

Faculté de Pharmacie

Année 2019

Thèse N°

Thèse pour le diplôme d'État de docteur en Pharmacie

Présentée et soutenue publiquement

le 11 octobre 2019

par

Marine Dupont

née le 4 juin 1991 à Limoges

Mise au point du séquençage haut débit des gènes des immunoglobulines et application à la recherche de BCRs stéréotypés de type « LLC » chez des patients exempts d'hémopathie

Thèse dirigée par Dr David RIZZO

Examineurs de la thèse :

Madame le Professeur Sylvie Rogez.....Présidente du Jury
Monsieur le Professeur Jean Feuillard.....Jury
Madame le Professeur Marie-Odile Jauberteau - Marchan Jury
Madame le Docteur Jasmine ChauzeixJury
Monsieur le Docteur David Rizzo.....Directeur de thèse



Faculté de Pharmacie

Année 2019

Thèse N°

Thèse pour le diplôme d'État de docteur en Pharmacie

Présentée et soutenue publiquement

le 11 octobre 2019

par

Marine Dupont

née le 4 juin 1991 à Limoges

Mise au point du séquençage haut débit des gènes des immunoglobulines et application à la recherche de BCRs stéréotypés de type « LLC » chez des patients exempts d'hémopathie

Thèse dirigée par Dr David RIZZO

Examineurs de la thèse :

Madame le Professeur Sylvie Rogez.....Présidente du Jury
Monsieur le Professeur Jean Feuillard.....Jury
Madame le Professeur Marie-Odile Jauberteau – Marchan.....Jury
Madame le Docteur Jasmine ChauzeixJury
Monsieur le Docteur David Rizzo.....Directeur de thèse





Liste des enseignants

Le 1^{er} septembre 2017

PROFESSEURS :

BATTU Serge	CHIMIE ANALYTIQUE
CARDOT Philippe	CHIMIE ANALYTIQUE ET BROMATOLOGIE
DESMOULIERE Alexis	PHYSIOLOGIE
DUROUX Jean-Luc	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE
FAGNERE Catherine	CHIMIE THERAPEUTIQUE - CHIMIE ORGANIQUE
LIAGRE Bertrand	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MAMBU Lengo	PHARMACOGNOSIE
ROUSSEAU Annick	BIOSTATISTIQUE
TROUILLAS Patrick	CHIMIE PHYSIQUE - PHYSIQUE
VIANA Marylène	PHARMACOTECHNIE

PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS DES DISCIPLINES PHARMACEUTIQUES :

PICARD Nicolas	PHARMACOLOGIE
ROGEZ Sylvie	BACTERIOLOGIE ET VIROLOGIE
SAINT-MARCOUX Franck	TOXICOLOGIE

ASSISTANT HOSPITALIER UNIVERSITAIRE DES DISCIPLINES PHARMACEUTIQUES :

CHAUZEIX Jasmine	HEMATOLOGIE (Renouvelé jusqu'au 1 ^{er} novembre 2018)
JOST Jérémy	PHARMACIE CLINIQUE (1 ^{er} novembre 2016 pour 2 ans)

MAITRES DE CONFERENCES :

BASLY Jean-Philippe	CHIMIE ANALYTIQUE ET BROMATOLOGIE
BEAUBRUN-GIRY Karine	PHARMACOTECHNIE

BILLET Fabrice	PHYSIOLOGIE
CALLISTE Claude	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE
CHEMIN Guillaume	BIOCHIMIE FONDAMENTALE
CLEDAT Dominique	CHIMIE ANALYTIQUE ET BROMATOLOGIE
COMBY Francis	CHIMIE ORGANIQUE ET THERAPEUTIQUE
COURTIOUX Bertrand	PHARMACOLOGIE, PARASITOLOGIE
DELEBASSEE Sylvie	MICROBIOLOGIE-PARASITOLOGIE-IMMUNOLOGIE
DEMIOT Claire-Elise	PHARMACOLOGIE
FROISSARD Didier	BOTANIQUE ET CRYPTOLOGAMIE
GRIMAUD Gaëlle	CHIMIE ANALYTIQUE ET CONTROLE DU MEDICAMENT
JAMBUT Anne-Catherine	CHIMIE ORGANIQUE ET THERAPEUTIQUE
LABROUSSE Pascal	BOTANIQUE ET CRYPTOLOGAMIE
LEGER David	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MARION-THORE Sandrine	CHIMIE ORGANIQUE ET THERAPEUTIQUE
MARRE-FOURNIER Françoise	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MERCIER Aurélien	PARASITOLOGIE
MILLOT Marion	PHARMACOGNOSIE
MOREAU Jeanne	MICROBIOLOGIE-PARASITOLOGIE-IMMUNOLOGIE
MUSUAMBA TSHINANU Flora	PHARMACOLOGIE
PASCAUD Patricia	PHARMACIE GALENIQUE – BIOMATERIAUX CERAMIQUES
POUGET Christelle	CHIMIE ORGANIQUE ET THERAPEUTIQUE
VIGNOLES Philippe	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE

Remerciements

A Madame le Professeur Sylvie Rogez, je vous remercie d'avoir accepté de présider ce jury et pour votre accompagnement tout au long de mon internat. Veuillez recevoir ma reconnaissance et mon profond respect.

A Monsieur le Professeur Jean Feuillard, je vous remercie d'avoir accepté de juger ce travail, veuillez recevoir ma reconnaissance et mon respect. Je vous remercie également pour vos conseils tout au long de mon internat et de m'avoir accueillie au sein de votre service pendant plusieurs semestres.

A Madame le Professeur Marie-Odile Jauberteau-Marchan, je vous remercie d'avoir accepté de juger ce travail et pour le semestre passé dans votre laboratoire. Veuillez trouver ici l'expression de mes sincères remerciements et de mon profond respect.

A Madame le Docteur Jasmine Chauzeix, je te remercie d'avoir accepté de juger ce travail et pour l'intérêt que tu portes à ce travail. Je tiens à t'exprimer toute ma gratitude et mon respect.

A Monsieur le Docteur David Rizzo, je te remercie d'avoir accepté de te lancer dans cette aventure avec moi. Merci pour ta patience, ta bienveillance, surtout ta disponibilité et ton optimisme qui ont permis la réussite de ce travail. Ce fut un plaisir et une chance de travailler avec toi. Pour cela, je tiens à t'exprimer ma reconnaissance, mon admiration et mon profond respect.

Je tiens à remercier Nathalie pour votre disponibilité, votre soutien au cours de ce travail, particulièrement à un moment critique.

A Alexis pour ton aide dans la mise en route de cette nouvelle technique.

A Paco pour ton aide et ton efficacité.

Aux biologistes du laboratoire d'hématologie du CHU de Limoges, Estelle avec ton dynamisme quotidien et les après-midis de lecture de myélogramme, Franck pour votre savoir particulièrement en cytologie et Magali pour tes connaissances en hémostase.

Je remercie particulièrement l'ensemble des techniciens du laboratoire d'hématologie du CHU de Limoges. Vous m'avez accueillie depuis mon 2^{ème} semestre. Petite dédicace toute particulière au secteur de biologie moléculaire, avec Mélanie, Adélaïde, Marie-Pierre, Marjorie, Marie-Candys et Virginie pour votre bonne humeur et votre aide qui m'ont permis d'accomplir ce travail.

Je remercie aussi le secteur de cytométrie en flux avec Patricia, Patrick, Stéphane et Anne de m'avoir laissé faire mes petits marquages et pour votre aide dans ce travail.

Je n'oublie pas les autres secteurs notamment la cytologie avec Alexia, Elodie, Sylvie, Marlène Jaja, Mumu, vous m'avez appris à reconnaître les cellules quand je suis arrivée en hématologie. Merci à l'hémostase Amaury, Delphine, Claire, Hilal pour votre apprentissage, je me rappellerai toujours du Von Kaulla et des cofacteurs ! Enfin, merci à la cytogénétique Mélanie, Catherine, Anne et Virginie pour votre disponibilité et votre savoir. Merci aussi à ceux que j'oublie...

Merci à tous les biologistes et technicien(ne)s des services dans lesquels j'ai réalisé mon internat, le service de Parasitologie-Mycologie, le service de Bactériologie-Virologie-Hygiène, le service d'Immunologie du CHU de Limoges ainsi que le service de Biochimie de Rangeuil du CHU de Toulouse.

Merci à l'équipe de recherche INSERM 1035, particulièrement Jean Max pour votre encadrement au cours de mon année recherche.

Un merci tout particulier à mes co-internes, notamment Nicolas, Caro, Léa, Malika et Guigui de m'avoir libéré du temps pour terminer ce travail. Je n'oublie pas les co-internes des semestres précédents : Virginie, Sarah, Chacha, François, Yoyo, Sylvain, Thibault, Florian, Marie-Céleste, Marc, Seb, Jacques, Audrey, Elodie, Damien, Alice, petit clin d'œil à la team rex (Charlycorne, Mymy, Thibaut, Claudia et MF), et à ceux que j'oublie ...

A Marie-Fleur, ma coupine d'internat, tout simplement merci pour ton soutien sans faille au cours de cet internat.

A Mylène, merci pour ta bonne humeur au quotidien et ton énergie.

Petite dédicace au groupe de travail qui s'est installé au laboratoire au cours de ces dernières semaines. Sans vous la fin aurait vraiment été encore plus dur. Courage Josselin ça sera bientôt ton tour. Merci pour votre bonne humeur et votre « précision » !

A Jérôme, tu m'as fait confiance tout au long de mon internat je te remercie.

Merci à toi Édith pour ton soutien tout au long de ces années d'internat.

A Amandine, ma grande sœur toujours disponible ; mais aussi Antoine ainsi que mon petit filleul Loulou et mes deux nièces Clara et Camillou que j'adore.

A mes deux mémés, merci d'être toujours aux petits soins avec moi. A pépé, tu es toujours avec moi, je ne t'oublie pas...

Papa, Maman merci à vous, vous m'avez toujours soutenu depuis le début de cette aventure. Sans vous, rien de tout ça n'aurait été possible. Merci pour votre amour et votre soutien.

Guillaume, pour ta patience, ton soutien et ta présence à mes côtés merci ... Ne t'aime ...

Droits d'auteurs

Cette création est mise à disposition selon le Contrat :

« **Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France** »

disponible en ligne : <http://creativecommons.org/licenses/by-nc-nd/3.0/fr/>



Table des matières

Liste des abréviations	18
Introduction	19
I. Structure et génétique des immunoglobulines.....	20
I.1. Structure des immunoglobulines.....	20
I.1.1. Les chaînes lourdes « <i>Heavy</i> »	20
I.1.2. Les chaînes légères « <i>Light</i> ».....	21
I.2. Organisation des gènes des immunoglobulines	22
I.3. Immunoglobuline membranaire et récepteur à l'antigène du lymphocyte B (BCR)	23
II. Ontogénie du lymphocyte B.....	24
II.1. Lymphopoïèse B indépendante de l'antigène	24
II.1.1. Stade pré-pro-B	25
II.1.2. Stade pro-B	25
II.1.3. Mécanisme de la recombinaison V(D)J	26
II.1.4. Stade pré-B	28
II.1.5. Stade B immature	29
II.1.6. Stade B transitionnel	29
II.2. Lymphopoïèse B dépendante de l'antigène	30
II.2.1. Réponse à l'antigène T dépendant.....	30
II.2.2. Réponse à l'antigène T indépendant.....	32
III. La Leucémie Lymphoïde Chronique (LLC).....	33
III.1. Définition	33
III.2. Épidémiologie.....	33
III.3. Diagnostic biologique	33
III.4. Pronostic	34
III.4.1. Classification clinico-biologique.....	34
III.4.2. Anomalies cytogénétiques de la LLC	35
III.4.3. Biologie moléculaire	35
III.4.4. Statut mutationnel des immunoglobulines	37
III.4.5. Répertoire IGHV des LLC et BCR stéréotypé	38
III.5. Histoire naturelle de la LLC.....	40
III.6. Traitements de la LLC	42
III.6.1. Les indications de traitement.....	42
III.6.2. Les traitements	43
III.6.3. Les résistances aux traitements ciblés.....	45
IV. Objectifs de la thèse	46
V. Matériels et Méthodes	47
V.1. Consentement.....	47
V.2. Constitution d'un pool de témoins sains et de la cohorte de patients.....	47
V.3. Cytométrie en flux.....	47
V.4. Tri cellulaire.....	47
V.5. Extraction d'ADN.....	47
V.6. Vérification du profil de clonalité par analyse de fragment	48
V.7. Vérification de la qualité des ADN utilisés	48
V.8. Préparation des bibliothèques	48

V.8.1. Principe de préparation des librairies par PCR	48
V.8.2. PCR-1	49
V.8.2.1. Amorces Leader VH	49
V.8.2.2. Amorces FR1 BIOMED-2	49
V.8.2.3. Mélange réactionnel et programme d'amplification	50
V.8.3. Étapes de purification	51
V.8.4. PCR-2	51
V.9. Qualification des librairies	52
V.9.1. Vérification de la présence d'un produit spécifique	52
V.9.2. Quantification ADN	52
V.10. Séquençage Illumina	52
V.10.1. Principe du séquençage Illumina	52
V.10.2. Préparation des librairies au séquençage	54
V.11. Analyse bio-informatique post-séquençage	54
V.11.1. Analyse de la qualité des <i>reads</i> obtenus	54
V.11.2. Assemblage des <i>reads</i>	55
V.11.3. Élimination des séquences d'amorces	55
V.11.4. Annotation des <i>reads</i> et identification des réarrangements V(D)J	55
V.11.5. Recherche des BCRs stéréotypés type « LLC »	56
V.11.6. Confirmation des <i>subsets</i> assignés avec l'outil Arrest/Assign Subsets	56
V.11.7. Phylogénie moléculaire	56
VI. Résultats	57
VI.1. Évaluation du pourcentage de lymphocytes B	57
VI.2. Évaluation de la performance du tri cellulaire et de l'extraction d'ADN	57
VI.3. Préparation des librairies	57
VI.4. Évaluation de la qualité et des performances du séquençage	58
VI.4.1. Analyse du nombre de <i>reads</i>	59
VI.4.2. Qualité de séquençage	59
VI.4.3. Analyse de la taille des <i>reads</i> obtenus	60
VI.4.4. Analyse des séquences <i>Unique Molecular Identifiers</i> (UMI)	61
VI.4.5. Évaluation du taux de séquences productives analysées	61
VI.4.6. Pool Témoins	62
VI.4.7. Pool LLC	62
VI.4.8. Confrontation avec le séquençage Sanger	64
VI.4.9. Gamme de dilution	64
VI.4.10. Analyse de l'échantillon LLC#1	66
VI.5. Application à la recherche de séquences stéréotypées type « LLC »	68
VI.5.1. Analyse des métriques	68
VI.5.2. Profil de clonalité	69
VI.5.3. Recherche des séquences stéréotypées	69
VI.5.3.1. Existence de CDR3 de type <i>subset</i> #2 sans IGHV3-21	71
VI.5.3.2. Diversité intra-clonale au sein du <i>subset</i> #2	72
VI.5.3.3. Diversité des séquences ayant la même séquence CDR3	74
VI.5.3.4. Regard critique sur l'analyse des données	75
VII. Discussion	77
VII.1. Mise en place du séquençage haut débit des gènes des immunoglobulines	77
VII.1.1. Aspects qualitatifs de la méthode	77
VII.1.2. Aspects quantitatifs de la méthode	77

VII.2. Recherche de séquences stéréotypées dans une cohorte de 19 patients exempts d'hémopathie.....	78
Conclusion et perspectives	81
Références bibliographiques	82
Annexes	87
Serment De Galien	91

Table des illustrations

Figure 1. Structure générale d'une immunoglobuline	20
Figure 2. Structure des gènes V dans le sens 5'-3'	21
Figure 3. Organisation des différents loci des immunoglobulines	22
Figure 4. Développement du lymphocyte B	25
Figure 5. Exclusion allélique au cours de la lymphopoïèse B	26
Figure 6. Schéma de structure des chaînes variables IGH avec les séquences RSS ...	27
Figure 7. Représentation schématique de la recombinaison V-D-J des chaînes lourdes	28
Figure 8. Le système immunitaire	30
Figure 9. Le microenvironnement du centre germinatif	32
Figure 10. La diversité mutationnelle retrouvée dans la LLC	36
Figure 11. Courbe de survie des groupes IGHV « muté » et « non muté »	37
Figure 12. Les critères de définition des subsets et exemples de motifs de CDR3	38
Figure 13. Proportion des patients ayant un BCR stéréotypé parmi les patients LLC ..	39
Figure 14. Classification pronostique	40
Figure 15. Modèle de développement et d'évolution de la LLC	41
Figure 16. Stratégie de traitement de la LLC d'après les critères de l'IWCLL 2018	43
Figure 17. Évolution des traitements de la LLC	44
Figure 18. Stratégie de préparation des librairies	48
Figure 19. Séquençage Miseq Illumina® par synthèse de nucléotides	53
Figure 20. Évaluation de la qualité du signal à chaque cycle	53
Figure 21. Stratégie d'analyse des données de séquençage	54
Figure 22. Composition d'un fichier Fastq	55
Figure 23. Stratégie de recherche des séquences stéréotypées et tableau des subsets majeurs de type "LLC"	56
Figure 24. Évaluation de la performance du tri cellulaire et de l'extraction d'ADN	57
Figure 25. Migration sur électrophorèse capillaire multi Na	58
Figure 26. Représentation graphique de la qualité de chaque base séquencée	60
Figure 27. Répartition de la taille des reads	60
Figure 28. Analyse d'une séquence de taille inférieure à 250	61
Figure 29. Problème d'amplification avec les UMIs	61
Figure 30. Comparaison du répertoire des gènes VH	62

Figure 31. Répartition des fréquences de LLC retrouvées dans le pool.....	63
Figure 32. Vérification du signal clonal de l'ADN de LLC #11.....	63
Figure 33. Estimation de la probabilité de ne pas détecter un clone entre 10^{-3} et 10^{-5} ..	65
Figure 34. Gamme de dilution de chaque clone LLC du pool	66
Figure 35. Assignation des subsets par ARRest/AssignSubsets.....	67
Figure 36. Analyse des métriques du séquençage avec le kit V3.....	68
Figure 37. Profil type clonalité des échantillons séquencés.....	69
Figure 38. <i>Subsets</i> non confirmés avec ARRest/Assign subsets	70
Figure 39. <i>Subsets</i> minoritaires dans la cohorte.....	70
Figure 40. <i>Subsets</i> majoritaires de la cohorte	71
Figure 41. Répartition des familles de gènes V non V3-21 ayant un motif <i>subset #2</i>	72
Figure 42. Répartition du nombre de <i>reads</i> de séquences stéréotypées pour le <i>subset #2</i> et la fréquence des CDR3 uniques.....	73
Figure 43. Séquences nucléotidiques des 16 séquences du Patient #1 ayant un CDR3 identique	74
Figure 44. Phylogénie moléculaire des clones ayant un CDR3 identique de <i>subset #2</i>	75
Figure 45. Motif de la séquence CDR3 du <i>subset #2</i>	76
Figure 46. Hypothèses de l'évolution du répertoire B liée au vieillissement.....	80

Table des tableaux

Tableau 1. Score de Matutes	34
Tableau 2. Classification de Rai et Binet	34
Tableau 3. Impact des anomalies cytogénétiques sur le pronostic	35
Tableau 4. Score pronostic avec les anomalies géniques et cytogénétiques	36
Tableau 5. Construction des amorces PCR-1	49
Tableau 6. Mix PCR-1	50
Tableau 7. Mix PCR-2	51
Tableau 8. Analyse du nombre de reads	59
Tableau 9. Abondance des clones attendue dans la gamme de dilution	64

Liste des abréviations

Ac : anticorps	Ig : immunoglobuline
ADN : acide désoxyribonuléique	IGHV : région variable des chaînes lourdes des Ig
Ag : antigène	IL : interleukine
AID : Activation Induced Cytidine Deaminase	IWCLL : International Workshop on Chronic Lymphocytic Leukemia
AMM : autorisation de mise sur le marché	L : Light
BCR : récepteur du lymphocyte B	LB : lymphocyte B
BTK : Bruton Tyrosine Kinase	LLC : leucémie lymphoïde chronique
CDR : complementary determining region	LT : lymphocyte T
CLP : progéniteur lymphoïde commun	MRD : Minimal Residual Disease
CMF : cytométrie en flux	NHEJ : Non homologous End-Joining system
CPA : cellule présentatrice d'antigène	pb : paire de bases
CSH : Cellules Souches Hématopoïétiques	PBS : tampon phosphate salin
DMSO : diméthylsulfoxyde	PLC : Phospholipase C
E2A : E box binding protein 2A	PU.1 : Purine Box Factor 1
EBF : Early B-cell Factor	RAG : Recombination Activation Gene
ERIC : European Research Initiative on Chronic lymphocytic leukemia	RSS : Recognition Signal Sequence
FR : framework region	TDL : temps de doublement
H : heavy	TdT : terminal déoxynucléotidyl transférase
HMG : High Mobility group proteins	XRCC4 : X-Ray Cross-complementing 4
HMS : hypermutation somatique	

Introduction

Chaque individu développe au cours de sa vie une immunité face aux différents agents extérieurs, notamment les bactéries ou les virus. Ils sont considérés comme des agents du « non soi ». L'immunité, du grec « *immunis* » qui veut dire « *protéger de* », est développée chez tous pour pouvoir lutter efficacement contre ces agressions.

