce nichée doit avoir une séquence la plus unique possible.

3- CGH-array

La CGH-array (pour Comparative Genomic Hybridization-array) est une technique de cytogénétique moléculaire. Elle repose sur l'hybridation d'un ADN test marqué par un fluorochrome (en général vert) en comparaison à un ADN normal (en général marqué par un fluorochrome rouge) sur une puce contenant un grand nombre de sondes ciblant les différents chromosomes ou régions d'intérêt dans certaines pathologies. L'hybridation est réalisée sur lame. Différentes résolutions existent.

Nous avons réalisé une analyse de CGH-array pour le patient n°1 de l'étude des remaniements atypiques impliquant les gènes des immunoglobulines. Cette technique a été réalisée avec la technologie Agilent (Agilent Technologies, Santa Clara, CA). Une puce 400k a été utilisée.

4- Séquençage haut débit (NGS)

Dans ce travail, nous avons utilisé deux technologies différentes pour la réalisation du NGS.

Concernant l'étude des translocations impliquant les gènes des immunoglobulines, une approche de séquençage du génome entier « paired-end » a été réalisée sur HiSeq 4000 Illumina (San Diego, CA). La réalisation de ce séquençage a été sous-traitée à Integragen (Evry, France).

Pour l'étude réalisée dans le cadre de l'étude de l'impact pronostique de l'électrophorèse des protéines sériques dans la LLC, deux panels ciblés de gènes ont été mis au point pour la technologie IonProton (ThermoFisher, Waltham, MA).

a- Technologie Illumina

i- Principe

La technologie de NGS Illumina repose sur un séquençage par synthèse, utilisant l'incorporation de nucléotides marqués par un fluorophore, suivie en temps réel par caméra CCD. L'utilisation de « paired-end », c'est-à-dire d'un séquençage dans les deux sens des fragments d'ADN, permet d'améliorer la détection de réarrangements.

ii- Application au séquençage « whole-genome »

Le séquençage « whole genome », c'est-à-dire du génome entier, a été réalisé pour les trois patients d'intérêt retenus dans l'étude de réarrangements atypiques impliquant les gènes des immunoglobulines. L'ADN génomique des patients a été utilisé pour la réalisation de librairies avec obtention de fragments d'environ 400 pb.

Pour les patients n°2 et 3, nous disposions de prélèvements largement infiltrés (plus de 90 % de cellules porteuses de l'anomalie), et une profondeur de lecture de 15-20x a été appliquée.

Concernant le patient n°1, nous avons utilisé de l'ADN extrait à partir de ganglion qui était plus infiltré que le sang. Sur ce prélèvement, environ 50 % des cellules étaient porteuses de l'anomalie. Après discussion avec Integragen, une profondeur de 15-20x a semblé suffisante pour la détection du remaniement chez ce patient.

Les données générées par le séquençage ont été alignées sur le génome de référence GRCh38.

iii- Analyse MANTA

L'analyse des données générées par le séquençage du génome entier pour la détection des remaniements chromosomiques a été réalisée grâce au logiciel MANTA. Il s'agit d'un logiciel développé par Illumina (San Diego, CA) permettant la détection de variants structuraux dont les insertions-délétions et les translocations chromosomiques.

Le logiciel MANTA utilise les fichiers bam (pour Binary Alignment Map) générés par le séquençage et procède en deux étapes[605]. Dans un premier temps, un graphique de toutes les associations de cassure dans le génome est construit. Dans un second temps, les composants de ce graphique sont travaillés pour la génération d'hypothèses de variants, l'assemblage, la génération d'un score et d'un rapport sous format vcf (pour variant call format). Ce logiciel permet d'obtenir un rapport en un temps limité (jusqu'à 4-5h contre jusqu'à 4-5 jours avec d'autres logiciels)[605].

b- Technologie Ion Proton

i- Principe

La Technologie IonProton a été développée par Life Technologies (ThermoFisher, Waltham, MA). Il s'agit d'une technique de NGS par synthèse, avec détection en temps réel de l'incorporation des bases entraînant la libération d'un proton. C'est la modification du pH qui est alors mesurée.

Dans notre cas, le séquençage nécessite la préparation de librairies par amplification clonale des régions d'intérêt à étudier. Pour cela, nous avons mis au point deux panels Ampliseq.

La réalisation des librairies a été faite avec le « Ion AmpliSeq Library Kit 2.0 » (ThermoFisher, Waltham, MA) selon les recommandations du fournisseur.

Les échantillons ont ensuite été analysés sur un séquenceur IonProton (ThermoFisher, Waltham, MA) au niveau de la plateforme de séquençage du CHU de Limoges.

ii- Définition des gènes d'intérêt

Nous avons mis au point un premier panel de NGS pour la détection des mutations d'intérêt diagnostique et/ou pronostique dans les lymphomes. Nous avons fait une revue de la littérature et suivi les recommandations des groupes de travail afin de retenir les gènes et régions d'intérêt.

Les gènes d'intérêt dans la LLC intégrés dans ce premier panel sont *TP53*, *ATM*, *NOTCH1*, *SF3B1*, *BIRC3*, *BRAF*, *FBXW7*, *MYD88*, *XPO1*, *POT1*, *EGR2*, *NKBIE, DDX3X, RPS15*, *PLCG2* et *BTK* (Tableau 9).

Le deuxième panel dessiné ciblait uniquement *TP53* et *ATM* pour la recherche de mutations.

iii- Définition d'amplicons de référence pour l'analyse de déséquilibres génétiques

Dans l'optique d'utiliser notre panel de NGS pour détecter les variations du nombre de copies dans la LLC, des régions génomiques spécifiques ont été intégrées dans le panel Ampliseq lymphoïde (Tableau 9). Le séquençage de certains gènes tels que *TP53* ou *ATM* permet de couvrir ces régions d'intérêt (détection de la del(17p) et de la del(11q)) avec suffisamment d'amplicons. D'autres gènes ont été ajoutés pour assurer la détection de la del(13q) (*DLEU1*, *DLEU2*, *MIR16-1*, *MIR15A* et *RB1*) et de la trisomie 12 (*POU6F1*, *MDM2*, 2 cibles utilisées en QMPSF). D'autres régions ont été ajoutées au niveau du chromosome 8 pour détecter les gains 8q (*EIF3H*, *MYC*), et au niveau des chromosomes 18 (*MYOM1*, *DCC* et *BCL2*) et 19 (*CACNA1A et CCNE1*) pour la détection de ces trisomies. Enfin, des régions contrôles supposées normales ont été intégrées (chromosome 3 avec *EPHA3*, *PROS1* et *TBL1XR1*, chromosome 4 avec *ZAR1*, chromosome 7 avec *SUMF2* et *ASL*, chromosome 9 avec un amplicon ciblant *JAK2*, et chromosome 22 avec un amplicon ciblant *CECR1*).