Il existe deux types d'immunité. Le premier type est l'immunité innée qui permet de lutter contre tous les antigènes (Ag) du « non soi », sans même avoir eu de contact avec eux au préalable. Elle regroupe plusieurs structures. Parmi elles, les différentes barrières anatomiques telles que la peau, des cellules douées de phagocytose comme les macrophages, les polynucléaires neutrophiles ; ainsi que des cellules capables d'induire une cytotoxicité (les lymphocytes NK).

Le second type est l'immunité adaptative, c'est-à-dire qu'elle est acquise au cours de la vie d'un individu. Elle met en jeu des cellules d'origine hématopoïétique telles que les lymphocytes B ou T et coopère étroitement avec l'immunité innée. Après une phase « d'apprentissage » (*i.e. : immunisation*), la réponse immune est rapide dès l'introduction d'un Ag dit « non soi » grâce à la présence de cellules dites de « mémoire ». Ces cellules peuvent persister de plusieurs semaines à plusieurs années dans l'organisme.

Les deux piliers de cette immunité acquise sont l'immunité adaptative humorale, dépendante de la production d'immunoglobulines (*i.e. : anticorps*) par les lymphocytes B et l'immunité adaptative cellulaire médiée par les lymphocytes T cytotoxiques.

Au cours de ce travail, nous allons nous intéresser aux cellules lymphoïdes, plus particulièrement aux lymphocytes B. Après avoir rappelé la structure et la génétique des immunoglobulines produites par les lymphocytes B, nous parlerons de l'ontogénie de ces derniers. Enfin, nous aborderons la question de la transformation maligne de ces cellules au travers du modèle de la leucémie lymphoïde chronique (LLC).

I. Structure et génétique des immunoglobulines

Les immunoglobulines (Ig) ou anticorps (Ac) sont produites par le lymphocyte B. Ce sont des glycoprotéines de 150kDa avec une structure bien définie.

I.1. Structure des immunoglobulines

Les immunoglobulines sont constituées de deux chaînes lourdes identiques appelées « Heavy » ou « H » et deux chaînes légères identiques nommées « Light » ou « L ».

Elles ont une structure hétéro-tétramerique. Pour maintenir l'intégrité de la structure de l'immunoglobuline, les chaînes lourdes et légères sont reliées entre elles de manière covalente par des ponts disulfures (cf. Figure 1).

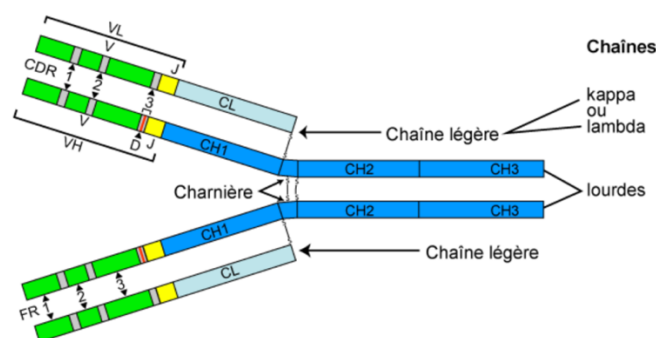


Figure 1. **Structure générale d'une immunoglobuline**

FR = Framework region, CDR = Complementary Determining Region, CH = domaine constant des chaînes lourdes, CL = domaine constant des chaînes légères, VH = domaine variable des chaînes lourdes, VL = domaine variable des chaînes légères

[d'après MP. Lefranc et al. 2002 (1)]

I.1.1. Les chaînes lourdes « Heavy »

Les chaînes lourdes sont constituées d'une région dite constante qui assure le rôle effecteur de l'anticorps et d'une région variable qui permet la reconnaissance de l'antigène. Ces deux régions sont assemblées par une région charnière.

La région constante permet l'activation du complément par sa liaison au domaine Fc. Elle est composée de 3 domaines appelés CH1, CH2 et CH3. Ces domaines sont séparés en 5 groupes ou « isotypes » ayant des caractéristiques distincts (IgM, IgG, IgA, IgD et IgE).

La région variable, comme dit plus haut, permet la fixation à l'antigène. Pour cela, cette partie a besoin d'être hétérogène d'une immunoglobuline à l'autre afin de pouvoir capter un maximum d'antigène. La diversité du répertoire des Ig est avant tout assurée par une recombinaison aléatoire de différents gènes V (pour Variable), D (pour Diversity) et J (pour Junction) (1–3). La séquence obtenue après l'assemblage des gènes V, D et J constitue la région variable. Elle est elle-même composée de 4 régions très conservées dites « régions

d'armature » ou « *Framework Region* » (FR1 à FR4) qui encadrent 3 régions hypervariables d'une dizaine d'acides aminés nommée « *Complementary Determining Region* » (CDR). Cette structure ainsi formée va pouvoir se lier à l'antigène dont elle est spécifique grâce aux régions CDRs (cf. Figure 1).

Enfin, en amont de cette région variable, il y a le domaine Leader. Il est séparé en deux exons, un exon de 40 à 50pb (L-PART1) et un second de 12pb (L-PART2). Tous deux sont séparés par un intron de 100 à 400pb (cf. Figure 2). Cette région contient une partie de la séquence du peptide signal nécessaire à l'induction de la transcription. Elle correspond à la région 5'UTR (« *Untranslated Transcribed Region* ») de l'ARNm codant pour les immunoglobulines (2).

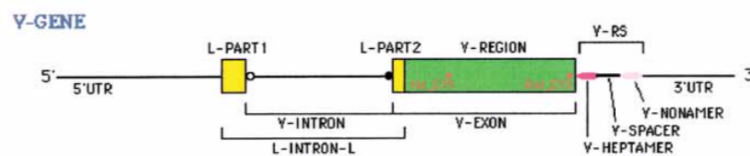


Figure 2. **Structure des gènes V dans le sens 5'-3'**

En 3'UTR, on retrouve la séquence RSS (« V-RS ») qui permet l'initiation de la recombinaison.

[d'après MP. Lefranc et al. 2002 (1)]

I.1.2. Les chaînes légères « *Light* »

Les chaînes légères se distinguent tout d'abord par l'absence de région charnière entre leur région variable et constante. Ensuite, contrairement aux chaînes lourdes, elles ne possèdent qu'un seul domaine constant. Ce domaine est composé de deux isotypes kappa et lambda. Enfin, leur partie variable n'est composée que des gènes de variabilité et de jonction. Il n'y a pas de gènes de diversité dans la partie variable des chaînes légères.

I.2. Organisation des gènes des immunoglobulines

Maintenant que nous avons détaillé la structure des Ig, nous allons nous intéresser à leur organisation. Comme évoqué précédemment, les Ig ont des chaînes lourdes et légères.

Les chaînes lourdes sont codées par un seul locus nommé IGH, situé sur le chromosome 14 (14q32.2). Il comprend 123 à 129 gènes V dont 38 à 46 sont fonctionnels, organisés en 7 sous-groupes. Pour les gènes D, il en existe 27 différents dont 23 fonctionnels. Enfin, il y a 9 gènes J dont 6 fonctionnels (2).

Pour les chaînes légères, le locus diffère en fonction de leur isotype kappa ou lambda. Les chaînes légères kappa sont codées par le locus « IGK » situé sur le chromosome 2 (2q1.2). Il comprend 76 gènes variables dont 34 à 37 fonctionnels qui appartiennent à 5 sous-groupes. Le domaine J est séparé en 5 gènes. Un gène unique code pour la partie constante. Enfin, les chaînes légères lambda sont codées par le locus « IGL » situé sur le chromosome 22 (22q11.2). Il comprend 70 à 74 gènes variables dont 29 à 35 sont fonctionnels qui appartiennent à 10 sous-groupes. La partie constante est codée par 7 gènes. Chaque partie constante est couplée à un domaine J en tandem (*cf.* Figure 3).

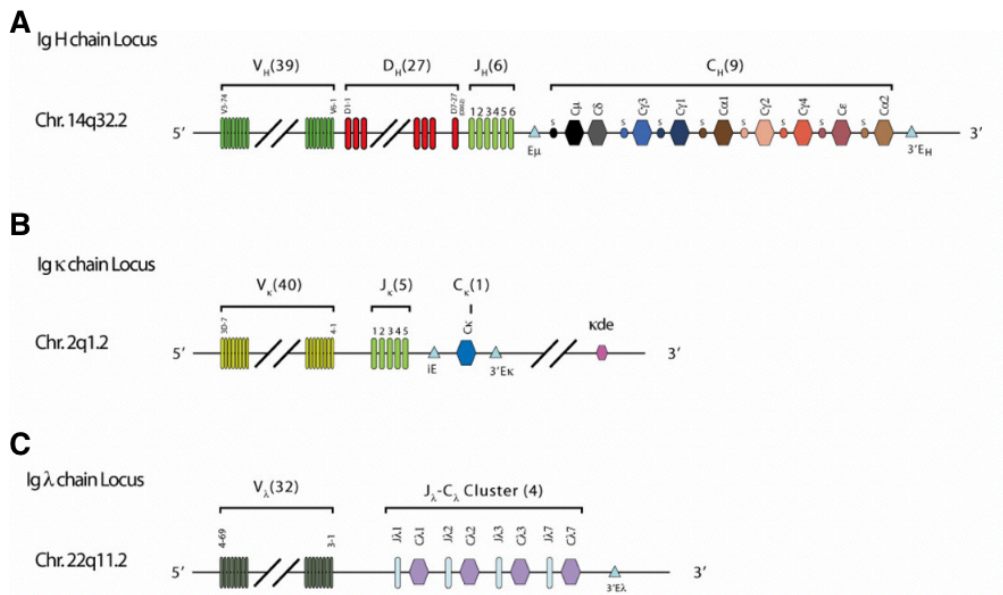


Figure 3. **Organisation des différents loci des immunoglobulines**

Les loci IGH (A), IGK (B) et IGL (C) représentés respectivement dans le sens 5'-3'.

[d'après Schroeder et Cavacini 2010 (3)]

Ceci entraîne une variabilité combinatoire d'environ 6 300 possibilités pour la chaîne lourde. De la même façon, chacune des chaînes légères kappa et lambda dispose de 200 combinaisons possibles.

Au total, la diversité combinatoire entre les gènes des chaînes lourdes et légères totalise environ 10^6 possibilités. A cela s'ajoute la variabilité due à l'insertion aléatoire de séquences nucléotidiques N et de séquences palindromiques P (microhomologie), lors des jonctions entre les différents gènes qui augmente la diversité d'un facteur de 10^3 à 10^4 . La dernière source de diversité combinatoire (d'environ 10^3) est l'acquisition de mutations somatiques sur les différentes régions des immunoglobulines, particulièrement au niveau des segments variables (dont on reparlera en détail plus loin).

Au total, on dispose d'un panel de 10^{12} combinaisons possibles d'immunoglobulines différentes. La limite de ce répertoire est le nombre de lymphocytes B dont dispose un individu (1).

I.3. Immunoglobuline membranaire et récepteur à l'antigène du lymphocyte B (BCR)

Toutes ces combinaisons évoquées ci-dessus vont permettre d'assurer une diversité des immunoglobulines synthétisées. Ces immunoglobulines jouent un rôle essentiel pour le lymphocyte B.

L'immunoglobuline va être intégrée à la membrane du lymphocyte B et sera associée à des molécules de co-stimulation telles que CD79a, CD79b (4). Cette association entre l'immunoglobuline membranaire et les molécules de co-stimulation forme le « *B cell receptor* » ou BCR.

La partie immunoglobuline du BCR conserve son rôle *princeps* de fixer l'antigène dont elle est spécifique. Les molécules de co-stimulation permettent l'initiation du signal lors de la fixation d'un antigène sur la partie immunoglobuline.

Le BCR joue un rôle essentiel lors de « l'éducation » du lymphocyte B au cours de sa maturation.

II. Ontogénie du lymphocyte B

Dans cette nouvelle partie, nous allons rappeler les différentes étapes de la maturation du lymphocyte B. Reprenons quelques généralités tout d'abord. Comme évoqué dans la première partie, les cellules lymphoïdes sont essentielles à l'immunité. Elles sont produites par les organes lymphoïdes primaires et terminent leur maturation dans les organes lymphoïdes secondaires. Elles circulent dans l'ensemble de l'organisme par voie sanguine ou lymphatique afin d'exercer leur rôle de surveillance immunitaire.

Le développement des cellules lymphoïdes est appelé la lymphopoïèse. A partir de la naissance, chez l'homme, elle débute dans la moelle osseuse (organe lymphoïde « primaire »). Cette première phase est appelée : lymphopoïèse B indépendante de l'antigène.

II.1. Lymphopoïèse B indépendante de l'antigène

C'est le début de la lymphopoïèse à partir d'une Cellule Souche Hématopoïétique (CSH) et à ce stade, il n'y a pas d'interaction avec les antigènes.

L'engagement de la CSH vers la différenciation lymphocytaire B est régulé par le microenvironnement médullaire qui regroupe des adipocytes, des cellules stromales et possède un réseau vasculaire spécifique. Selon les cytokines produites, elles orientent la différenciation des CSH vers un lignage cellulaire donné, notamment grâce à l'induction de facteurs de transcription spécifique.

Les principaux facteurs de transcription impliqués dans la lymphopoïèse sont Ikaros et PU.1 (« *Purine Box Factor 1* »). Leur association va permettre l'engagement de la cellule souche hématopoïétique vers le lignage B pour se différencier en cellule pré-pro-B (5,6).

Comme le montre la Figure 4, la lymphopoïèse se réalise en plusieurs étapes.

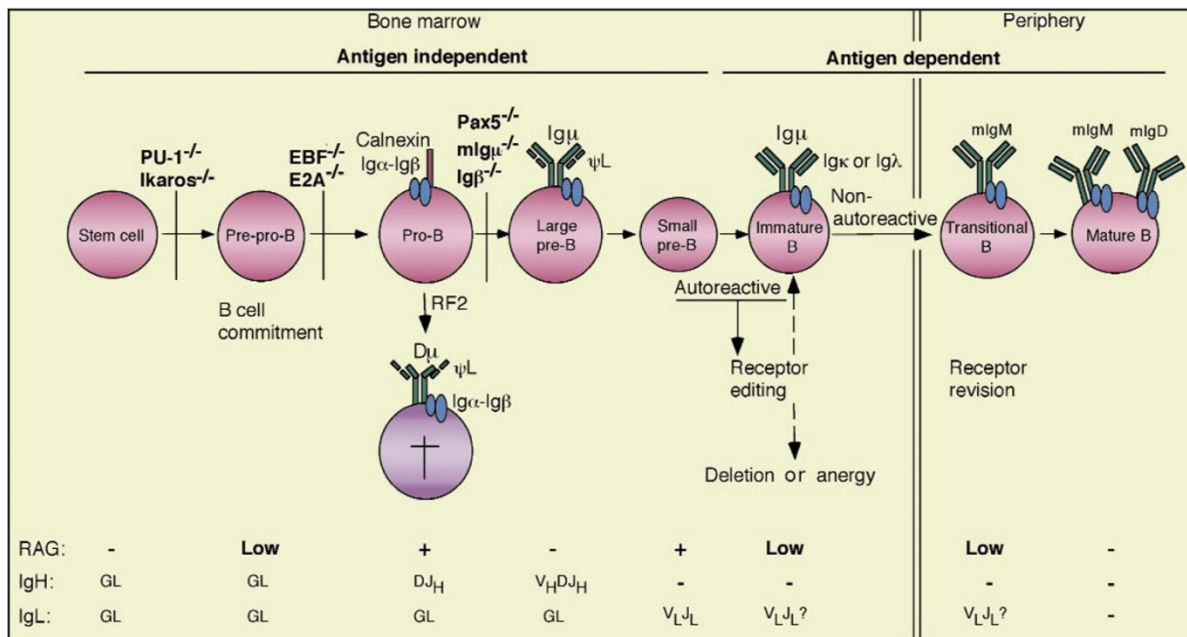


Figure 4. Développement du lymphocyte B

Représentation des différents stades du lymphocyte B lors de la lymphopoïèse. Les premières étapes ont lieu dans la moelle osseuse et sont indépendantes de la présence d'un Ag. On retrouve les stades : cellule souche, pré-pro-B, pro-B, large-pré-B, small-pré-B. Le stade B immature est le dernier stade du lymphocyte B dans la moelle osseuse. La moelle libère le lymphocyte B « transitionnel » qui poursuivra sa maturation dans les organes lymphoïdes secondaires.

[d'après Meffre et al. 2000 (7)]

II.1.1. Stade pré-pro-B

Le stade pré-pro-B est le stade le plus précoce du lymphocyte B. La cellule est engagée dans le lignage B et le réarrangement des gènes des immunoglobulines n'a pas encore eu lieu. La poursuite de la lymphopoïèse est possible grâce au facteur de transcription PAX-5 qui va induire un engagement irréversible vers le lignage B au détriment des lignages T et myéloïde (7).

II.1.2. Stade pro-B

A ce stade, les facteurs de transcription EBF (*Early B-cell Factor*) et E2A (*E box binding protein 2A*) vont induire l'expression de marqueurs spécifiques du lignage B comme le CD19 (5). Une fois la cellule engagée définitivement dans la lymphopoïèse B, les enzymes RAG1 et RAG2 s'expriment et vont initier le réarrangement des gènes des immunoglobulines (V, D et J).

Le processus commence d'abord au niveau du locus des chaînes lourdes. La première étape est la jonction D_H-J_H puis V_H-D_HJ_H. Cette recombinaison a lieu tout d'abord sur un seul allèle. Si le réarrangement est productif, l'autre allèle ne sera pas réarrangé. Ce phénomène est appelé l'exclusion allélique.

Dans le cas contraire, en l'absence de réarrangement productif sur le premier allèle, le second allèle subira la recombinaison V(D)J. Si les deux réarrangements sont non productifs, alors la cellule est éliminée par induction d'apoptose (cf. Figure 5).

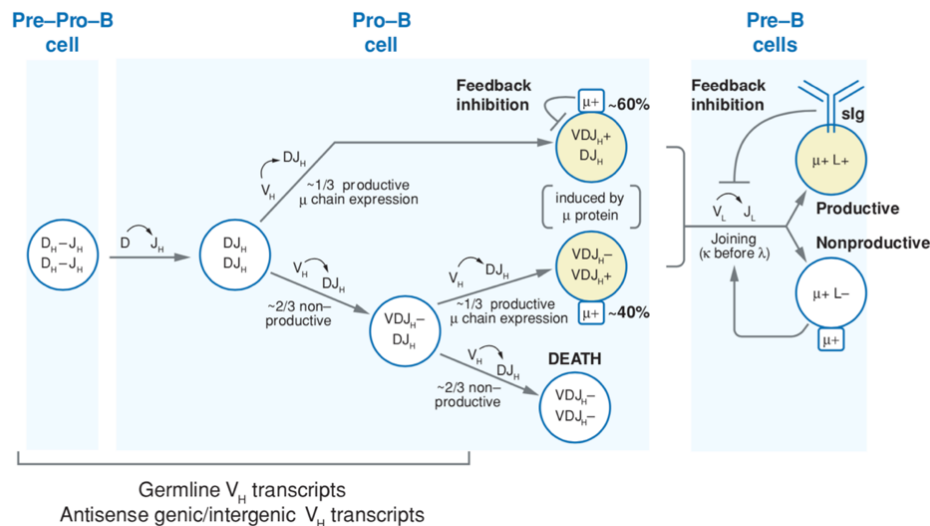


Figure 5. **Exclusion allélique au cours de la lymphopoïèse B**

Les réarrangements sont réalisés successivement sur les deux allèles de chaque locus, jusqu'à obtenir un réarrangement productif.

[d'après Jung et al., 2006 (8)]

II.1.3. Mécanisme de la recombinaison V(D)J

La recombinaison des régions variables des immunoglobulines est un remaniement de l'ADN germlinal. C'est un phénomène complexe.

La première étape est une cassure double brin de l'ADN. Cette cassure est contrôlée par les enzymes RAG1 et RAG2 ainsi que par les protéines HMG1 et HMG2 (« *High Mobility group proteins* ») (9,10).

Ce phénomène nécessite la présence d'une séquence signal de recombinaison : la séquence « *Recognition Signal Sequence* » ou « *RSS* » (11). Elle est composée d'un heptamère et d'un nonamère conservé séparés par 12 ou 23 nucléotides peu conservés définissant ainsi des RSS 12-mer et RSS 23-mer.

Une séquence RSS 23-mer est présente en 3' des gènes V et une séquence RSS 12-mer est présente en 5' des gènes J. En revanche les gènes D sont flanqués d'une séquence RSS 12-mer et RSS 23-mer des côtés 5' et 3' respectivement (cf. Figure 6).

Les différentes séquences RSS s'assemblent seulement si leur taille est différente, c'est-à-dire qu'une séquence RSS 12-mer ne peut s'associer qu'avec une séquence RSS 23-mer. En effet, deux séquences RSS 12-mer ou deux séquences RSS 23-mer ne peuvent s'associer entre elles. Cette règle permet d'éviter les réarrangements non désirés.

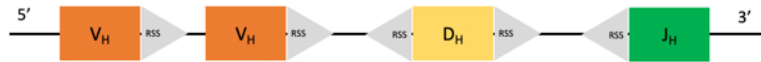


Figure 6. **Schéma de structure des chaînes variables IGH avec les séquences RSS**

Les régions RSS sont retrouvés en 3' des gènes V et 5' des gènes J.
 Pour les gènes D, une séquence RSS est flanquée en 3' et 5'.

Lors de la formation des chaînes lourdes, le choix des différents segments V_H - D_H - J_H est aléatoire. La première étape de la formation des régions variables des chaînes lourdes se fait grâce à l'assemblage des segments J_H et D_H . Ensuite, ce réarrangement D_H - J_H s'associe à un segment V_H . Le tout forme un exon V_H - D_H - J_H codant pour la partie variable qui est ensuite associé aux exons codant la région constante. Après épissage, les ARN messagers matures sont traduits en protéine (cf. Figure 7).