Chromosome	Gène	Région	Référence	Cible pour CNV seulement	Cible pour mutation récurrente dans la LLC	Cible pour mutation récurrente dans autres lymphomes
Chr1	SPEN	E11	NM_015055			X
Chr1	ARID1A	from E1 to 20	NM_006015			X
Chr1	NOTCH2	from E26 to 28 + E34	NM_024408			X
Chr2	REL	from E1 to E10 (with UTR)	NM_001291746			X
Chr2	XPO1	E14, E15	NM_003400		X	
Chr2	CXCR4	E1, E2	NM_003467			X
Chr2	SF3B1	from E1 to E25	NM_012433		X	
Chr3	MYD88	from E3 to E5	NM_001172567		X	
Chr3	EPHA3	E1, E3, E5, E7, E9, E11, E13, E15, E17	NM_005233	X		
Chr3	PROS1	E2, E4, E6, E8, E9, E10, E11, E13, E15	NM_000313	X		
Chr3	TBL1XR1	E3, E6, E7, E8, E12	NM_024665	X		
Chr3	KLHL6	E6, E7	NM_130446			X
Chr4	ZAR1	from E1 to E4	NM_175619	X		
Chr4	RASSF6	from E1 to E11	NM_201431			X
Chr4	PTPN13	from E2 to E48	NM_080685			X
Chr4	FAT4	E1	NM_024582			X
Chr4	PCDH10	E1	NM_032961			X
Chr4	FBXW7	from E2 to E12	NM_033632		X	
Chr5	EGR1	E1, E2	NM_001964			X
Chr6	IRF4	from E2 to E9	NM_002460			X
Chr6	NFKBIE	from E1 to E6	NM_004556		X	
Chr6	PRDM1	from E1 to E7	NM_001198			X
Chr6	TNFAIP3	from E2 to E9	NM_006290			X
Chr7	CARD11	from E2 to E9	NM_032415			X
Chr7	SUMF2	from E1 to E9	NM_015411	X		
Chr7	ASL	E1, E2, E4, E6, E8, E10, E12,E14, E15, E16, E17	NM_000048	X		
Chr7	POT1	from E5 to E19	NM_015450		X	
Chr7	BRAF	E11, E15	NM_004333		X	
Chr7	EZH2	E16, E18	NM_004456			X
Chr8	IKBKB	from E2 to E22	NM_001556			X
Chr8	EIF3H	from E1 to E8	NM_003756	X		
Chr8	MYC	from E1 to E3	NM_002467			X
Chr9	JAK2	E7	NM_004972	X		
Chr9	NOTCH1	from E25 to E28 + E34	NM_017617		X	
Chr10	EGR2	E3, E4	NM_001136178		X	
Chr11	SWAP70	from E1 to E3	NM_015055			X
Chr11	BIRC3	from E6 to E9	NM_001165		X	
Chr11	ATM	from E1 to E62	NM_000051		X	
Chr12	SOX5	from E1 to E15	NM_006940			X
Chr12	KMT2D	from E1 to E54	NM_003482			X
Chr12	POU6F1	from E2 to E6	NM_002702	X		
Chr12	MDM2	E1, E3, E5, E7, E9, E11	NM_002392	X		
Chr12	DTX1	E2	NM_004416			X
Chr13	RB1	E2, E4, E6, E8, E10, E12, E14, E16, E18, E20, E22, E24, E26	NM_000321	X		
Chr13	DLEU2	E1, E2, E4, E5, E7, E9, E10		X		
Chr13	Mir16-1	E1		X		
Chr13	Mir15A	E1		X		
Chr13	DLEU1	from E1 to E6		X		
Chr14	TRAF3	E3, E9, E10, E12	NM_145725			X
Chr15	SIN3A	E4, E8	NM_001145357			X
Chr16	CREBBP	from E1 to E31	NM_004380			X
Chr16	PLCG2	from E18 to E23	NM_002661		X	
Chr17	ATP2A3	E2, E5, E8, E11, E14, E17, E20, E23	NM_174953	X		
Chr17	TP53	from E2 to E11	NM_000546		X	
Chr17	MAP3K14	E12	NM_003954	X		
Chr17	CD79B	from E1 to E6	NM_001039933			X
Chr18	MYOM1	E2, E6, E8, E10, E12, E14, E16, E20, E22, E24, E26, E28, E30, E32, E34, E36, E38	NM_003803	X		
Chr18	DCC	E1, E3, E5, E7, E9, E11, E13, E15, E17, E19, E21, E23, E25, E29	NM_005215	X		
Chr18	BCL2	E2, E3	NM_000633			X
Chr18	RPS15	from E1 to E4	NM_001018		X	
Chr19	CACNA1A	E1, E4, E7, E10, E13, E16, E19, E41, E44, E47	NM_000068	X		
Chr19	KLF2	from E1 to E3	NM_016270			X
Chr19	CCNE1	from E1 to E11	NM_001238	X		
Chr19	CD79A	from E1 to E5	NM_001783			X
Chr22	CECR1	E2	NM_001282228	X		
Chr22	MAPK1	from E1 to E8	NM_002745			X
Chr22	XBP1	from E1 to E5	NM_005080			X
ChrX	DDX3X	from E1 to E17	NM_001356		X	
ChrX	ZMYM3	from E2 to E25	NM_001171162			X
ChrX	BTK	from E14 to E16	NM_000061		X	