De la même façon, pour les chaînes légères kappa et lambda, il y a un assemblage des chaînes V_L - J_L de manière aléatoire pour constituer une région variable fonctionnelle.

Les fragments d'ADN non utilisés pour les réarrangements sont délétés sous forme d'ADN circulaire (4).

Lors de l'assemblage des différents segments, des phénomènes de réparation de l'ADN créent une variabilité dans les zones de jonction entre les gènes associés. Ce phénomène introduit de la diversité jonctionnelle.

Comme précisé ci-dessus, la première étape lors de la recombinaison est une cassure double brin de l'ADN. Ceci entraîne l'intervention d'un système ubiquitaire qui permet la réparation, le « *Non homologous End-Joining system* » ou « NHEJ ». C'est un système enzymatique qui permet la réparation de jonctions d'extrémités non homologues (12). Lors de la cassure double brin, il y a recrutement des protéines Ku qui sont très abondantes dans les cellules et se lient aux extrémités de l'ADN (13). Pour terminer la réparation, c'est ADN ligase IV qui permet la liaison entre les extrémités des brins d'ADN.

Lors de cette réparation, spécifiquement au cours de la lymphopoïèse, la TdT (« *Terminal Déoxynucléotidyl Transférase* ») intervient et va insérer plusieurs nucléotides au niveau des jonctions entre les gènes V, D et J recombinaison (13). La TdT est exprimée très tôt au cours de la lymphopoïèse B, dès le stade pro-B. Ces insertions de N nucléotidiques apportées par la TdT apportent une diversité jonctionnelle.

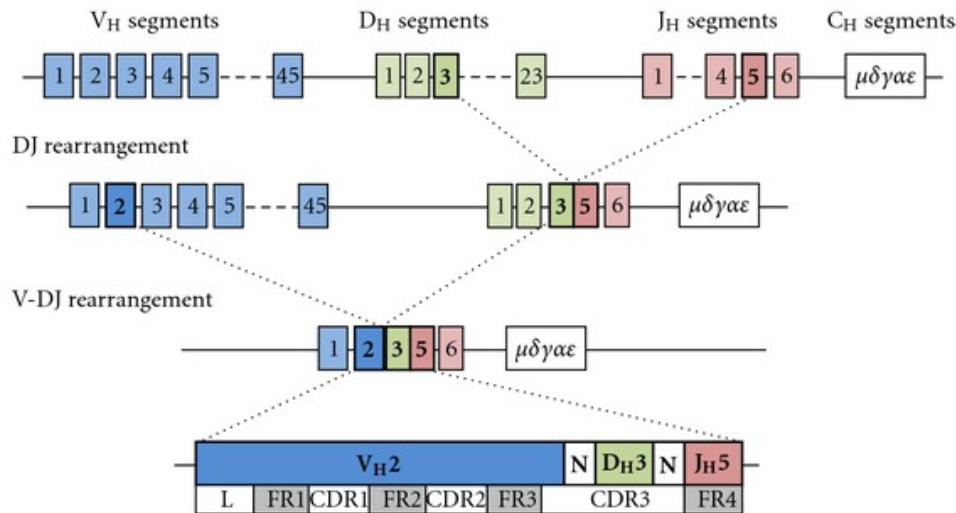


Figure 7. **Représentation schématique de la recombinaison V-D-J des chaînes lourdes**

La première étape permet l'association d'un gène D_H et J_H, ici D3 et J5.

La seconde étape est l'association V_H avec le segment D_H-J_H, ici V2 avec D3-J5.

Enfin, une fois les segments V_H-D_H-J_H assemblés, la séquence complète est réarrangée. Entre chaque segment, V_H-D_H-J_H il y a des régions N qui sont ajoutées à chaque réarrangement.

[d'après Raess et Bagg, 2012 (14)]

II.1.4. Stade pré-B

A ce stade, le pré-BCR (chaîne lourde μ associée à la pseudo-chaîne légère ψL) est présent à la surface des cellules. La transduction de son signal réprime l'expression des gènes RAG1 et RAG2 afin d'inhiber le réarrangement du second allèle IGH, dans le cas de la présence d'un réarrangement fonctionnel sur le premier allèle.

Le signal du pré-BCR permet la production d'une protéine, SLP65 (15) induisant un rétrocontrôle négatif de l'inhibition des gènes RAG (16). Le signal du pré-BCR est donc inhibé et les gènes RAG ré exprimés. Le réarrangement des chaînes légères peut commencer, il débute par le locus IGK. En cas de réarrangement Kappa non fonctionnel, le locus IGL est réarrangé. Lorsqu'une chaîne légère fonctionnelle est enfin produite, l'expression des gènes RAG1 et RAG2 est inhibée (cf. Figure 5).

II.1.5. Stade B immature

Après la production de réarrangements fonctionnels des chaînes lourdes et légères, l'immunoglobuline formée s'associe au CD79a et CD79b pour former le BCR fonctionnel. A ce stade, le lymphocyte B exprime une IgM de surface. Les lymphocytes B restent dans ce compartiment immature, au sein de la moelle osseuse, pendant une moyenne de 3 à 4 jours (7). A partir de ce stade, les lymphocytes B vont être en contact avec les Ag du « soi ». Ceux qui reconnaissent ces auto-Ag sont appelés « LB autoréactifs ». Ils vont soit éditer leur BCR, c'est-à-dire modifier le réarrangement V(D)J exprimé, soit être éliminés par déclenchement de l'apoptose ou inactivés par induction d'un état d'anergie (7). Les cellules B qui survivent et qui ne sont pas autoréactives quittent la moelle osseuse pour poursuivre leur maturation dans les organes lymphoïdes secondaires.

II.1.6. Stade B transitionnel

A ce stade, le lymphocyte B a quitté la moelle osseuse et poursuit sa maturation dans les organes lymphoïdes secondaires.

Le lymphocyte B transitionnel a une courte durée de vie. Seulement 10 à 30% des B transitionnels vont évoluer en B matures de longue durée de vie. Le nombre de B transitionnel peut être variable au cours du temps contrairement au nombre de B matures. En effet, au cours d'infection, le nombre de B transitionnel diminue de manière transitoire. La perte sera compensée par un rebond d'hyperproduction (7).

II.2. Lymphopoïèse B dépendante de l'antigène

Le lymphocyte B transitionnel va coloniser les organes lymphoïdes secondaires représentés par les ganglions lymphatiques, les amygdales, les follicules lymphatiques des muqueuses et la pulpe blanche de la rate (*cf.* Figure 8). A ce niveau, le lymphocyte B pourra terminer sa maturation de manière T dépendante ou T indépendante, selon le type d'antigène.

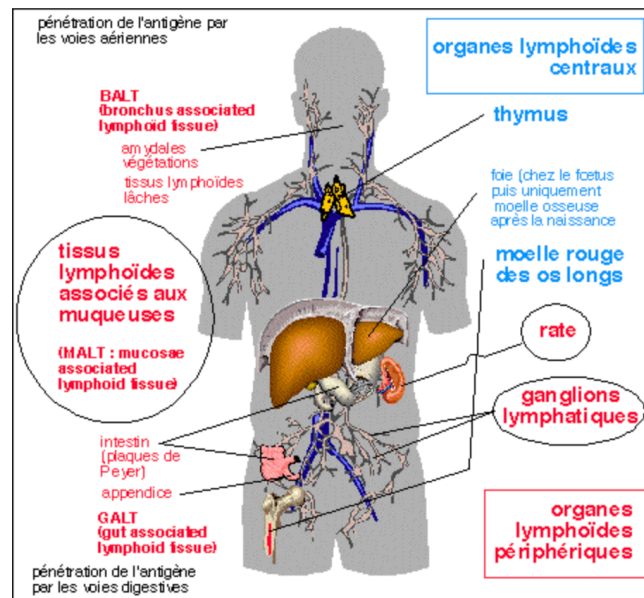


Figure 8. Le système immunitaire

*Il comprend les organes centraux (thymus et moelle rouge des os longs) et les organes périphériques (ganglions lymphatiques, rate et organes lymphoïdes associés aux muqueuses) mais aussi les appareils circulatoires (sanguin et lymphatique)
[d'après Immunologie : organes, cellules et molécules (17)]*

II.2.1. Réponse à l'antigène T dépendant

Lors de la réponse à un Ag T dépendant, les lymphocytes B impliqués dans la réponse immune vont migrer jusqu'aux zones ganglionnaires riches en lymphocytes T CD4+. Au contraire les cellules B ne réagissant pas avec les cellules T sont repoussées dans la zone du manteau. On retrouve aussi à ce niveau des cellules présentatrices d'antigène (CPA) telles que les cellules dendritiques qui vont pouvoir présenter les antigènes aux lymphocytes B et T. Les antigènes présentés par les CPA vont interagir avec les molécules de co-stimulation de la cellule B et ainsi activer le lymphocyte B via son BCR. Ce microenvironnement formé, va favoriser la formation de structure appelée follicule primaire qui évolue en follicule secondaire (18).

Après plusieurs jours de stimulation des cellules B, le centre germinatif apparaît comme avoir deux zones bien distinctes : la zone sombre et la zone claire (*cf.* Figure 9). Dans la zone sombre, on retrouve les centroblastes qui sont les cellules B ayant une activité proliférique intense. La zone claire contient les centrocytes, de nombreuses cellules dendritiques, cellules T et macrophages. Les centrocytes sont de plus petites cellules B qui ne sont pas en cours de division.

C'est au sein de la zone sombre, où la prolifération est importante que les cellules B vont pouvoir subir le phénomène d'hypermutation somatique. Ce phénomène va apporter de la diversité au niveau des régions variables des Ig (IgV) dans le but d'augmenter leur affinité pour les antigènes, notamment ceux présentés par les cellules dendritiques du centre germinatif. Ensuite, les cellules B avec leur IgV modifié vont pouvoir migrer vers la zone claire pour subir la commutation de classe et terminer ainsi leur maturation en cellules mémoires ou plasmocytes (18).

L'hypermutation somatique (HMS) et la commutation de classe sont faites à l'aide d'une enzyme : AID ou « *Activation Induced Cytidine Deaminase* ». AID catalyse la désamination ciblée des résidus de désoxycytidine. Cela transforme un couple C/G en un couple U/G. La réparation du mésappariement va entraîner l'apparition de mutations.

L'hypermutation somatique (HMS) est un processus qui modifie la région variable des immunoglobulines (IgV). L'HMS introduit principalement des échanges mononucléotidiques, mais également de petites délétions et duplications (18).

La commutation de classe est le mécanisme qui permet à la chaîne lourde IgM ou IgD produite au cours de la maturation du lymphocyte B d'être modifiée en IgG, IgA ou IgE.

Cette commutation modifie les fonctions effectrices de l'immunoglobuline. Ceci est nécessaire pour éliminer certains types d'antigènes, tout en conservant la spécificité de l'immunoglobuline définie par la région variable.

C'est un processus qui implique à nouveau des cassures double brin et la réparation par le complexe NHEJ. Ces cassures surviennent dans des régions répétitives spécifiques (régions *switch*) qui précèdent les gènes de la région constante des immunoglobulines (18).

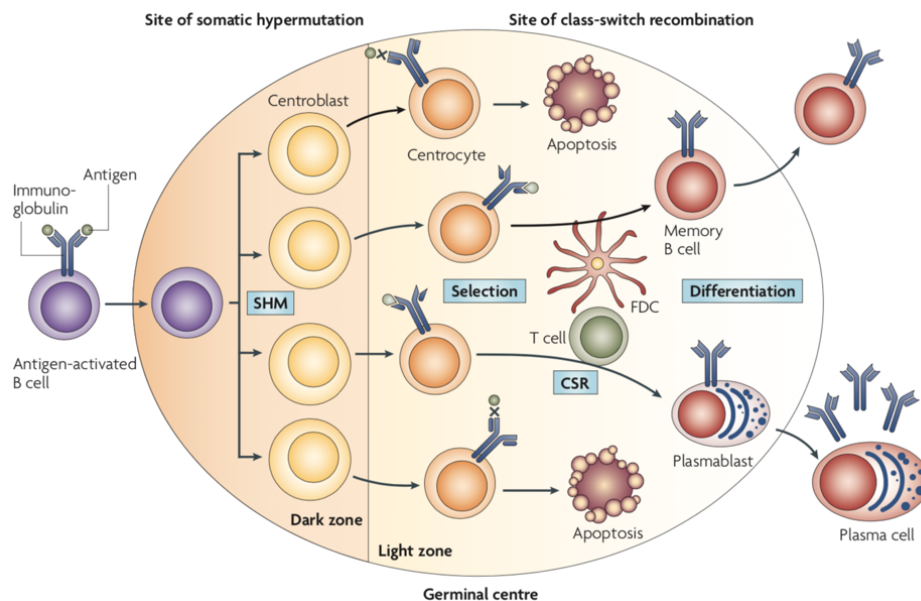


Figure 9. Le microenvironnement du centre germinatif

Environnement du lymphocyte B après l'activation antigénique. Il va subir l'HMS au niveau de la zone sombre, lieu de prolifération intense et migrer dans la zone claire où il pourra être sélectionné selon son affinité envers les Ag, subir la commutation de classe et se différencier en lymphocyte B mémoire ou plasmocyte. [d'après Klein et Dalla-Favera 2008 (18)]

II.2.2. Réponse à l'antigène T indépendant

A ce niveau, les cellules B seront plutôt stimulées par des Ag de type glucidique produits par les micro-organismes tels que le lipopolysaccharide ou les Ag contenus dans la capsule des germes encapsulés. Après interactions avec les Ag, ces cellules B évoluent en plasmocytes producteurs d'IgM. Ces cellules seront les « sentinelles » lors de l'induction de la réponse innée en cas d'infection (19).

Nous allons maintenant nous concentrer sur la transformation leucémique du lymphocyte B au travers du modèle de la leucémie lymphoïde chronique (LLC).

III. La Leucémie Lymphoïde Chronique (LLC)

III.1. Définition

La leucémie lymphoïde chronique (LLC) est un lymphome B indolent caractérisé par l'accumulation monoclonale de lymphocytes B matures CD5+ avec phase leucémique systématique.

D'après la classification OMS 2016, elle est classée dans la famille des « Tumeurs à cellules lymphoïdes B matures, groupe LLC/Lymphome lymphocytaire » (20). C'est une pathologie dont le diagnostic est réalisé sur des critères biologiques.

III.2. Épidémiologie

L'institut de veille sanitaire a réalisé un état des lieux de 1982 à 2012 sur l'incidence des LLC/Lymphome lymphocytaire. Elle estime à 4 464 le nombre de nouveaux cas de cette entité en France en 2012 avec un ratio homme/femme égal à 2.

L'âge médian au diagnostic est de 72 ans avec une médiane de LLC clinique à 64 ans. C'est une pathologie d'évolution chronique qui a une survie pouvant aller de 1 an à plus de 20 ans. La LLC est exceptionnelle avant 40 ans. La survie globale à 5 ans est supérieure à 80% tous stades confondus (21).

III.3. Diagnostic biologique

Le diagnostic de LLC est posé sur des critères biologiques bien définis. Le premier critère essentiel est la présence d'une hyperlymphocytose isolée > 5 G/L évoluant depuis plus de 3 mois chez un adulte de plus de 40 ans. Cette anomalie nécessite des investigations biologiques complémentaires.

A l'examen cytologique, on retrouve une population lymphocytaire monomorphe de petite taille avec une chromatine mottée. Associée à ces lymphocytes matures, on retrouve des « ombres de Gümprécht », ce sont des cellules apparaissant comme lysées sur le frottis sanguin (22). Ceci pourrait être lié au fait que dans les lymphocytes clonaux, l'expression de la vimentine est diminuée. La vimentine est une protéine du cytosquelette assurant la rigidité et l'intégrité du lymphocyte, ce qui explique donc la présence de cellules *a priori* lysées sur le frottis sanguin. Ce n'est que l'expression d'un artéfact mécanique causé lors de l'étalement des cellules pour la réalisation du frottis sanguin (23).

Devant cette orientation cytologique, on réalise un immunophénotypage par cytométrie en flux des cellules lymphoïdes afin de rechercher la présence d'un clone. Dans le cas de la LLC, les critères de cette population sont parfaitement définis grâce au score de Matutes (*cf.* Tableau 1).

Ce score a été établi par *Matutes et al.* en 1994 (24). Ils ont proposé un score de classification avec le CD5, le CD23, FMC7, l'expression de surface des immunoglobulines et le CD22. Trois ans plus tard, en 1997, *Moreau et al.* (25) complètent ce score avec le CD79b qui permet d'augmenter son pouvoir discriminant envers les LLC par rapport aux autres syndromes lymphoprolifératifs.

Pour affirmer le diagnostic de LLC, ce score doit être supérieur ou égal à 4 (20). Lorsque ce score est égal à 3, on peut évoquer une LLC atypique, mais si ce score est inférieur à 3 le diagnostic de LLC est peu probable.

	1 POINT	0 POINT
CD5	+	-
CD23	+	-
EXPRESSION DE SURFACE DES IMMUNOGLOBULINES	Normale ou faible	Forte
FMC7	-	+
EXPRESSION CD79B/CD22	Faible	Forte

Tableau 1. **Score de Matutes**
[d'après Matutes et al. et Moreau et al. (24,25)]

III.4. Pronostic

III.4.1. Classification clinico-biologique

Pour établir le pronostic clinique des patients atteints de LLC, il existe deux classifications. Dans les pays anglo-saxons, on utilise préférentiellement la classification de Rai établie en 1975 (26) et en France, la classification de Binet établie en 1977 (27). Une étude a montré que ces deux scores étaient comparables (28). Ces deux classifications sont basées sur la présence d'un syndrome tumoral et/ou de cytopénies associées. La classification de Rai est organisée en 4 stades de 0 à IV et la classification de Binet est en 3 stades de A à C, elles sont résumées dans le Tableau 2.

Classification de Rai		Classification de Binet	
Stade 0	Lymphocytose sanguine > 5G/L ou une lymphocytose médullaire > 30%	A	Pas d'anémie, ni thrombopénie et moins de 3 aires ganglionnaires atteintes
Stade I	Critères du stade 0 + la présence d'adénopathies	B	Au moins 3 aires ganglionnaires atteintes sans anémie, ni thrombopénie
Stade II	Critères du stade 0 + splénomégalie et/ou hépatomégalie		
Stade III	Critères du stade 0 + anémie	C	Association d'une anémie et/ou thrombopénie avec des adénopathies
Stade IV	Critères du stade 0 + thrombopénie		

Tableau 2. **Classification de Rai et Binet**

III.4.2. Anomalies cytogénétiques de la LLC

Il y a des anomalies cytogénétiques récurrentes dans la LLC. En 2000, *Döhner et al.* ont proposé une classification pronostique basée sur les quatre anomalies les plus fréquentes (cf. Tableau 3).

La délétion 13q est l'anomalie ayant la prévalence la plus élevée et retrouvée chez 30 à 50% des patients au diagnostic (29–31). Elle implique la perte de deux micro-ARNs régulant l'apoptose et est associée à un pronostic favorable lorsqu'elle est isolée.

La trisomie 12 est retrouvée chez 10 à 18% des patients au diagnostic souvent associée à une cytologie atypique (23). La délétion 11q est présente chez 6 à 20% des patients, elle entraîne une délétion du gène *ATM* (« *Ataxia–Telangiectasia Mutated* ») qui code pour une protéine en amont de la protéine p53. Cette délétion peut aussi impliquer la perte du gène *BIRC3* qui a un rôle dans la résistance aux chimiothérapies classiques (32).

Enfin, on retrouve la délétion 17p impliquant une perte du gène *TP53* (gène suppresseur de tumeur) chez moins de 10% des patients au diagnostic (26,27,30). Chez les patients réfractaires ou en rechute, sa fréquence atteint 30 à 50% des patients (33,34).

PRONOSTIC	ANOMALIES CYTOGENETIQUES
FAVORABLE	del(13q14) isolée
INTERMEDIAIRE	trisomie 12 del(11q)
DEFAVORABLE	del(17p)

Tableau 3. **Impact des anomalies cytogénétiques sur le pronostic**
[d'après *Döhner et al.* 2000 (29)]

III.4.3. Biologie moléculaire

Le développement du séquençage haut débit a permis de mettre en évidence une très grande diversité de mutations dans la LLC (cf. Figure 10) (35). Cette complexité génomique reste aujourd'hui assez difficile à analyser dans son ensemble. On distingue des gènes plus fréquemment altérés dont *TP53* (7 - 15% des cas), *SF3B1*(15%), *ATM* (9%), *MYD88* (4 - 10%), *NOTCH1* (4 – 12%) et *BIRC3* (~ 5%) (31,36). Parmi les autres gènes altérés, on retrouve *ZMYM3*, *MAPK1*, *FBXW7*, *DDX3X* ou *EGR2* dont certains semblent identifier des sous-groupes de patients particuliers (36,37).

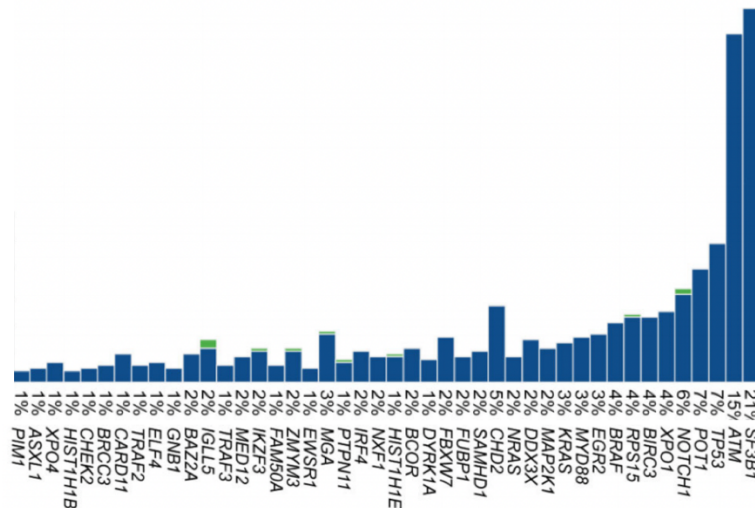


Figure 10. **La diversité mutationnelle retrouvée dans la LLC**

La fréquence de mutations de différents gènes retrouvés mutés dans une cohorte de 538 patients LLC au diagnostic (en vert les mutations synonymes et en bleu les missenses)
[adapté de Landau et al. 2015 (33)]

En 2014, Rossi et al. ont proposé une stratification pronostique des patients en prenant en compte à la fois les mutations et les anomalies cytogénétiques (cf. Tableau 4). Au total, cette classification en trois groupes plaçait les patients avec une anomalie de TP53 et/ou BIRC3 (mutation ou délétion) dans le groupe avec un pronostic défavorable. A l'inverse, les patients avec une del(13q) isolée ou/et une trisomie 12 sans autre anomalie constituaient le groupe de pronostic le plus favorable.

RISQUE PRONOSTIC	ANOMALIES
PATIENT « HAUT RISQUE »	Anomalie de TP53 (mutation ou délétion) et/ou anomalie de BIRC3 (mutation ou délétion)
PATIENT « RISQUE INTERMEDIAIRE »	Mutation NOTCH1 et/ou mutation SF3B1 et/ou del 11q Absence d'anomalie de TP53 et BIRC3
PATIENT « FAIBLE RISQUE »	Trisomie 12 et absence de mutations Del13q14 isolée

Tableau 4. **Score pronostic avec les anomalies géniques et cytogénétiques**

[d'après Rossi et al. 2014 (31)]

Au-delà du diagnostic initial, la probabilité d'apparition d'une nouvelle mutation est de 25% tous les 10 ans au cours de la maladie (31). De plus, les patients ayant une trisomie 12 ou une cytogénétique normale ont un risque de développer une anomalie de « haut risque » multiplié par deux (*i.e.* : acquisition d'anomalie de *TP53* ou *BIRC3*) (31). Au contraire, les patients ayant une *del(13q14)* isolée ont une faible probabilité de développer une anomalie de « haut risque ».

Comme les mutations de *TP53*, la prévalence des mutations de *NOTCH1*, *SF3B1* et *BIRC3* est plus élevée chez les patients réfractaires.

III.4.4. Statut mutationnel des immunoglobulines

L'impact du statut mutationnel des régions variables des chaînes lourdes des immunoglobulines (IGHV) a été mis en évidence en 1999, de manière indépendante par deux équipes *Hamblin et al.* et *Damle et al.* (38,39).

Les deux équipes ont regardé le taux de mutation des séquences IGHV des patients atteints de LLC. Les patients étaient classés en groupe « IGHV non muté » lorsque l'homologie avec la séquence germinale était supérieure ou égal à 98% et un groupe « IGHV muté » lorsque l'homologie était inférieure à 98%. Dans les deux études, le groupe de LLC « IGHV non muté » représentait environ 45 – 50% des patients. Ils présentaient des stades cliniques plus avancés ainsi qu'une espérance de vie sans traitement plus courte (*cf.* Figure 11).

Aujourd'hui encore, l'identification du statut mutationnel IGHV au seuil de 2% reste le marqueur le plus informatif, y compris face aux nouvelles données du séquençage haut débit. La détermination de ce statut a fait l'objet d'une uniformisation des pratiques à l'échelle européenne (40). Depuis peu, le statut mutationnel IGHV dans la LLC est également un paramètre décisionnel dans la stratégie thérapeutique (*cf.* Traitements de la LLC page 42).

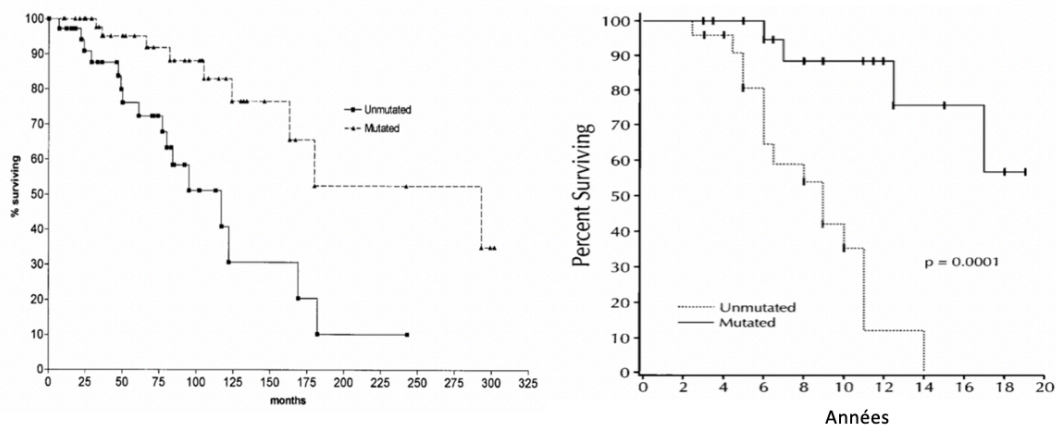


Figure 11. **Courbe de survie des groupes IGHV « muté » et « non muté »**

A gauche, l'étude d'*Hamblin et al.* avec une cohorte de 84 LLC dont 38 patients « IGHV non muté ». A droite, celle de *Damle et al.* avec une cohorte de 47 LLC dont 24 patients « IGHV non muté » [d'après *Hamblin et al.* et *Damle et al.* (38,39)]

III.4.5. Répertoire IGHV des LLC et BCR stéréotypé

Depuis quelques années, il a été mis en évidence que dès les stades précoces du lymphocyte B, un biais de sélection des gènes V a lieu (41,42). En effet, trois familles sont surreprésentées dans le répertoire normal des immunoglobulines, les familles VH1, VH3 et VH4. Pendant toute sa lymphopoïèse, le lymphocyte B subit différentes sélections qui restreignent le nombre de possibilités d'immunoglobulines. Il est apparu que certaines recombinaisons sont plus fréquentes que d'autres. En effet, on retrouve préférentiellement des combinaisons qui acquièrent des mutations somatiques tel que les VH4-34, et certaines qui conserveront leur séquence germinale comme par exemple les VH1-69.

Dans le cas de la LLC il a été montré tout d'abord que le répertoire des immunoglobulines était assez restreint (23,32,43). En 1998, certaines équipes ont pu remarquer l'existence de BCR quasi similaires dans une proportion significative de patients atteints de LLC (44). En 2004, *Messmer et al.* vont définir ces BCRs « similaires » comme des BCRs « stéréotypés » (45). En étudiant plusieurs groupes de patients LLC, ils ont pu établir des critères permettant d'identifier ces séquences stéréotypées entre les différents patients. Les caractéristiques principales de ces séquences ont rapidement évolué en une quinzaine d'années et sont résumées dans la Figure 12A (45–47). La nature du gène VH et son statut mutationnel ainsi que la séquence en acide aminé du CDR3 sont les principaux paramètres permettant de définir un BCR stéréotypé (cf. Figure 12B).

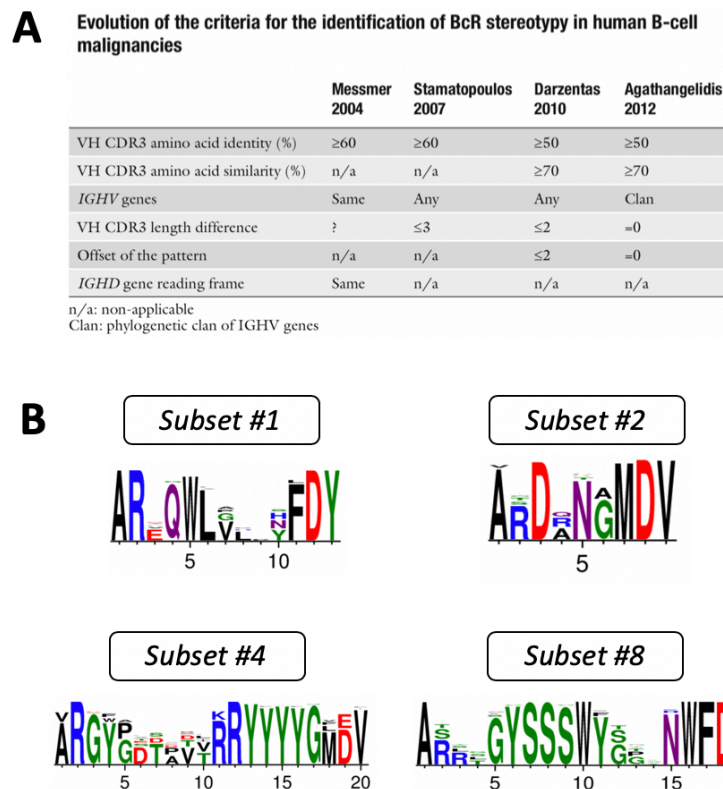


Figure 12. Les critères de définition des *subsets* et exemples de motifs de CDR3

[D'après Darzentas et Stamatopoulos, 2013 (48,49)]

Les données les plus récentes retrouvent la présence d'un BCR stéréotypé (*subset*) chez 30% des patients. On distingue actuellement 19 « *subsets majeurs* » que l'on retrouve dans 12% des cas. Les « *subsets mineurs* » sont estimés à une centaine environ et sont présents chez 18% des patients (47,49) (cf. Figure 13).

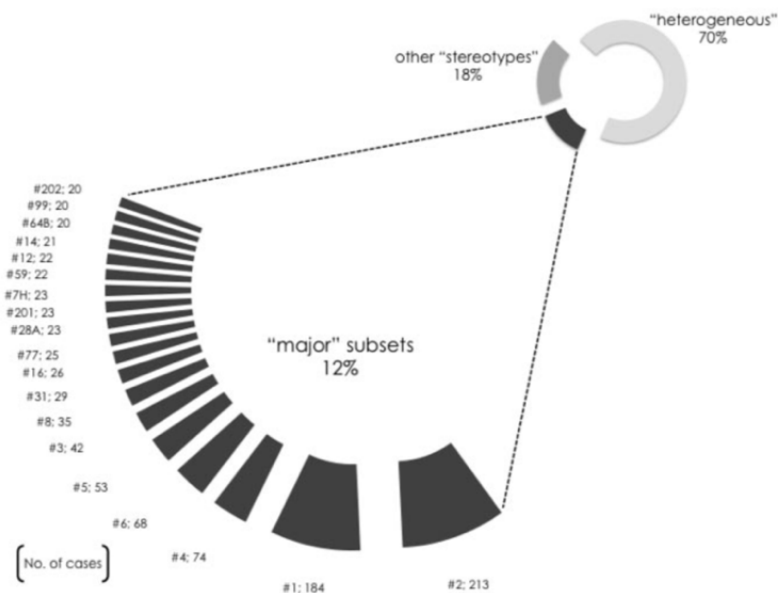


Figure 13. Proportion des patients ayant un BCR stéréotypé parmi les patients LLC

Représentation de la fraction de patients LLC avec un stéréotype majeur (« *major subsets* »), ceux avec un stéréotype mineur (« *other stereotypes* ») et ceux n'ayant pas de séquence stéréotypée (« *heterogeneous* »)

[d'après Agathangelidis et al. 2012 (47)]

Il a pu être établie une relation entre pronostic et certains *subset* ainsi que des profils mutationnels et anomalies cytogénétiques associés. Pour le *subset* #1, on retrouve des mutations de *TP53*, *NOTCH1* ainsi que la délétion 13q associée. Pour le *subset* #2 qui est le plus fréquent (22% des *subsets* majeurs), on retrouve une del(11q) et des mutations du gène *SF3B1*. Le *subset* #4 a peu d'anomalies génétiques associées et un profil IGHV muté. Enfin le *subset* #8 est associé à une trisomie 12 et des mutations de *NOTCH1*.

L'association des différentes mutations et/ou anomalies cytogénétiques permettent d'évaluer le pronostic de certains *subsets*. Les *subsets* #1, #2 et #8 sont plutôt de pronostic défavorable, au contraire du *subset* #4 qui est de pronostic favorable (49) (cf. Figure 14).

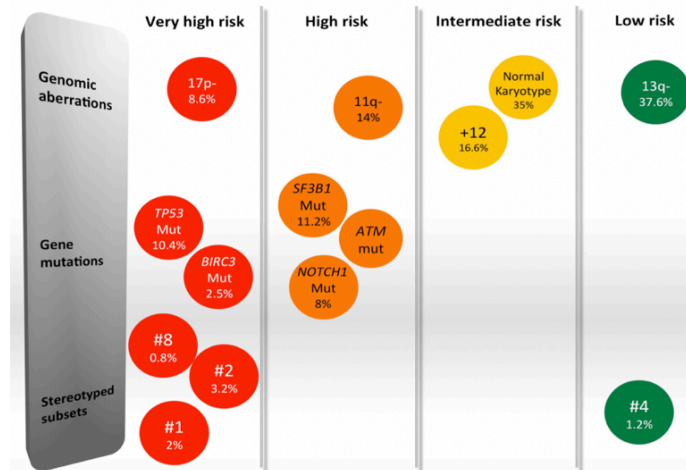


Figure 14. **Classification pronostique**

Les anomalies cytogénétiques, moléculaires et la catégorisation du BCR dans un subset permet d'établir un stade pronostic [d'après Stamatopoulos et al., 2017(49)]

III.5. Histoire naturelle de la LLC

Avec toutes les caractéristiques cytogénétiques et moléculaires de la LLC se pose la question de la cellule d'origine de cette pathologie. Pendant de nombreuses années, la LLC était présentée comme étant une pathologie liée à une cellule dépourvue de fonction d'apoptose, avec une surexpression des protéines anti-apoptotiques tels que *Bcl2*, *Mcl*, *survivin* (32,50). L'évolution de la maladie est reflétée par la balance entre la vitesse de prolifération du clone et sa sensibilité à la mort cellulaire (32). L'expansion clonale a lieu dans les ganglions lymphatiques et la moelle osseuse où un microenvironnement particulier (pseudofollicule) est favorable à la stimulation du clone *via* le BCR et les autres récepteurs de co-stimulation (51,52). Il a pu être montré que ce – dernier exerçait un rôle essentiel dans la genèse et le maintien de cet expansion clonale (53). En effet, lorsqu'il est activé il met en jeu les kinases (SYK et BTK) et phosphorylent la phospholipase C (PLC γ 2) qui permettent l'activation des facteurs de transcription NF κ B et NFAT qui vont promouvoir la prolifération cellulaire nécessaire à la survie des cellules leucémiques (23).

Dans le cas de la LLC, l'origine de cette stimulation antigénique du BCR est encore mal connue. Plusieurs hypothèses sont faites, l'une d'elle est la stimulation antigénique par des virus latents ou des bactéries commensales (32). En effet, la LLC serait le résultat d'une perpétuelle stimulation du BCR par l'agent infectieux. Une autre hypothèse avancée serait la stimulation par un auto-Ag (comme la vimentine, la filamine) qui favoriserait l'expansion clonale (43,54,55). La présence d'un répertoire biaisé et de séquences stéréotypées dans les cellules clonales de LLC soutiennent l'hypothèse d'un nombre restreint d'antigènes capable de stimuler le clone. La nature de ces auto-Ag semble être commune aux différents patients LLC.

Nous avons vu précédemment que les clones LLC pouvaient être classés en deux groupes distincts, les LLC ayant la région variable des chaînes lourdes mutées et les LLC non mutées. Or, au cours de la maturation physiologique des lymphocytes B, les mutations somatiques des régions variables des immunoglobulines sont principalement acquises au sein du centre germinatif.

L'existence de ces deux groupes pourrait être en faveur d'une double origine de la cellule leucémique. D'un côté la cellule de « LLC mutée » où sa maturation se termine avec l'aide des lymphocytes T au sein du centre germinatif et à l'inverse, la cellule de « LLC non mutée » où la fin de sa maturation se fait de manière T indépendante au niveau de la zone marginale donc hors des centres germinatifs (*cf.* Figure 15). Certains auteurs ont mis en évidence que les cellules B de la zone marginale et de la LLC étaient fonctionnellement très proches et avancent l'hypothèse que la cellule de LLC aurait une origine unique et indépendante des centres germinatifs (56,57). Cette double origine de la cellule de LLC est aussi remise en cause par le faible taux de mutation constant dans les clones mutés contrairement aux autres lymphomes non hodgkiniens où l'expérience du centre germinatif est confirmée. Il est possible que des mutations du gène V surviennent sans stimulation par les lymphocytes T en dehors des centres germinatifs. Ces mutations peuvent favoriser une auto-réactivité (32,58).

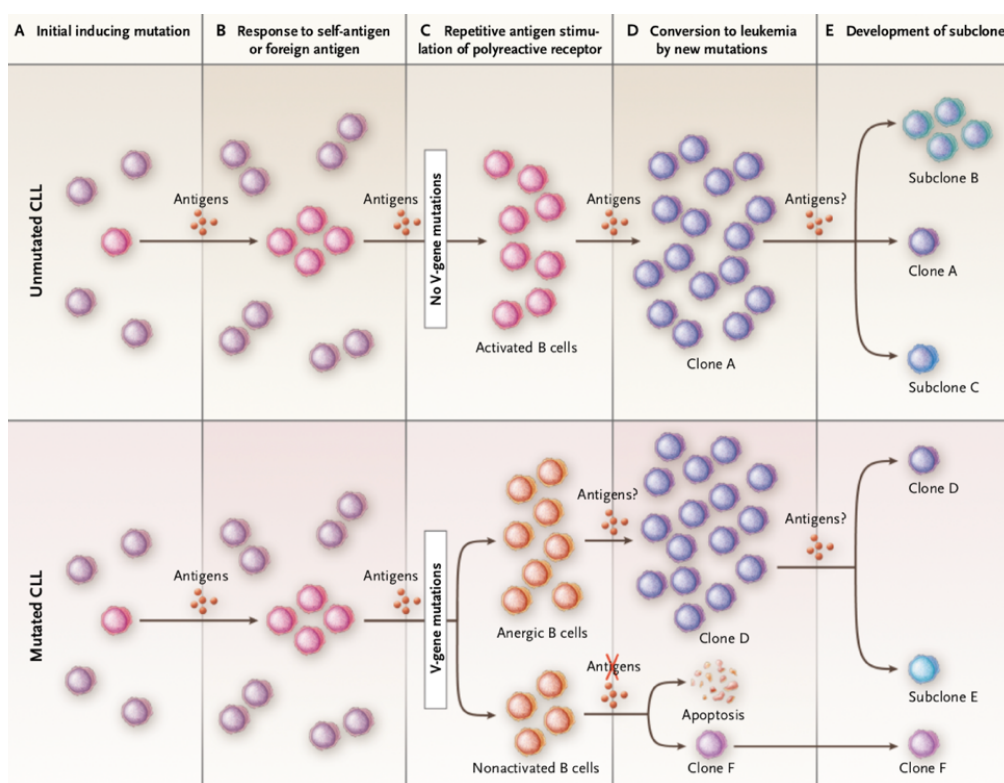


Figure 15. **Modèle de développement et d'évolution de la LLC**

En haut de la figure est représentée l'évolution de la cellule clonale « non mutée ». De la même façon, dans la partie inférieure de la figure, on retrouve le modèle « muté » avec la possibilité d'une perte d'affinité de l'antigène après mutation somatique des gènes V [d'après Chiorazzi et al. 2005 (32)]

Pour terminer, de nombreuses équipes ont évoqué la probabilité d'altérations géniques très précoces, dès le stade de CSH et qui seraient à l'origine de la maladie. En 1997, *Gahn et al.* mettent en évidence la trisomie 12 dans les CSH CD34+, CD19- (59-60). Plus récemment, d'autres équipes ont recherché la présence de mutations de *TP53*, *NOTCH1*, *SF3B1*, *FBXW7*, *MYD88* et *XPO1* dans les CSH CD34+, CD19- de patients atteints de LLC. Ils ont pu observer des mutations de *NOTCH1* et *XPO1* avec des charges alléliques comparables aux CD34-, CD19+ (61,62). Ces résultats confirment la présence de ces mutations au stade de la CSH et leur probable implication dans l'origine de la maladie. Pour les mutations de *SF3B1*, *TP53* et *FBXW7* les charges alléliques retrouvées dans les CD34+, CD19- étaient faibles donc ne confirment pas une altération de ces gènes dès le stade de CSH dans la LLC.

Enfin *Marsilio et al.* ont mis en évidence des mutations du gène *SF3B1* dans des cellules de LLC (Lin-, CD34+, CD38low, CD45RA-, CD90+/CD90-) avec des fréquences alléliques très faibles. Ils se sont alors posé la question d'une contamination des différents stades cellulaires par des stades plus matures. Ils soulignent les difficultés d'interprétation et les limites méthodologiques possibles de ces approches.

III.6. Traitements de la LLC

Le diagnostic de LLC n'implique pas obligatoirement la mise en place d'un traitement spécifique. La prise en charge consistera à faire une surveillance « *Wait and Watch* » afin de dépister un éventuel signe de progression. C'est une pathologie qui peut évoluer lentement selon les groupes de patients. Certains groupes peu évolutifs peuvent avoir une survie approchant la survie de la population générale.