Tableau 9 : Liste des cibles inclues dans le panel NGS lymphoïde.

Concernant le panel réduit à *TP53* et *ATM*, le but était la détection de la del(11q) et de la del(17p) uniquement. Des gènes contrôles ont été ajoutés : au niveau du chromosome 11 (*SWAP70*, gène situé en 11p, permettant de différencier une del(11q) d'une monosomie 11), au niveau du chromosome 17 (*MAP3K14*, *CD79B*, gènes situés en 17q). Les gènes contrôles

normaux sont *JAK2* sur le chromosome 9, *CECR1* sur le chromosome 22. De plus, un gène situé sur le chromosome Y (*SRY*) a été intégré afin d'améliorer la normalisation, y compris pour des puces déséquilibrées en terme de sex-ratio, et de détecter automatiquement les hommes et les femmes.

iv- Mise au point de panels NGS

Le panel Ampliseq lymphoïde commandé a été développé sur la plateforme en ligne AmpliSeq Designer. Il couvre 221,6 kb. 70 gènes ont été intégrés (soit exome, soit régions ciblées). Il s'agit d'un design en 2 pools avec un total de 1206 amplicons. Afin d'améliorer la couverture obtenue, en particulier au niveau de certains gènes d'intérêt, nous avons choisi une option de spécificité relâchée. La couverture théorique obtenue est de 97,68 %.

Un premier test avait été réalisé avec 5 patients par puce. La profondeur obtenue au niveau de *TP53* semblait suffisante avec cette configuration (minimum de 100 lectures nécessaires d'après les recommandations[353]) : la profondeur moyenne obtenue par amplicon était alors d'environ 9000X. Dans cette configuration, les régions d'intérêt les moins bien couvertes (telles que l'exon 5 de *TP53*) ont été séquencées avec une profondeur minimale d'environ 300X. Nous avons ensuite testé nos différents échantillons avec une configuration avec 7 à 8 patients par puce, permettant d'obtenir une profondeur de lecture moyenne d'environ 7000X par amplicon, avec un minimum de 100X pour l'exon 5 de *TP53*.

Le panel Ampliseq réduit à *TP53* et *ATM* a été développé de la même façon. Il couvre 30,22 kb. Il s'agit d'un design en 2 pools avec un total de 172 amplicons. La couverture théorique obtenue est de 96,39 %. La spécificité retenue est haute, la couverture des régions d'intérêt étant théoriquement suffisante. 50 patients ont été passés simultanément par puce. La couverture au niveau de *TP53* était suffisante (minimum de 100 lectures).

v- Logiciel CovCop

Le logiciel CovCop a été développé par les Drs Paco Derouault et Anne-Sophie Lia de l'EA6309 de la faculté de Médecine de Limoges. Son principe a été publié dans la revue Bioinformatics en 2016[606]. Il a été spécialement développé pour la détection des variations du nombre de copies après NGS basé sur la réalisation de librairies par amplicons (amplification de gènes ou régions géniques ciblées) telles que celles obtenues avec nos panels Ampliseq. Le principe de