III.6.1. Les indications de traitement

Les indications de traitement sont définies par l'IWCLL (« *International Workshop on CLL* ») et revues en 2018 (63) :

- L'insuffisance médullaire progressive avec l'apparition ou l'aggravation de cytopénie(s) (anémie et/ou thrombopénie)
- La présence d'une splénomégalie volumineuse (6 cm au-dessous de la marge costale), progressive ou symptomatique
- La présence d'adénopathies volumineuses (au moins 10 cm dans le plus long diamètre), progressives ou symptomatiques
- Une progression de la lymphocytose de plus de 50% sur 2 mois ou un temps de doublement (TDL) inférieur à 6 mois (pour les patients ayant une lymphocytose initiale inférieur à 30G/L, le TDL ne doit pas être utilisé pour définir à lui seul une indication de traitement)
- Une cytopénie auto-immune (anémie ou thrombopénie) ne répondant pas aux corticoïdes
- La présence de signes généraux tel que la perte de poids de 10% ou plus, non volontaire dans les 6 mois précédents, une asthénie significative, une fièvre de plus de 38°C pendant au moins 2 semaines sans signe d'infection, des sueurs nocturnes depuis plus d'un mois sans signe d'infection

Si un de ces critères est retrouvé lors du diagnostic ou apparaît au cours du suivi, le patient bénéficiera d'un traitement spécifique. Celui-ci sera différent, en fonction de la présence de certaines anomalies moléculaires et/ou cytogénétiques qui peuvent caractériser chaque patient atteint de LLC, comme nous l'avons décrit précédemment. Aujourd'hui, il y a 3 critères indispensables qui vont définir le traitement, la mutation et/ou délétion *TP53*, le statut mutationnel des IGHV et enfin, l'état général du patient au moment où le traitement est envisagé. Cette stratégie est résumée dans l'arbre décisionnel de la Figure 16.

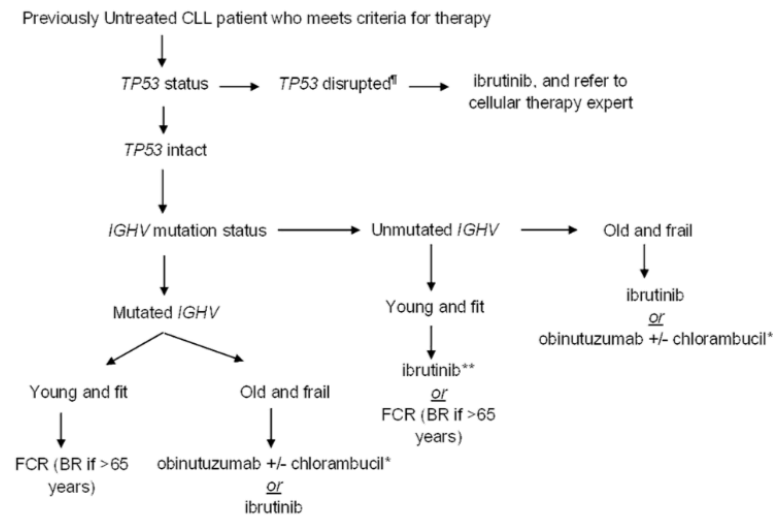


Figure 16. **Stratégie de traitement de la LLC d'après les critères de l'IWCLL 2018**
[d'après Parikh et al., 2018 (64)]

III.6.2. Les traitements

Pendant environ une quarantaine d'années, les traitements de la LLC n'ont que peu évolués (cf. Figure 17). Ils ont été restreints à l'utilisation de drogues dites « cytotoxiques » qui ont une action directe sur l'ADN. Les analogues des purines inhibent sa synthèse et les agents alkylants de la famille des moutardes azotées inhibent sa transcription et sa réplication. L'inconvénient majeur de ces molécules est qu'elles agissent sur toutes les cellules en division, c'est-à-dire sur les cellules tumorales comme sur les cellules non tumorales. La conséquence principale est l'importance des effets secondaires de ces drogues à court et à long terme.

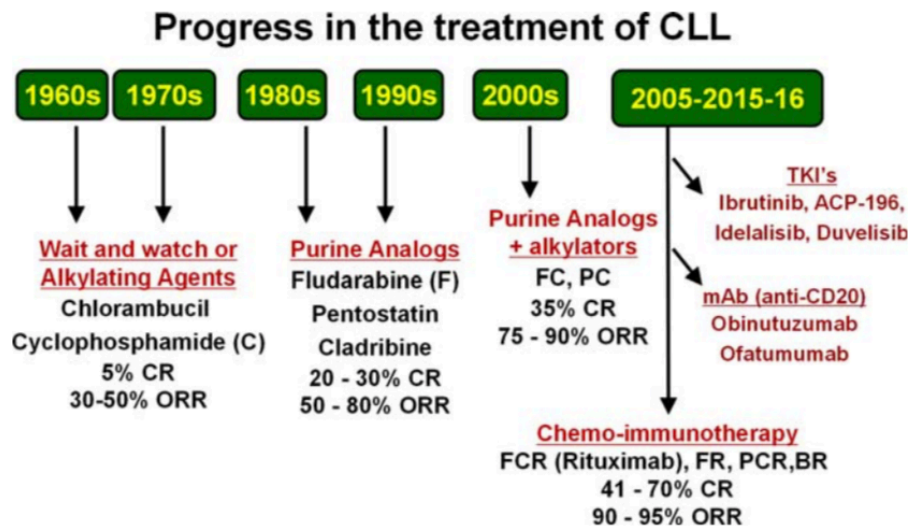


Figure 17. Évolution des traitements de la LLC

[d'après Rai et Jai, 2016 (65)]

Devant cet inconvénient commun aux principales drogues utilisées dans les cancers, il a été développé des drogues ciblées permettant de limiter la dégradation des cellules non tumorales. On voit apparaître au début des années 2000, une nouvelle molécule permettant de cibler l'antigène CD20, c'est le Rituximab. Cette molécule inhibe la cellule tumorale de manière spécifique. C'est un anticorps monoclonal anti-CD20 chimérique composé de deux fragments. Un fragment Fab qui cible l'Ag CD20 du lymphocyte B et un fragment Fc qui entraîne la lyse du lymphocyte (cytotoxicité dépendante du complément et des anticorps). Par la suite, d'autres anticorps monoclonaux se développent comme l'obinutuzumab qui est entièrement humanisé à la différence du rituximab chimérique. Un autre anticorps monoclonal humanisé a été développé, un anti-CD52, l'alentuzumab qui se lie aux antigènes CD52 présents à la surface des lymphocytes B et T majoritairement.

En fonction de l'âge du patient, différentes combinaisons de drogues sont utilisées. Pour les patients de moins de 65 ans, le traitement classique est l'association FCR [Fludarabine – Cyclophosphamide – Rituximab]. Pour les patients de plus de 65 ans, le traitement classique est l'association RB [Rituximab – Bendamustine].

Par la suite, il a été développé de nombreux inhibiteurs de tyrosines kinases. Parmi eux, deux molécules sont qualifiées d'inhibiteurs du BCR. D'une part, l'ibrutinib qui cible la Bruton kinase (BTK). Comme évoqué précédemment, la signalisation du BCR, dont BTK est un acteur, permet le maintien de la prolifération et de la survie de la cellule clonale. L'ibrutinib forme une liaison covalente au niveau du site actif de BTK, ce qui entraîne une inhibition enzymatique prolongée de celle-ci. L'ibrutinib possède deux AMM dans la LLC, tout d'abord en monothérapie chez des patients adultes non précédemment traités et la seconde est en monothérapie ou en association à la bendamustine et au rituximab (BR) chez les patients ayant reçu au moins un traitement antérieur (www.vidal.fr).

D'autre part, l'idélalisib cible la PI3K δ qui est hyperactive dans les lymphopathies B. PI3K δ a un rôle central dans la cascade de signalisation de la prolifération cellulaire tumorale. L'idélalisib est toujours associé à un anticorps monoclonal anti-CD20 et possède deux AMM dans la LLC. La première concerne les patients ayant eu un traitement antérieur et la seconde est le traitement de première intention des patients avec une anomalie de *TP53* (mutation ou délétion) et qui ne sont éligibles à aucun autre traitement (www.vidal.fr).

Dernièrement, une autre classe de thérapie ciblée est apparue, les inhibiteurs de *Bcl2* tel que le vénétoclax. Nous l'avons évoqué dans la physiopathologie de la LLC, il y a une surexpression des facteurs anti-apoptotiques. Cette nouvelle molécule se lie directement au domaine BH3 de *Bcl2*, ce qui permet l'initiation de la perméabilisation de la membrane mitochondriale externe et l'activation des caspases nécessaire à l'apoptose.

III.6.3. Les résistances aux traitements ciblés

De récentes études montrent l'apparition de résistances acquises vis-à-vis des nouvelles molécules telles que l'ibrutinib, l'idélalisib ou le vénétoclax. Les mécanismes de résistances à l'ibrutinib reposent sur la sélection de clones tumoraux porteurs de mutations de BTK ou de PLC γ 2 (66). La résistance à ces traitements favorise une évolution de la maladie, ce qui encourage l'acquisition de nouvelles mutations (67). Concernant la résistance au vénétoclax, une mutation de *Bcl2* a été mise en évidence. Elle entraîne une diminution de l'affinité du vénétoclax pour le *Bcl2* (68,69).

IV. Objectifs de la thèse

Comme nous avons pu l'évoquer précédemment, la question de la cellule à l'origine de la LLC n'est pas encore clairement élucidée. Les données récentes de séquençage haut débit ont révélé qu'il existe d'assez nombreuses mutations conductrices pouvant promouvoir le développement de ce syndrome lymphoprolifératif telles les mutations touchant *TP53*, *NOTCH1*, *SF3B1* ou encore *MYD88* (31,36). De manière très intéressante, plusieurs équipes viennent de démontrer que chez certains patients, ces mutations sont également présentes dans les cellules souches hématopoïétiques (CSH) (61,62,70). Elles seraient donc acquises très tôt, au cours du processus de lymphomagenèse. Ainsi, il existe vraisemblablement une histoire naturelle de la LLC débutant dès le stade CSH.

Nous savons d'autre part que le répertoire IgHV des LLC est « biaisé », différent du répertoire B normal et que chez 30% des patients il peut être identifié un BCR stéréotypé (*subset*) à la surface des cellules tumorales (46,48). Ces données révèlent donc que la sélection par le BCR, et donc la spécificité antigénique, semble jouer un rôle important dans le développement du clone de LLC. De plus, comme vu précédemment, la présence de certains BCRs stéréotypés est très fortement associée à certaines altérations géniques.

Cependant, ce BCR est acquis très tôt dans au cours de la lymphopoïèse et sa région variable est modelée aux stades immatures très proches de la CSH, dans la moelle osseuse (à l'exception des hypermutations somatiques). Nous nous sommes posés la question d'une autre histoire naturelle possible de la LLC, qui débiterait par la production d'un BCR « favorisant » la LLC et qui serait un BCR stéréotypé.

L'objectif principal de ce travail a été de rechercher la présence de séquences stéréotypées type « LLC » chez des patients exempts d'hémopathie. Le but principal était de savoir si ces séquences étaient produites chez des patients dits « sains » puis éventuellement à quelle fréquence elles pouvaient être retrouvées (sont-elles ubiquitaires ? sont-elles retrouvées uniquement chez certains individus ?).

Pour répondre à cette question, nous avons donc mis au point le séquençage haut débit des gènes des immunoglobulines à partir d'ADN ainsi qu'un *pipeline* bio-informatique pour l'analyse des données (script R). La validation de cette méthode a constitué l'objectif secondaire de mon travail de thèse.

V. Matériels et Méthodes

V.1. Consentement

Un consentement et/ou un formulaire de non-opposition à la recherche biomédicale a été obtenu pour chaque patient et témoin. Les analyses ont été effectuées sur les reliquats de tubes de numération qui ont été requalifiés par le Centre de Ressources Biologiques (CRBioLim) de notre institution, auprès duquel une demande de session a été réalisée.

V.2. Constitution d'un pool de témoins sains et de la cohorte de patients

Le pool de témoins a été réalisé à l'aide de 12 sujets âgés de 24 à 38 ans, 6 femmes et 6 hommes ayant une lymphocytose moyenne à 2,5 G/L.

Nous avons constitué une cohorte de 19 patients (11 hommes et 8 femmes) ayant une lymphocytose entre 1,7 et 3 G/L, âgés de 55 à 76 ans avec un âge médian à 66 ans.

Tous les sujets composants le pool de témoins et la cohorte n'avaient pas d'antécédent d'hémopathie maligne, de radiothérapie, de chimiothérapie ni de prise de médicaments modulateurs du système immunitaire. De plus, l'absence de clone B circulant a été vérifiée au préalable par cytométrie en flux.

V.3. Cytométrie en flux

Le protocole utilisé est détaillé en Annexe 1. Les anticorps utilisés pour l'immunophénotypage des lymphocytes B sont le CD19 – PE Cy5.5 Immunotech [clone J4.119], le CD5 – APC Immunotech [clone BL1a] et enfin le Kappa FITC / Lambda PE Dual color Dako [clone A8B5 + N10.2 poly lapin poly F(ab')₂].

L'analyse est réalisée sur un Navios, Beckman Coulter® (Brea, CA) et retraitée avec le logiciel Kaluza version 2.1 (Beckman Coulter®).

V.4. Tri cellulaire

Un tri positif des lymphocytes B est fait sur 3 à 4mL de sang total, prélevé sur tube EDTA. Le kit utilisé est le kit EasySep™ HLA Chimerism Whole Blood CD19 Positive Selection Kit de manière automatisée avec le RobotSep de Stemcell® (Grenoble, France). Le kit et l'automate ont été utilisés selon les recommandations du fournisseur.

V.5. Extraction d'ADN

L'extraction d'ADN est faite sur l'automate Promega – Maxwell® (San Diego, CA). Le kit utilisé est le kit d'extraction à partir de sang total Promega – Maxwell® RSC Whole Blood DNA Kit. Nous avons utilisé le kit selon les recommandations du fournisseur.

Les concentrations d'ADN ont été mesurées par Nanodrop® et Qubit® 2.0 Fluorometer (kit BR – Broad Range).

V.6. Vérification du profil de clonalité par analyse de fragment

Le profil de clonalité des ADN clonaux a été vérifié par une analyse de fragment selon le protocole consensuel BIOMED-2 (71). Brièvement, c'est une amplification réalisée avec les amorces localisant le locus FR1 des immunoglobulines, grâce à l'AmpliTaq Gold™ (Thermofischer, Waltham, MA). L'analyse est faite à l'aide du Genemapper Software® (Thermofischer).

V.7. Vérification de la qualité des ADN utilisés

Il s'agit d'une PCR de contrôle de la qualité des ADN fournie dans le kit TRG Clonality Assay Invivoscribe® (San Diego, CA). Elle permet d'obtenir des tailles de fragments à 100, 200, 300, 400, 500 et 600pb. La présence et la taille de chaque pic permet de valider l'absence de dégradation de l'ADN.

V.8. Préparation des librairies

V.8.1. Principe de préparation des librairies par PCR

La stratégie utilisée a été développée précédemment au sein de notre équipe de recherche « Contrôle des réponses immunes B et des lymphoproliférations » (CRIBL) – UMR 7276/INSERM 1262 dirigée par le Pr Michel Cogné. Avec l'aide du Dr Alexis Saintamand, nous l'avons adaptée à l'amplification des réarrangements IgH humains à partir des régions V-Leader/FR1 et sur une matrice ADN.

Le principe est de réaliser une première PCR afin d'amplifier les séquences d'intérêts (PCR-1). Ensuite une 2nde PCR est faite, permettant de terminer la construction des adaptateurs et d'ajouter les index pour identifier chaque échantillon lors du séquençage (PCR-2) (cf. Figure 18).

Cette stratégie permet également d'introduire des barcodes moléculaires ou UMI (« *Unique Molecular Identifiers* ») lors de la première PCR. Ce sont des séquences courtes servant de code – barre moléculaire. Ils sont ajoutés aux fragments d'ADN amplifiés dans le but de pouvoir identifier chaque molécule d'ADN amplifiée. Ils sont aussi utiles pour réduire les erreurs de séquençage et les biais quantitatifs introduits par amplification.

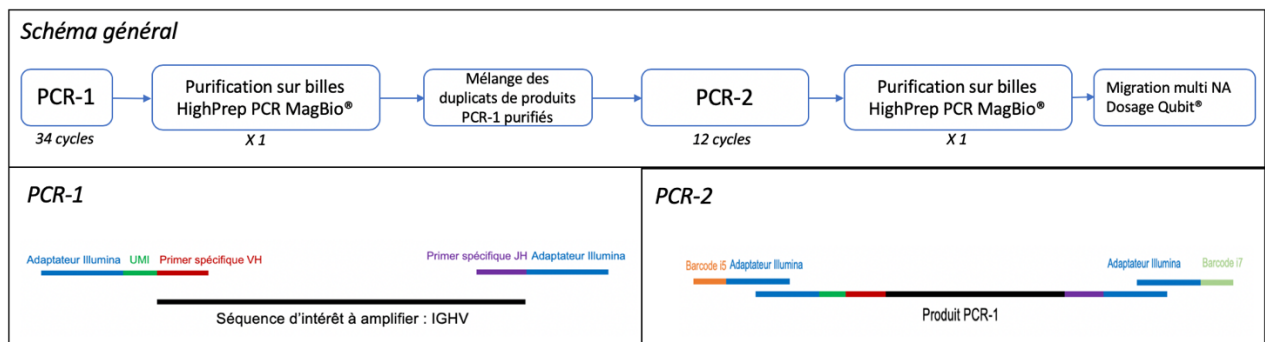


Figure 18. Stratégie de préparation des librairies

V.8.2. PCR-1

V.8.2.1. Amorces Leader VH

Nous avons testé une approche avec des amorces ciblant la région Leader des IGHV en utilisant une combinaison de 17 amorces sens VH et 1 amorce anti-sens JH consensus (cf. Annexe 2). Ces différentes amorces ont été désignées et sont utilisées au sein du laboratoire pour la recherche des mutations somatiques des patients LLC. En revanche, elles n'avaient jamais été utilisées au sein d'une PCR multiplex.

Le principal intérêt des amorces Leader est d'obtenir la séquence totale du gène VH réarrangé. Ainsi, nous pouvons gagner en précision quant au taux de mutations somatiques et à l'identification des différents allèles IgHV.

V.8.2.2. Amorces FR1 BIOMED-2

Une seconde approche a été testée en utilisant les amorces consensuelles BIOMED-2 ciblant le locus FR1 des immunoglobulines (71) (cf. Tableau 5). Ces amorces ont été largement validées et sont utilisées en PCR multiplex.

5' – 3'	Adaptateur Illumina	UMI	Séquence spécifique	Taille
VH1	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GGCCTCAGTGAAGGTCTCCTGCAAG	68
VH2	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GTCTGGTCTTACGCTGGTGAAACCC	68
VH3	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CTGGGGGTCCCTGAGACTCTCCTG	68
VH4	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CTTCGGAGACCTGTCCCTCACCTG	68
VH5	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CGGGGAGTCTCTGAAGATCTCCTGT	68
VH6	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	TCGAGACCTCTCACTCACCTGTG	68
JH consensus	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG		CTTACCTGAGGAGACGGTGACC	56

Tableau 5. Construction des amorces PCR-1

Dans la colonne « séquence spécifique » : ce sont les amorces BIOMED-2 qui permettent d'amplifier les régions d'intérêts.

Dans la colonne « UMI » : la séquence des UMIs ajoutés. Pour l'amorce JH consensus, il n'a pas été ajouté de séquence UMI.

Dans la colonne « Adaptateur Illumina » : la première partie des adaptateurs Illumina.

V.8.2.3. Mélange réactionnel et programme d'amplification

La PCR-1 a été réalisée avec l'AmpliTaq Gold™ qui est une ADN polymérase optimisée pour l'amplification des gènes des immunoglobulines. Dans le but de ne pas saturer les capacités de l'enzyme et d'obtenir la meilleure sensibilité de notre méthode, nous avons choisi de réaliser la PCR-1 en duplicat avec une quantité maximale d'ADN fixée à 250ng par tube de PCR.

Pour les ADNs de patients ne permettant pas d'obtenir la quantité optimale de 500ng, le volume maximum d'ADN (soit 15µL/tube) a été utilisé pour réaliser le mélange réactionnel (cf. Tableau 6).

Les conditions classiques déjà utilisées au sein du laboratoire pour la réalisation du mélange réactionnel de cette PCR n'étant pas optimal, nous avons fait varier plusieurs paramètres avant de définir les conditions à utiliser. Nous avons fait varier la concentration de chaque amorce (5µM – 10µM – 20µM), la concentration de l'ADN polymérase (AmpliTaq Gold™ à 1 unité/tube ou 5 unités/tube) et enfin l'apport du DMSO pour favoriser la spécificité d'hybridation des amorces.

Nous avons finalement retenu l'utilisation d'amorces à 5µM, de la polymérase pure (5 unités par tube) et l'ajout de DMSO. Le mélange réactionnel de la PCR-1 est présenté dans le Tableau 6.

Pour le choix du programme, nous avons utilisé le programme classique des amorces BIOMED-2 (déjà utilisé au sein du laboratoire). Ce programme d'amplification à une première étape de dénaturation initiale de 7 minutes à 95°C, ensuite 34 cycles avec une dénaturation de 30 secondes à 95°C, une hybridation d'une minute à 60°C et une élongation d'une minute et 30 secondes à 72°C. L'élongation finale dure 10 min à 72°C.

Réactifs	Volume / tube (µL)
Amorces VH1 à VH6 (5µM)	1
JH consensus primer (5µM)	1
dNTP (10µM)	1
Tampon 10X Gold	5
MgCl ₂	3
Taq Gold pure	1
DMSO	3
Eau	Qsp 50

Tableau 6. Mix PCR-1

V.8.3. Étapes de purification

Les purifications sont réalisées avec des billes HighPrep PCR Magbio®. Ces étapes sont nécessaires pour éliminer les amorces résiduelles et les produits d'amplification non spécifiques de petite taille comme les dimères d'amorces.

Le volume relatif de billes utilisé pour chaque purification est de 1,8X. Le protocole utilisé est celui recommandé par le fournisseur. Les étapes de purification ont été réalisées par série de 6 patients.

V.8.4. PCR-2

Cette étape a pour but d'ajouter un index aux différents échantillons afin de pouvoir les identifier lors du séquençage. Nous avons utilisé un double indexage. La référence des index utilisés est NEXTERA XT index kit V2 Illumina®.

Les amorces de la seconde PCR contiennent la seconde partie des adaptateurs et les index nécessaires à l'identification de chaque échantillon (cf. Figure 18).

Pour cette PCR-2, le mélange réactionnel a été établi d'après les données fournisseurs de l'ADN polymérase utilisée : la Phusion® High-Fidelity DNA Polymerase (cf. Tableau 7).

Le programme PCR est composé d'une première étape de dénaturation initiale de 30 secondes à 98°C, puis 12 cycles avec une dénaturation de 10 secondes à 98°C, une hybridation de 30 secondes à 62°C et enfin une élongation de 30 secondes à 72°C. L'élongation finale dure 5 min à 72°C.

Il est nécessaire pour cette seconde étape de ne pas avoir trop de cycles de PCR pour limiter l'amplification de produits non spécifiques.

Réactif	Volume / tube (µL)
Index i5 (sens) (10µM)	1
Index i7 (anti-sens) (10µM)	1
dNTP (10µM)	0,5
Phusion Buffer 5X	5
Phusion DNA Polymerase	0,25
Produit PCR-1	2
Eau	15,25
Volume total = 50 µL	

Tableau 7. Mix PCR-2

V.9. Qualification des bibliothèques

V.9.1. Vérification de la présence d'un produit spécifique

Une migration par électrophorèse capillaire avec le multi NA de Shimadzu® (Kyoto, Japon) est réalisée après la purification des produits de PCR-1 et PCR-2.

Après PCR-1, la taille des produits PCR est d'environ 400pb.

Après PCR-2, la taille des produits PCR est comprise entre 450 et 500pb.

V.9.2. Quantification ADN

Un dosage fluorimétrique (Qubit® 2.0 fluorometer, kit ds DNA HS assay– High Sensitivity) est réalisé pour quantifier l'ADN de chaque bibliothèque.

V.10. Séquençage Illumina

V.10.1. Principe du séquençage Illumina

Le séquençage Illumina® (San Diego, CA) est fait par synthèse de nucléotides.

Comme expliqué précédemment, la bibliothèque correspond à un pool de fragments d'ADN cibles à séquencer. Ils sont flanqués des adaptateurs i5 et i7 en 5' et en 3' respectivement. Les adaptateurs permettent de fixer les ADN à séquencer sur la cellule de lecture (*i.e.* : *flowcell*) et chaque index ajouté permet d'identifier les bibliothèques provenant de différents échantillons biologiques (multiplexage). Une fois dénaturée en ADN simple brin à la soude (NaOH), seule une faible quantité de bibliothèque est déposée sur la *flowcell*. Les molécules d'ADN à séquencer se répartissent aléatoirement sur la surface et « à distance » les unes des autres afin d'éviter la « surcharge » de la *flowcell*. De manière à obtenir un signal d'intensité suffisante du signal pour la lecture de la séquence, chaque molécule d'ADN fixée est amplifiée par PCR directement sur la *flowcell* grâce à un système d'amorçage de proche en proche. On obtient à partir d'une molécule cible initiale un « *cluster* » de molécules reproduites à l'identique (*cf.* Figure 19A). Elles sont porteuses de la même séquence nucléotidique exactement. Le séquençage commence par l'extrémité libre des molécules flanquées de l'adaptateur i5 (amorce de séquençage universelle) alors que l'adaptateur i7 permet l'ancrage sur la *flowcell*.

La technologie Illumina® offre la possibilité de réaliser un séquençage « *paired – end* ». Dans ce cas, les molécules du *cluster* sont retournées après la première lecture (*read 1*) pour fixer cette fois-ci l'adaptateur i5 sur la *flowcell* et produire une séquence (*read 2*) à partir de l'adaptateur i7. Dans ce type de séquençage, nous avons le choix de la taille de lecture dans les deux sens du même brin d'ADN initial. Nous pouvons par exemple pour une séquence de 300pb amplifiée obtenir une lecture de 300pb dans chaque sens. Au contraire, pour une séquence de 900pb amplifiée, il est possible d'obtenir une lecture de 500pb dans un sens et 100pb dans l'autre sens, tout dépend de la région d'intérêt choisie à séquencer.

Une fois les clusters formés, ils vont être analysés. A chaque cycle, il y a ajout des 4 bases (A, T, G et C) qui sont chacune marquées par un fluorochrome différent. Une fois l'une d'elle liée à la séquence, il y a une excitation lumineuse puis l'acquisition d'une image permettant de définir qu'elle base a été lue. Ceci est recommencé à chaque cycle (cf. Figure 19B). C'est une chimie spécifique de la technologie Miseq, la stratégie de marquage des nucléotides est différente sur les autres appareils Illumina®.

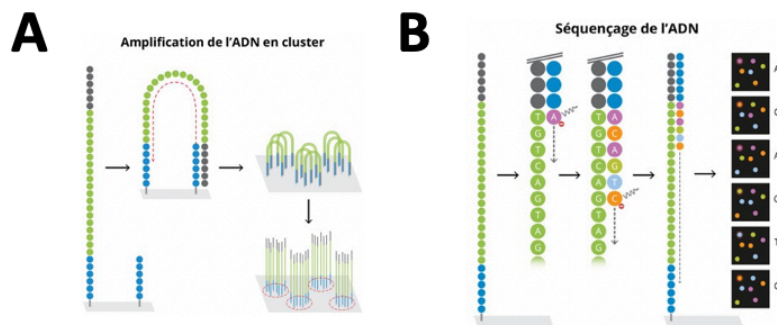


Figure 19. Séquençage Miseq Illumina® par synthèse de nucléotides

- A. Représentation d'un amplicon fixé sur la flowcell qui est multiplié pour former des clusters.
- B. La synthèse d'un brin complémentaire permettant l'émission d'un fluorochrome à chaque nouvelle base incorporé. Chaque base possède un fluorochrome différent.

A chaque cycle est évalué la pureté du signal (C) pour chaque cluster qui représente 1 000 molécules identiques. Le signal doit être identique pour les 1 000 molécules. Ainsi, le rapport des intensités des différentes bases lues détermine la pureté du signal (cf. Figure 20). Lorsque valeur est proche de 1, cela signifie que le signal est pur, c'est-à-dire qu'une seule base est lue. Lorsque ce score est inférieur à 0,6 deux fois, le *cluster* est éliminé et ne sera pas disponible à l'analyse car considéré comme ayant une qualité insuffisante.

Les *reads* qui passent ce filtre appelé *passing filter* sont considérés comme ayant un signal de lecture des bases assez « pur », permettant une analyse correcte des séquences. Un seuil de 85% des *reads* totaux passant ce *passing filter* est considéré comme acceptable.

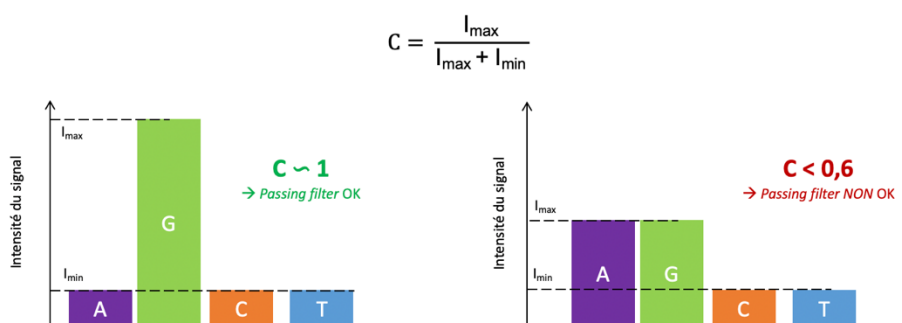


Figure 20. Évaluation de la qualité du signal à chaque cycle

V.10.2. Préparation des bibliothèques au séquençage

Dans ce travail, nous avons réalisé les séquençages avec deux types de kit.

Le premier kit utilisé est le kit Illumina® Nano V2 MiSeq (500 cycles). On peut obtenir une densité de clusters comprise entre 865 et 965 k/mm². Le séquençage dure environ 17h et le nombre de *reads* obtenu est de 1 million en *single read* ou de 2 millions en *paired-end*. Le pool de bibliothèques est utilisé à 10pM avec 15% de PhiX.

Le PhiX est une solution de contrôle qui permet un maintien de la diversité des lectures de bases sur la *flowcell*. Si les bibliothèques ont des séquences trop homogènes entre elles, le séquenceur aura des difficultés de lecture car ce sera toujours les mêmes bases lues. Pour contrer ce phénomène, le PhiX est nécessaire. La quantité de PhiX ajoutée est d'autant plus importante que les bibliothèques sont homogènes. Nous avons fixé sa quantité à 15% d'après les recommandations des ingénieurs d'Illumina®.

Le second kit utilisé est le kit Illumina® V3 MiSeq (600 cycles). On peut obtenir une densité de clusters comprise entre 1 200 et 1 400 k/mm². Le séquençage dure environ 56h et le nombre de *reads* obtenu est de 22 à 25 millions en *single read* ou de 44 à 50 millions en *paired-end*. Le pool de bibliothèques est utilisé à 12pM avec 15% de PhiX.

V.11. Analyse bio-informatique post-séquençage

L'analyse se déroule en 5 grandes étapes successives (cf. Figure 21). La première étape est le démultiplexage de l'ensemble des *reads*. Cette étape est réalisée directement par le séquenceur lui-même à la fin du séquençage. Elle permet d'éliminer les index de chaque *read* pour pouvoir identifier les échantillons. On obtient ainsi des fichiers fastq contenant les séquences sans index qui sont identifiées selon la feuille de travail établie au préalable.

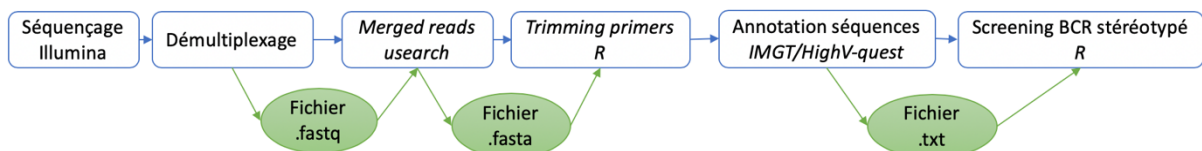


Figure 21. Stratégie d'analyse des données de séquençage

V.11.1. Analyse de la qualité des *reads* obtenus

Un fichier Fastq est un format universel utilisé en séquençage haut débit. Il s'agit d'un fichier texte consensuel permettant de stocker les données de séquençage et leur score qualité associé (cf. Figure 22). La séquence et le score sont codés par un caractère ASCII.

V.11.5. Recherche des BCRs stéréotypés type « LLC »

La recherche est faite à l'aide d'un script R. Celui-ci est divisé en plusieurs étapes successives (cf. Figure 23). Les critères définissant chaque *subset* ont été établis grâce aux données de la littérature (46,47). Ils ont été regroupés dans le tableau de la Figure 23.

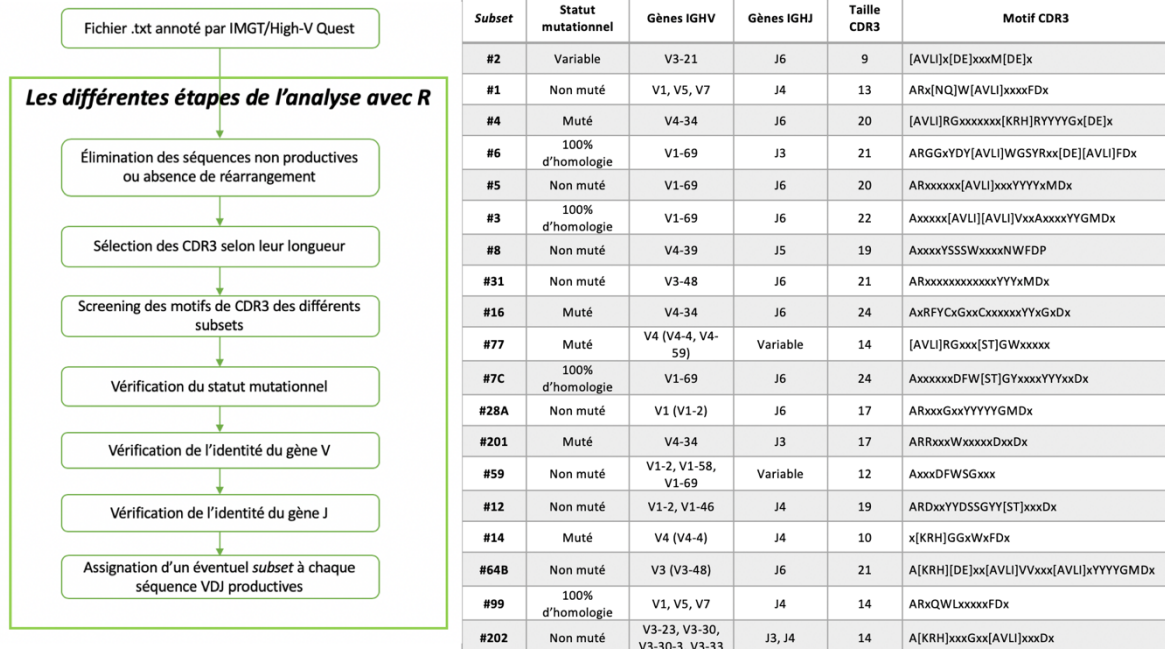


Figure 23. Stratégie de recherche des séquences stéréotypées et tableau des *subsets* majeurs de type "LLC"

[le tableau est adapté d'après Agathangelidis et al., 2012 (47) et ARRest/Assign Subsets (<http://tools.bat.infspire.org/arrest/assignsubsets/>)]

V.11.6. Confirmation des *subsets* assignés avec l'outil ARRest/Assign Subsets

Pour chaque séquence analysée comme étant une séquence stéréotypée de type « LLC », nous avons soumis ces mêmes séquences sur la plateforme en ligne ARRest/Assign subsets (<http://tools.bat.infspire.org/arrest/assignsubsets/>). Cet outil nous a permis de valider notre approche par screening de séquence par « identité ». ARRest/Assign subsets a été développé par l'équipe de *Bystry et al.* pour faciliter l'assignation des séquences clonales aux différents *subsets* (72). Grâce à une cohorte de plusieurs milliers d'échantillons de patients LLC (> 20 000), ils ont pu développer avec une statistique bayésienne un score de probabilité d'appartenance de la séquence à un *subset* défini (49).

V.11.7. Phylogénie moléculaire

Pour observer la phylogénie moléculaire potentielle des séquences identifiées comme *subset*, nous avons utilisé le logiciel Seaview® version 4.7.

VI. Résultats

VI.1. Évaluation du pourcentage de lymphocytes B

Pour chaque échantillon utilisé, nous avons vérifié l'absence de clone B circulant par cytométrie en flux. Nous avons pu évaluer le pourcentage de lymphocytes B pour chaque échantillon. En moyenne, la population étudiée a 3% de lymphocytes B parmi les leucocytes totaux (de 0,7 à 6%).

VI.2. Évaluation de la performance du tri cellulaire et de l'extraction d'ADN

Après tri cellulaire, nous avons obtenu un nombre entre 120 000 à 591 000 cellules avec un nombre médian de cellules à 215 000 (cf. Figure 24A). Les rendements obtenus varient entre 15 et 60% avec médiane de 32%.

Les concentrations d'ADN des cellules CD19+ triées varient de 1,91ng/μL à 33,2ng/μL au Qubit® (cf. Figure 24B). Les ratios A280/A260 au Nanodrop varient entre 1,6 et 1,7. Il n'y avait pas de relation directe évidente entre la quantité de cellules B triées et la concentration des ADN obtenus (cf. Figure 24B).

Pour réaliser chaque librairie, nous avons utilisé au maximum de 500ng d'ADN de lymphocytes B triés. Pour les 19 patients de la série, nous avons utilisé entre 46 et 500ng d'ADN de lymphocytes B triés.

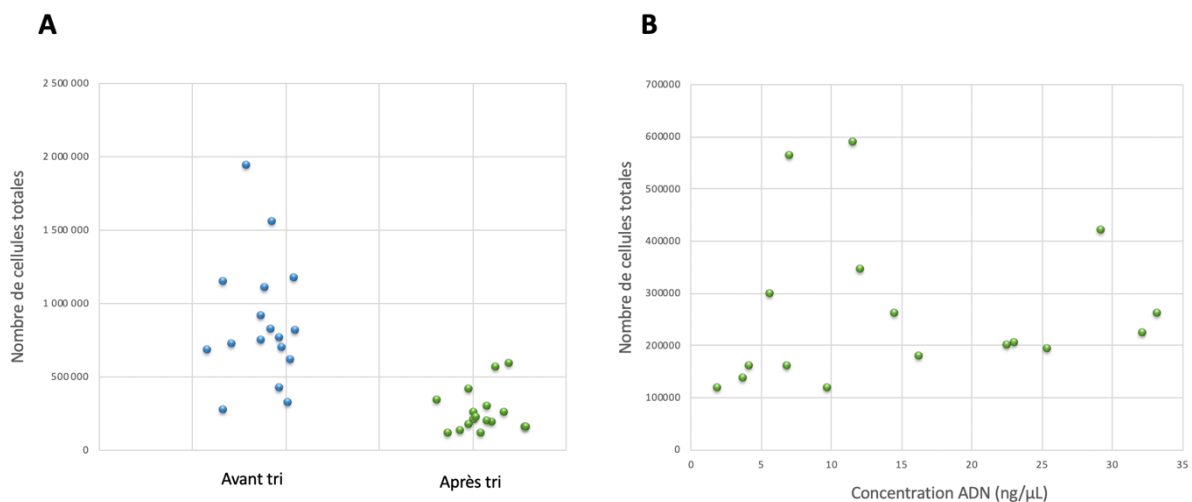


Figure 24. Évaluation de la performance du tri cellulaire et de l'extraction d'ADN

VI.3. Préparation des librairies

Les librairies ont été réalisées avec deux PCRs successives ainsi qu'une purification après chacune. Tout d'abord, nous avons voulu utiliser des amorces ciblant la région Leader des immunoglobulines. L'utilisation de ces amorces isolées dans une PCR fonctionne. En revanche, l'ensemble des 17 amorces VH dans le même mélange réactionnel a mis en évidence des difficultés techniques. Pour y remédier, nous avons modifié la concentration des amorces (5μM – 10μM – 20 μM), utilisé différentes ADN polymérase (Phusion® High-Fidelity

DNA Polymerase, Takara PrimeSTAR® Max DNA Polymerase et l'AmpliAq Gold™). Malgré ces modifications de mélange réactionnel, lors de la vérification de l'amplification par migration électrophorétique, une bande non spécifique persiste après la seconde purification à une taille d'environ 200pb pour une taille de la région d'intérêt entre 500 et 700pb (cf. Figure 25).

En parallèle, nous avons testé les amorces FR1. Avec celles-ci, nous n'avons pas observé ce phénomène d'amplification non spécifique, nous avons obtenu une seule bande spécifique à environ 450 – 500pb comme attendu. Nous avons donc poursuivi la suite de la préparation des bibliothèques avec les amorces FR1.

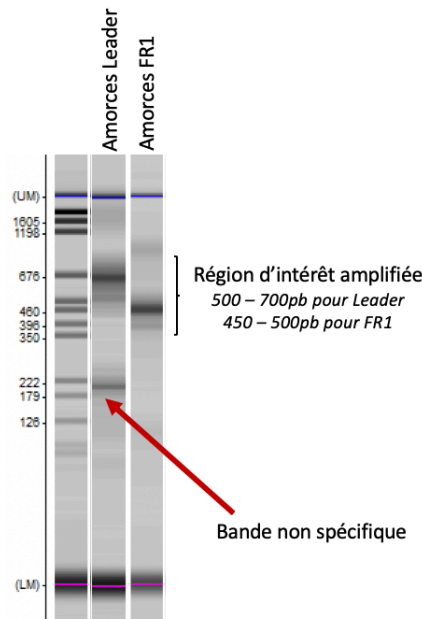


Figure 25. **Migration sur électrophorèse capillaire multi Na**

Exemple d'un échantillon polyclonal amplifié avec les amorces Leader et FR1.

Les concentrations d'ADN des bibliothèques obtenues varient de 5,54ng/μL à 8,04ng/μL c'est-à-dire 19,5 à 27,4nM après conversion. Chaque bibliothèque a été diluée pour obtenir un pool homogène à 4nM nécessaire au séquençage.

VI.4. Évaluation de la qualité et des performances du séquençage

Dans un premier temps, nous avons évalué la méthode de préparation des bibliothèques en réalisant un séquençage de 5 échantillons avec un kit Nano V2 Illumina®.

Ces 5 échantillons étaient composés d'un pool témoins (12 ADN polyclonaux), un pool de LLC (14 ADN de LLC différents), deux dilutions du pool de LLC dans le pool témoins (une dilution au 1/2 et une dilution au 1/5) et un ADN clonal de LLC avec une séquence stéréotypée (LLC#1). Pour chaque échantillon, nous avons vérifié les différentes métriques du séquençage, c'est-à-dire le nombre de *reads* obtenus par échantillon, la taille des *reads* ainsi que la qualité des séquences.