ce logiciel est basé sur l'analyse du nombre de lectures par amplicon et par patient au sein d'une même puce de séquençage. Plusieurs étapes de correction et de normalisation sont réalisées. Tout d'abord, les nombres de lectures par amplicon de chaque échantillon sont corrigées par le total des amplicons de chaque librairie. Une correction par rapport au pourcentage en GC et la longueur de chaque amplicon est ensuite réalisée. En effet, ces paramètres influencent l'efficacité d'amplification par PCR. L'étape de normalisation est ensuite appliquée : les amplicons ayant la variance la plus faible sont sélectionnés automatiquement comme amplicons de référence. Un ratio normalisé est alors généré pour chaque amplicon en divisant leur nombre de lectures par la médiane des lectures des amplicons de référence sélectionnés automatiquement. En l'absence d'anomalie de nombre, le ratio normalisé est centré sur la valeur 1. Ce ratio prendra la valeur théorique de 1,5 en cas de trisomie ou de gain d'un allèle et tendra vers la valeur de 0,5 en cas de délétion mono-allélique.

Il est possible d'assigner à chaque patient le sexe (homme ou femme) afin d'éviter de mauvaises interprétations d'amplicons situés sur le chromosome X. Il est également possible d'intégrer des amplicons situés sur le chromosome Y pour permettre une détection automatique. CovCop corrige alors les valeurs des amplicons situés sur le chromosome X en multipliant par 2 les valeurs obtenues chez les hommes. Ceci peut induire un biais qui reste limité si le nombre d'amplicons situés sur le chromosome X est faible. Une autre façon de s'affranchir de ce biais est de réaliser des puces ne contenant que des échantillons provenant d'hommes ou de femmes. Le seuil de détection des variations du nombre de copies peut être ajusté par l'utilisateur. Un minimum de 4 amplicons successifs variant dans le même sens est nécessaire pour considérer un locus comme potentiellement délété ou dupliqué.

Cov'Cop a été initialement développé pour la recherche d'anomalies constitutionnelles. Un minimum de 3 échantillons par puce comprenant au moins 100 amplicons chacun est nécessaire pour un fonctionnement optimal de Cov'Cop dans cette indication.

vi- Utilisation de la fréquence allélique des variants pour la détection des CNV

Le NGS permet d'identifier de nombreux SNP. La fréquence allélique de ces variants (ou VAF ou Variant Allelic Frequency) dans l'échantillon est d'environ 50 % en cas d'hétérozygotie, et de 100 % en cas d'homozygotie. Une valeur déviante de 50 % peut être observée en cas de CNV. Nous avons donc utilisé la VAF déterminée par NGS pour la détection des CNV. Le panel NGS utilisé est le panel de 70 gènes.

Nous avons mis au point et validé un algorithme (script R) permettant de sélectionner automatiquement les SNP détectés par NGS à partir des fichiers vcf comprenant les différents variants rapportés par NGS, et annotés à partir des bases de données des SNP (1000Genomes

et dbSNP). Le critère de sélection des SNP est une fréquence allélique dans la population générale supérieure à 1 %. Certains SNP ayant des valeurs aberrantes (c'est-à-dire des valeurs éloignées systématiquement de 50 ou 100 %, en lien avec des difficultés de séquençage) ont été éliminés. Les données issues des 82 patients de l'étude sont utilisées pour calculer la moyenne et l'écart-type des différents SNP situés sur un même amplicon. Des courbes ROC (pour Receiving Operator Characteristic) ont été construites afin de déterminer le seuil de normalité pour les amplicons. Celui-ci a été choisi afin d'obtenir une bonne spécificité et a été fixé à +/- 2,5 écart-types par rapport à la moyenne. Pour les SNP représentés moins de 5 fois dans la base de données, la moyenne et l'écart-type des SNP de tous les échantillons est appliquée. Le choix d'un seuil à +/- 2,5 écart-type a été fixé afin de privilégier la spécificité du test par rapport à la sensibilité de détection d'une anomalie.

Dans ce script nous avons inclus une représentation graphique des VAF avec les bornes de normalité par amplicon et l'analyse visuelle rapide des résultats, facilitant le repérage des anomalies présentes dans l'échantillon. Un package R est en cours de réalisation.