Un second séquençage a été réalisé avec un kit V3 Illumina®. Cette fois – ci les échantillons séquencés étaient 3 dilutions du pool de LLC dans le pool témoins (1/50, 1/100, 1/500) et les 19 patients de la cohorte (cf. VI.5 page 68).

VI.4.1. Analyse du nombre de reads

Le nombre de *read* par échantillon est homogène et optimal. Pour un séquençage avec une puce nano V2, nous avons obtenu au total 2 063 506 *reads* avec une lecture en *paired – end*, correspondant à 1 031 753 de *clusters*. Le rendement théorique pour une puce de ce type est d'environ 2 000 000 de *reads* en *paired – end*. Nous avons donc utilisé cette puce au maximum de ces capacités.

Au total, 94% des *reads* ont passé le « *passing filter* », ce qui est synonyme d'une très bonne qualité de séquençage.

La moyenne du nombre de *reads* est optimale (cf. Tableau 8). Elle est à 206 351 *reads paired-end* pour une moyenne théorique à 250 000 *reads*. Un seul échantillon a présenté un nombre de *reads* inférieur à 200 000 *reads* (156 515 *reads*).

L'assemblage de chaque *read* réalisée avec l'outil *usearch* nous a permis d'obtenir en moyenne 86% des *reads* avec un *merged* correct. Pour rappel, les conditions requises pour que l'assemblage des deux *reads* soit réalisé sont, $\geq 80\%$ d'*overlap* entre les *reads* et moins de 2 *missmatch*s sont autorisés. La principale cause de rejet des séquences lors de l'assemblage des *reads* est la présence de plus de 2 *missmatch*s (cf. Tableau 8).

	POOL LLC	DILUTION 1/5	DILUTION 1/2	LLC #1	POOL TEMOINS
Nombre de <i>reads</i> totaux par échantillon (x 1 000 <i>reads</i>)	156	207	211	212	243
Proportion de <i>reads merged</i> (%)	78	86	83	94	88
Nombre de <i>reads</i> obtenus après <i>merged</i> (x 1 000 <i>reads</i>)	122	178	176	200	214
Nombre de <i>reads</i> éliminés avec > 2 <i>missmatch</i> s (x 1 000 <i>reads</i>)	32	27	33	8	25

Tableau 8. **Analyse du nombre de reads**

Analyse du nombre de reads pour un séquençage de 5 échantillons sur une puce nano V2. La première ligne indique le nombre de séquences obtenues dans un sens de lecture.

VI.4.2. Qualité de séquençage

La qualité de lecture des bases lors du séquençage est optimale. La majorité des bases lues ont un score compris entre Q30 et Q40 (cf. Figure 26). Ce qui a pu être observé, c'est qu'au niveau du *read* 2, il y a une perte de qualité systématique à partir de la moitié du *read*. Le score de qualité reste supérieur à Q20.

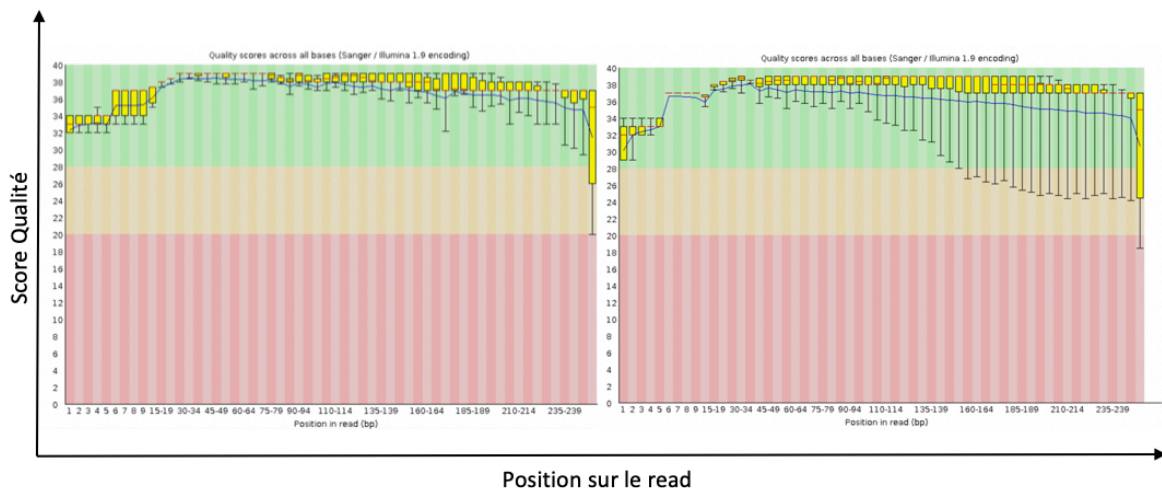


Figure 26. **Représentation graphique de la qualité de chaque base séquencée**

Calcul de la moyenne (et écart-type) du score de qualité de chaque base séquencée. En vert est représenté la tranche de score qualité entre Q30 et Q40. C'est la catégorie optimale, c'est-à-dire que le risque d'erreur de séquençage est entre 1 base sur 1 000 (Q30) à 1 base sur 10 000 (Q40). La majorité des séquences lues ont une précision comprise entre 99,9% et 99,99%.

Pour la tranche orange, la précision est entre 99% (Q20) et 99,9% (Q30).

VI.4.3. Analyse de la taille des reads obtenus

Pour la puce V2 nano, la taille des reads lue est de 250. Nous avons remarqué que moins de 10% des reads avaient une taille inférieure à 250 (cf. Figure 27). Nous avons donc analysé ses séquences de petite taille pour connaître leur nature, il s'est avéré que c'était des dimères d'amorces (cf. Figure 28).

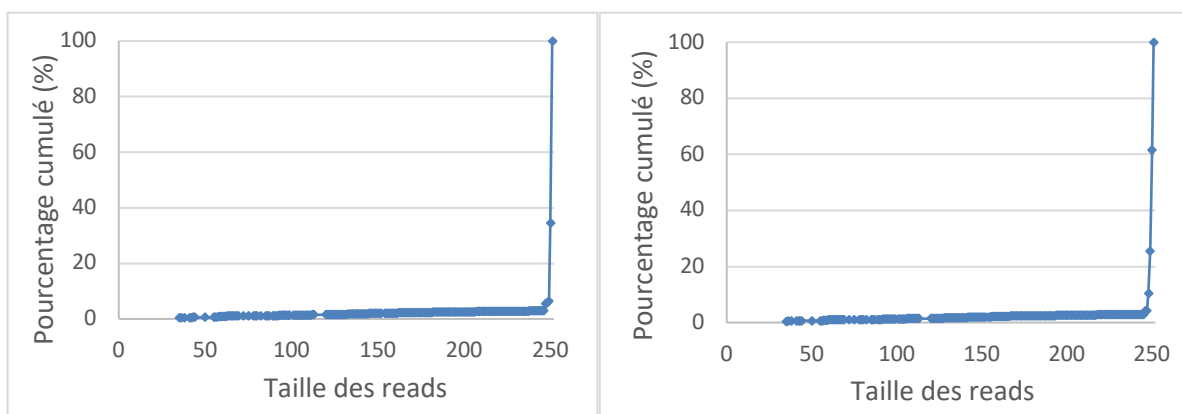


Figure 27. **Répartition de la taille des reads**

Exemple pour un même échantillon avec le read 1 à gauche et le read 2 à droite.

En ordonné, est représenté le pourcentage cumulé de chaque taille de read. Il n'y a que quelques reads (< 10%) qui ont une taille inférieure à 250. Ce résultat est homogène entre la lecture du read 1 et du read 2.

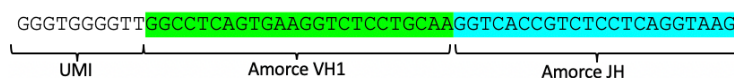


Figure 28. **Analyse d'une séquence de taille inférieure à 250**

Exemple d'analyse d'une séquence de taille 56. On peut voir que cette séquence ne contient que l'amorce V sens (amorces V + UMI) et l'amorce J anti-sens

VI.4.4. Analyse des séquences *Unique Molecular Identifiers* (UMI)

Nous avons ajouté dans les amorces de la PCR-1 des séquences UMIs de 10 nucléotides. Nous les avons analysées à l'aide d'un script R. Pour cela, nous avons récupéré la séquence de 10 nucléotides en amont des amorces V sens.

Sur la totalité des reads analysés, moins de 1% des reads n'ont pas d'UMI de taille correcte, c'est-à-dire à une taille inférieure à 10 nucléotides.

En revanche, nous avons entre 85 et 90% d'UMI unique par échantillon. En clair, cela signifie que nous avons plus de séquences d'UMI unique que de cellules théoriquement analysables. Ce résultat nous a permis de mettre en évidence un problème de design dans la PCR-1.

Pour rappel, l'UMI est inséré en amont de la séquence VH spécifique dans l'amorce de la PCR-1. Prenons un exemple, la séquence 1 est amplifiée par l'amorce VH1 + UMI-1 au cycle 1 de PCR. Au cycle 2, il est hautement probable que ce fragment soit amplifié par une autre amorce VH1 + UMI-2 et quand même amplifier la même séquence que précédemment. Ce n'est donc pas une amplification provenant d'une autre cellule (*cf.* Figure 29). Pour utiliser au mieux les UMIs, il faudrait une étape de ligation des fragments d'ADN aux UMIs en amont de l'amplification. Nous n'avons donc pas pu utiliser l'analyse des UMI pour ce travail.

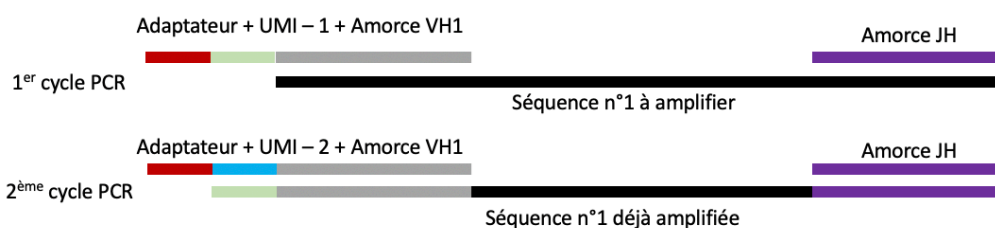


Figure 29. **Problème d'amplification avec les UMIs**

VI.4.5. Évaluation du taux de séquences productives analysées

Après annotation des séquences par la plateforme en ligne IMGT/HighV-QUEST, nous avons pu voir qu'en moyenne, pour chacun des échantillons, nous avons 80% de séquences amplifiées qui sont productives. Ces proportions sont en accord avec les données de la littérature (7) et d'autres équipes ont pu obtenir des résultats similaires (73,74). De plus, ceci est un bon indicateur de la qualité des séquences obtenues.

VI.4.6. Pool Témoins

Le pool témoins est composé de 12 ADN différents de lymphocytes B triés. Les échantillons utilisés sont ceux du groupe « témoins » (cf. V.1). Nous avons pu établir le profil des différents gènes V retrouvés pour les différents échantillons séquencés. Notre pool de 12 témoins sains nous a permis d'obtenir un répertoire dit « normal ». Nous l'avons comparé à différentes publications afin de vérifier l'absence de biais technique (73–75). Nous l'illustrons dans la Figure 30 réalisée à l'aide des données de *Boyd et al.* (75).

On retrouve une prédominance des gènes VH3-23, VH3-30, VH4-59 dans les données de la littérature et dans notre pool témoins. En revanche, les familles VH1-2 et VH1-18 sont prédominantes contrairement aux données de *Boyd et al.*, à l'inverse, VH1-69 est moins présent dans notre pool témoins. Malgré quelques différentes de répartitions pour 3 familles, nous avons validé que notre technique n'introduisait de biais lors de l'amplification des différents réarrangements V(D)J.

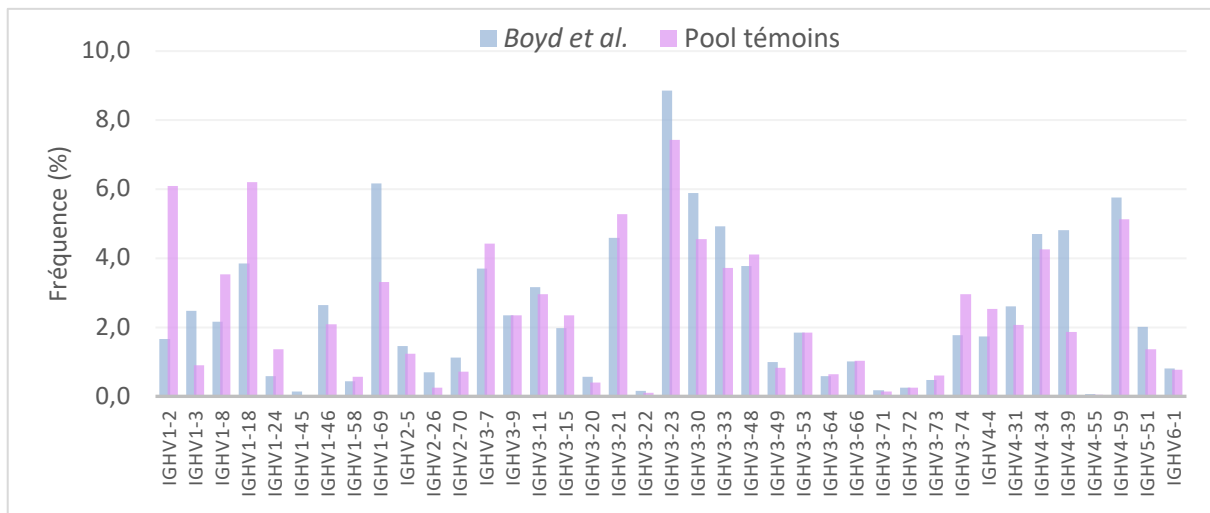


Figure 30. **Comparaison du répertoire des gènes VH**

Fréquences des gènes VH du pool témoins comparées à un répertoire obtenu par Boyd et al. 2010 chez 8 sujets sains (75)

VI.4.7. Pool LLC

Un mélange de 14 ADN de LLC a été constitué. Dans ce pool, nous avons choisi des séquences V(D)J différentes préalablement identifiées par séquençage Sanger (cf. Annexe 3). Le choix des différents ADN a été fait pour nous permettre de valider l'amplification des gènes V de différentes familles. En effet, les PCR multiplex peuvent conduire à des défauts d'amplification de certaines cibles.

En moyenne chaque échantillon de LLC contenait 70% de cellules clonales. Les ADN ont été extraits après réalisation d'un ficoll. Pour avoir un pool homogène, nous avons introduit 250ng de chaque clone LLC. Nous avons obtenu un pool contenant 4 736ng d'ADN dont 70% de cellules clonales de 14 LLC différentes. L'abondance attendue de chaque séquence clonale est de 5% dans ce pool (cf. Figure 31).

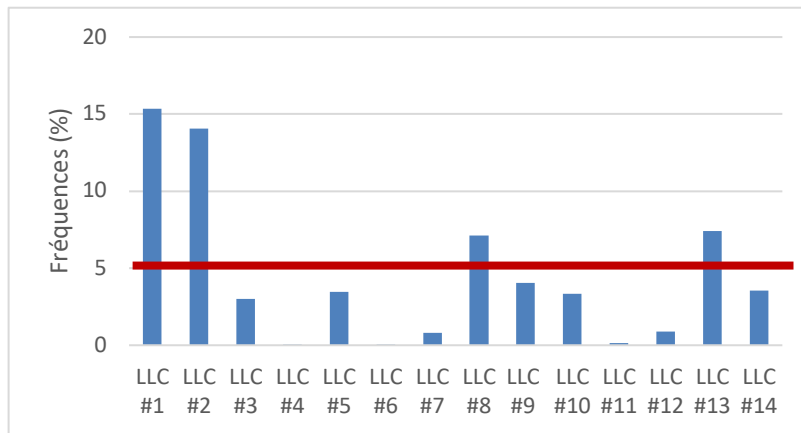


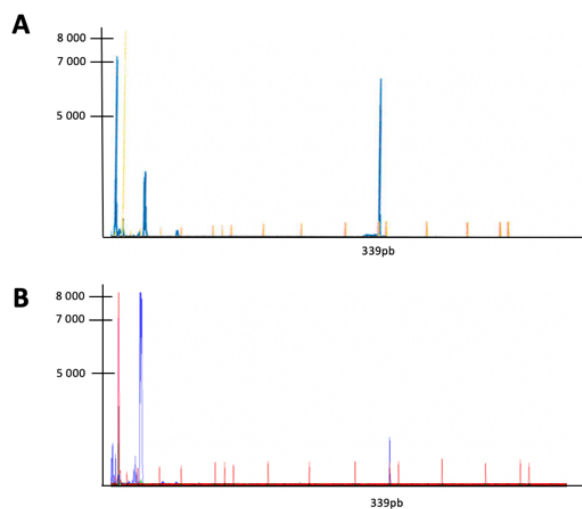
Figure 31. Répartition des fréquences de LLC retrouvées dans le pool

Histogramme représentant les fréquences de chaque LLC du pool. La ligne rouge est la fréquence théorique, c'est-à-dire 5%.

Devant l'absence des séquences des échantillons #4, #6 et #11, nous avons vérifié la clonalité de ces ADN par analyse de fragment, ainsi qu'une PCR permettant d'évaluer la qualité de l'ADN. Pour ces 3 échantillons, le signal de la PCR de clonalité était plus faible qu'au diagnostic (cf. Figure 32).

En revanche, concernant la qualité des ADN, nous n'avons pas pu mettre en évidence de différence entre les échantillons testés.

Grâce à cette analyse de fragments, nous avons pu mettre en évidence une absence d'amplification liée à l'échantillon et non à la technique utilisée. Ces échantillons seront exclus de l'analyse.



VI.4.8. Confrontation avec le séquençage Sanger

Pour les 11 LLC analysables en NGS, nous avons retrouvé la même séquence clonale que celle identifiée avec le séquençage Sanger.

VI.4.9. Gamme de dilution

Nous avons réalisé une gamme de dilution de notre de pool LLC (14 ADN clonaux) dans le pool témoins dans le but d'établir le seuil de sensibilité de la technique. Nous avons choisi 5 dilutions : 1/2, 1/5, 1/50, 1/100 et 1/500. Les échantillons « pur », « 1/2 », « 1/5 » ont été séquencés avec le kit V2 et les échantillons « 1/50 », « 1/100 », « 1/500 » avec le kit V3.

Les abondances des séquences clonales vont de 5.10^{-2} à 1.10^{-4} (cf. Tableau 9). Chaque échantillon de la gamme de dilution a été amplifié à partir de 500ng d'ADN.

Gamme de dilution	Abondance attendue / clone
Pool de 14 LLC	5.10^{-2}
Dilution 1/2	$2,5.10^{-2}$
Dilution 1/5	1.10^{-2}
Dilution 1/50	1.10^{-3}
Dilution 1/100	5.10^{-4}
Dilution 1/500	1.10^{-4}

Tableau 9. **Abondance des clones attendue dans la gamme de dilution**

Avec 500ng d'ADN de lymphocytes B, on étudie 76 923 cellules B. En ayant une profondeur de séquençage moyenne de 200 000 reads (V2) et 900 000 reads (V3) ainsi, sans biais d'amplification on peut en théorie obtenir 2 à 3 reads par cellules (V2) et 11 à 12 reads par cellules (V3). Ce « seuil » théorique est à nuancer par les fluctuations d'échantillonnage.

Nous avons réalisé une modélisation par application de la loi Binomiale pour évaluer la probabilité de ne pas détecter un clone entre 10^{-3} et 10^{-5} . D'après la Figure 33, nous avons pu voir que pour un clone présent à un taux de 10^{-4} avec les profondeurs de séquençage obtenue (750 000 reads en moyenne après *merged*) la probabilité de ne pas le détecter était d'une chance sur mille. Le facteur limitant pour détecter un clone à une fréquence inférieure à 10^{-5} est la quantité d'ADN initiale, il est donc nécessaire d'augmenter la quantité de cellules amplifiées. Dans la littérature, les techniques de séquençage haut débit réalisant ce type de quantification utilise une quantité d'ADN total de l'ordre du µg au minimum (76–78).

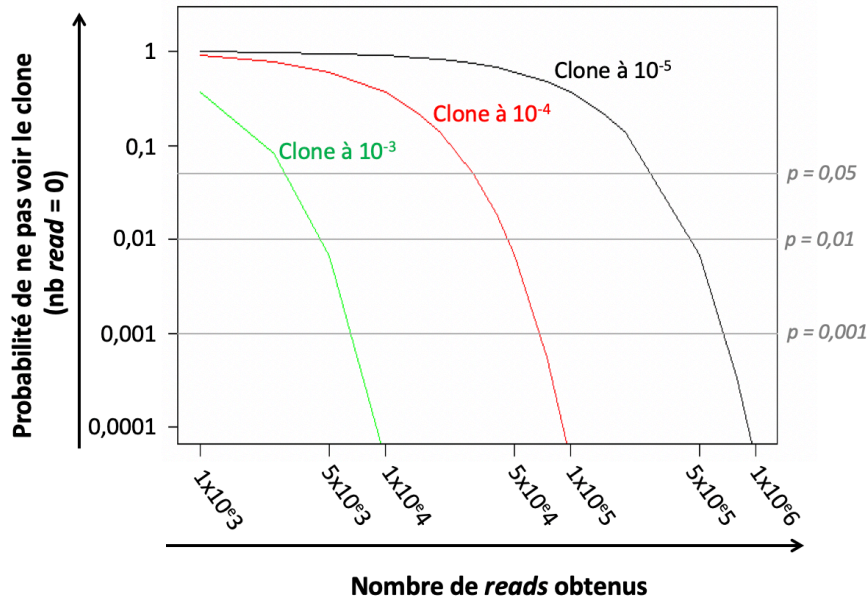


Figure 33. Estimation de la probabilité de ne pas détecter un clone entre 10^{-3} et 10^{-5}

Nous avons analysé chaque ADN de manière indépendante (cf. Figure 34). Nous avons pu voir que les fréquences de chaque clone diminuaient de manière linéaire. On peut retrouver une petite particularité pour les ADN #7 et #12. Pour ces échantillons, aux faibles dilutions, les fréquences sont inférieures à celles attendues. Il y a vraisemblablement un phénomène de compétition lors de la PCR qui empêche la bonne amplification de ces clones. En revanche, à partir de la dilution au 1/50^{ème}, on s'affranchit de cette compétition car l'apport du répertoire normal par le pool témoins favorise la diversité et les clones peuvent être amplifiés correctement.

A la dilution 1/500^{ème}, les occurrences pour chaque LLC varient entre 14 et 134 reads, ce qui est significatif. Nous n'avons pas atteint la limite de linéarité à cette dilution.

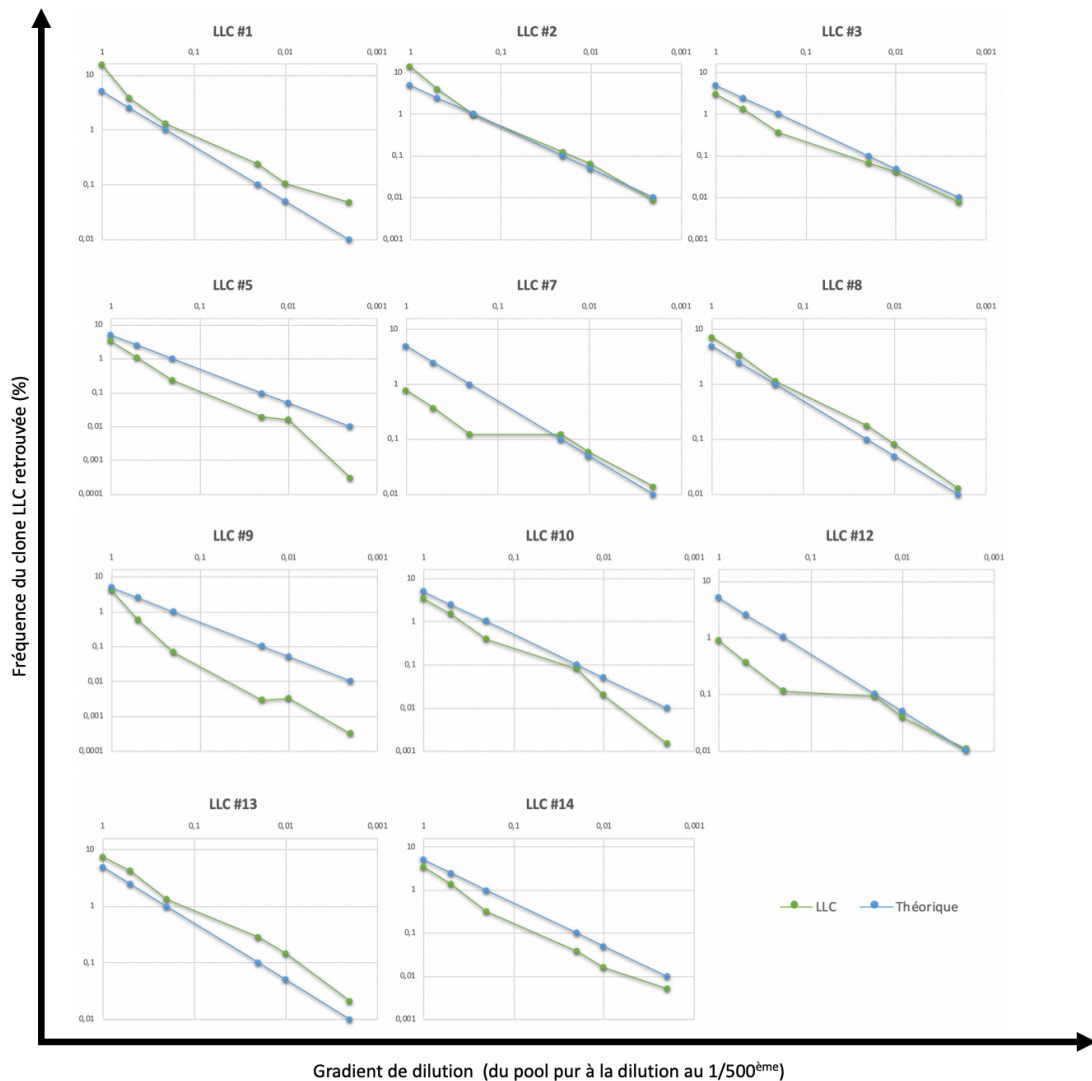


Figure 34. Gamme de dilution de chaque clone LLC du pool

En bleu est représenté la fréquence théorique attendue et en vert la fréquence du clone obtenue. Nous observons une linéarité très proche des valeurs théoriques particulièrement après la dilution au 1/50^{ème} (0,02).

VI.4.10. Analyse de l'échantillon LLC#1

Une librairie a été réalisée à partir de 500ng d'ADN de l'échantillon LLC#1 possédant une séquence stéréotypée du *subset #2*. Le nombre de *reads* obtenu est de 200 766 *reads*. Nous avons pu mettre en évidence que 92% des *reads* obtenus étaient la séquence clonale identifiée au préalable par séquençage Sanger.

Grâce à l'analyse bio-informatique (R) mise en place, nous avons pu identifier cette séquence clonale comme séquence stéréotypée du *subset #2*.

Nous avons comparé notre résultat avec l'outil ARRest/AssignSubsets qui nous a confirmé l'identification du *subset #2* pour cette séquence clonale (cf. Figure 35).

ARResT/AssignSubsets

assigning new members to existing subsets of stereotyped antigen receptor sequences

we're running ARResT/AssignSubsets - please follow our progress below...

(?) monitoring the resources used (your quota: 300 sec and 1000 megabytes RAM)
 (?) checking IMGT accessibility
 (?) running ARResT/SeqCure with your sequences...
 (!) more sequences than we allow (50), skipping and moving on
 (=) [ARResT/SeqCure report](#)
 (?) model is running...

(=) 50 / 50 / 50 were assigned / 'healthy' / submitted

DISCLAIMER - there is no guarantee that ARResT/AssignSubsets will be able to properly assign all your sequences to subsets, please bear this in mind when making decisions, especially important ones on e.g. clinical care, and especially with 'borderline'- or 'low'-confidence assignments. To help us improve ARResT/AssignSubsets, please [contact us](#).

[plain-text-formatted results table](#) (best viewable in a spreadsheet), or see below [click to open/close quick help](#)

assignment frequencies table

CLL#2	CLL#1	CLL#4	CLL#6	CLL#5	CLL#3	CLL#8	CLL#31	CLL#16	CLL#77
2.8%	2.4%	1.0%	0.9%	0.7%	0.6%	0.5%	0.4%	0.3%	0.3%

50

CLL#7H	CLL#28A	CLL#201	CLL#12	CLL#59	CLL#14	CLL#64B	CLL#99	CLL#202
0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%

assignment report table

label [+ heat map, if appl.]	SeqCure	subset	confidence	score
M06560:5:000000000-D6W8B:1:1101:11982:1900	warning	CLL#2	extreme	82.07

Figure 35. Assignment des subsets par ARResT/AssignSubsets

ARResT/AssignSubsets assigne directement le subset trouvé à la séquence. Il donne un commentaire sur la confiance qu'il a en son assignation, ici « extreme » qui est corrélée au score calculé : 82.07.

Au total, grâce à l'analyse de ces différents échantillons, nous avons pu valider la qualité des bibliothèques réalisées et le séquençage. Notre pool témoins nous a permis d'obtenir le profil du répertoire « normal » attendu dans une population *à priori* « saine ». La linéarité de la technique testée avec la dilution d'un pool de 14 LLC dans un pool témoins n'est pas atteinte à 1.10^{-4} . Enfin, notre analyse bio-informatique est validée tant pour rechercher des séquences clonales connues que les séquences stéréotypées.

VI.5. Application à la recherche de séquences stéréotypées type « LLC »

Dans un second temps, nous avons séquencés 19 patients sains à la recherche de BCRs stéréotypés. Nous avons vérifié les différentes métriques pour valider la qualité du séquençage obtenu.

VI.5.1. Analyse des métriques

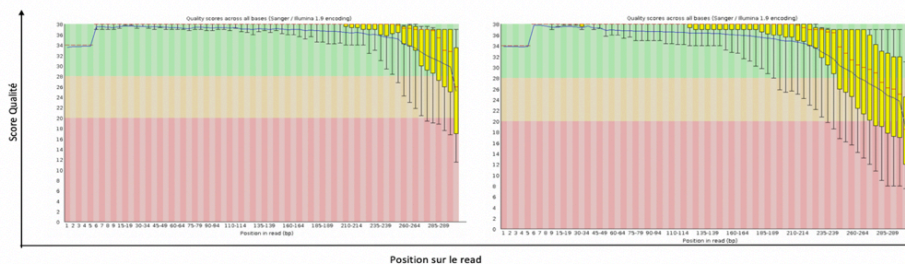
Le nombre de *reads* obtenus par échantillon est en moyenne de 933 000 pour une moyenne théorique à 1 000 000 (cf. Figure 36A).

L'analyse de la qualité de lecture des bases obtenues sur le séquençage avec le kit V3 nous montre une analyse de très bonne qualité sur les *reads* 1. Seulement les 50 dernières bases lues ont un score entre Q20 et Q30. En revanche, pour le *read* 2, la qualité diminue fortement. Sur les 70 dernières bases lues, le score diminue jusqu'à être inférieur à Q20. Pour chaque échantillon, l'analyse de la qualité des bases lues (*read* 1 et *read* 2) est homogène. Le *read* 1 a une excellente qualité et le *read* 2 a toujours une dernière partie de moins bonne qualité (cf. Figure 36B).

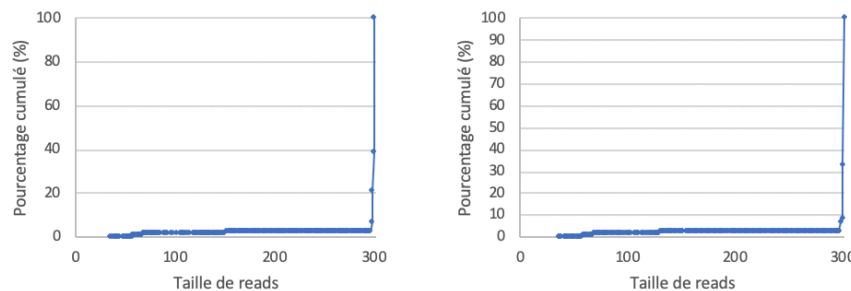
Concernant la taille des *reads*, comme pour le séquençage avec le kit V2, moins de 10% des *reads* ont une taille inférieure à 300pb (cf. Figure 36C). Lors de leur analyse, nous avons pu observer que ces séquences étaient des dimères d'amorces.

	MOYENNE	MIN – MAX
NOMBRE DE READS TOTAUX PAIRED-END PAR ÉCHANTILLON (X 1000 READS)	933	604 – 1 142
NOMBRE DE READS OBTENUS APRÈS MERGED (X 1000 READS)	724	476 – 892
PROPORTION DE READS MERGED (%)	78	73 – 80
NOMBRE DE READS ÉLIMINÉS PAR LE MERGED CAR > 2 MISMATCHS (X 1000 READS)	195	120 – 231

A. Tableau résumant le nombre de *reads* obtenus en moyenne ainsi que les rendements après assemblage des séquences



B. Qualité des *reads* obtenus avec FASTQC report, exemple d'un échantillon (*read* 1 à gauche et *read* 2 à droite)



C. Analyse de la répartition de la taille des *reads* pour un même échantillon (*read* 1 à gauche et *read* 2 à droite). En ordonné, est représenté le pourcentage cumulé de *reads* selon la taille. Moins de 10% des *reads* ont une taille inférieure à 300.

Figure 36. Analyse des métriques du séquençage avec le kit V3

VI.5.2. Profil de clonalité

Pour chaque échantillon recueilli, un dépistage de clone B a été réalisé par CMF. Pour chaque échantillon, une médiane de 2 500 cellules B ont été analysées [1 300 – 8 400 cellules B]. Nous avons confirmé l'absence de clone par CMF pour tous les échantillons.

A l'aide de l'analyse des séquences obtenus, nous avons pu établir un « profil de clonalité » pour chaque échantillon. Ce profil est proche de celui généré par l'analyse de fragment selon le protocole BIOMED-2 utilisé en routine (71). Pour cela, nous avons regroupé chaque séquence en fonction de sa taille. Nous en avons établi une représentation graphique qui mime les résultats obtenus par analyse de fragment (*cf.* Figure 37). Tous les échantillons testés ont un profil régulier, sauf le patient #11 où nous avons observé un profil irrégulier.

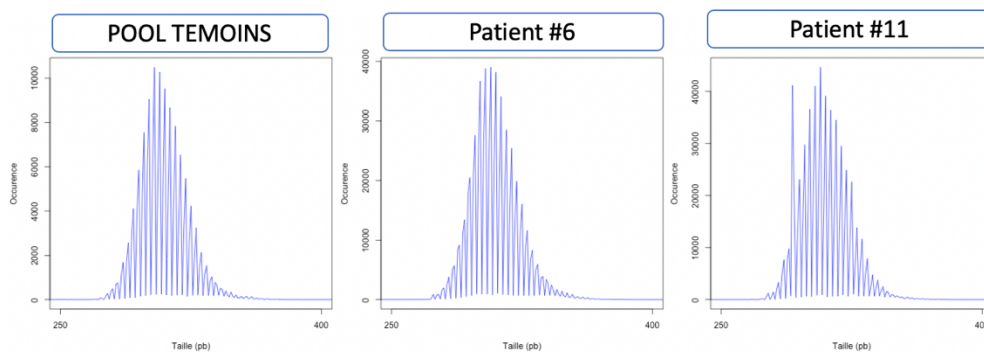


Figure 37. Profil type clonalité des échantillons séquencés

A gauche, le pool de 12 ADN témoins, au centre un patient représentatif des 19 patients séquencés. Tous deux montrent un profil régulier de la répartition de taille des séquences. A gauche, le patient #11 montre un profil irrégulier avec une séquence plus courte prédominante. Cette irrégularité n'avait pas été observé par CMF.

VI.5.3. Recherche des séquences stéréotypées

En moyenne, après annotation des séquences par IMGT HighV-Quest, 82% des reads obtenus ont une séquence avec un réarrangement fonctionnel.

Nous avons appliqué notre stratégie d'identification des *subsets* sur les 19 patients de la cohorte (*cf.* Figure 23). Avec cette technique, 4 *subsets* n'ont pas été retrouvés chez nos patients, le *subset* #4, #6, #16 et #201.

Pour ce qui est des séquences identifiées comme *subset* avec notre stratégie, nous les avons soumises en ligne sur ARRest/Assign subsets pour confirmation.

Pour 5 *subsets* en particulier, le #8, #12, #31, #77 et #202, nous n'avons confirmé que très peu des séquences avec ARRest/Assign subsets (*cf.* Figure 38). Ceci a pu nous permettre d'observer que les motifs CDR3 de certains *subsets* notamment les *subsets* #31 et #202 ne sont pas assez spécifiques pour pouvoir les identifier au mieux avec notre méthode (recherche d'identité de séquence).



Figure 38. **Subsets non confirmés avec ARRest/Assign subsets**

En vert, le nombre de reads identifiés avec R et en bleu confirmés par ARRest/Assign subsets

Les subsets #1, #7C, #14, #28A, #64B et #99 ont été identifiés avec notre stratégie d'analyse et confirmés sur ARRest/Assign subsets dans 34 à 100% des cas. Ces subsets sont retrouvés chez 5 à 37% des patients de la cohorte (cf. Figure 39).



Figure 39. **Subsets minoritaires dans la cohorte**

En vert, le nombre de reads identifiés avec R et en bleu confirmés par ARRest/Assign subsets

Enfin, les *subsets* #2, #3, #5 et #59 ont été retrouvés chez 63 à 95% des patients de la cohorte. Les séquences ont été confirmées à 100%, 100%, 45% et 86% par ARRest/Assign subsets respectivement (cf. Figure 40). Ces quatre *subsets* semblent être ubiquitaires au sein de cette cohorte.

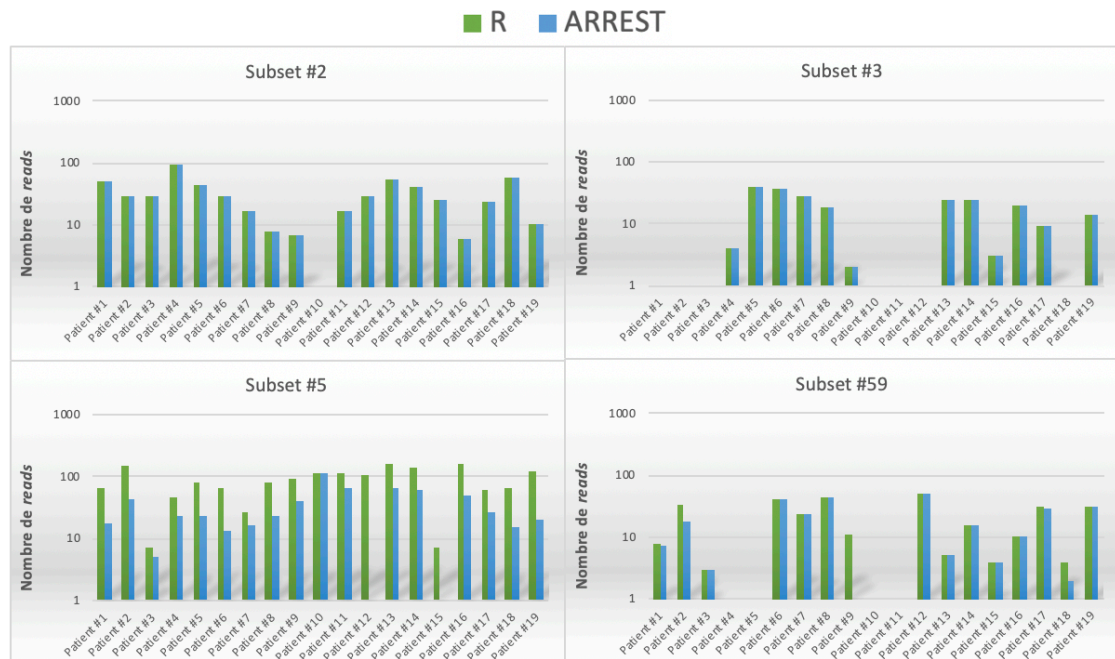


Figure 40. **Subsets majoritaires de la cohorte**

En vert, le nombre de reads identifiés avec R et en bleu confirmés par ARRest/Assign subsets

Pour la suite de l'analyse des résultats, nous avons choisi de nous concentrer sur le *subset* #2 de par sa fréquence et sa valeur pronostique très particulière dans la LLC.

VI.5.3.1. Existence de CDR3 de type *subset* #2 sans IGHV3-21

Lors de la recherche des séquences de CDR3 ayant un motif particulier de *subset*, nous avons pu observer que ces CDR3 « particuliers » n'étaient pas spécifique d'un gène V et/ou de gènes J. Ainsi pour le *subset* #2, de nombreux gènes V sont associés à sa séquence CDR3 « particulière » et non pas uniquement le gène V3-21 qui le caractérise (cf. Figure 41). Il apparait donc que la génération de CDR3 très court semble fréquente.

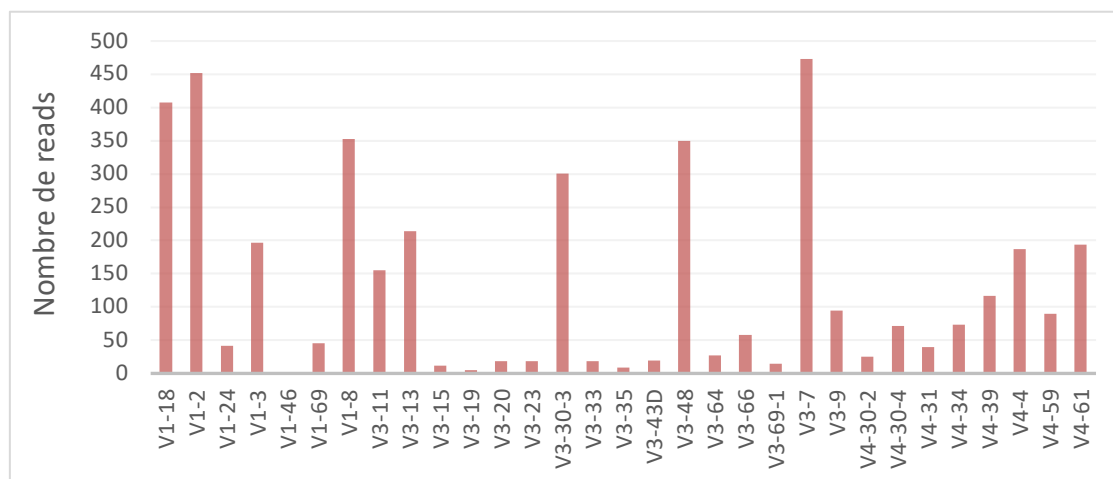


Figure 41. Répartition des familles de gènes V non V3-21 ayant un motif *subset #2*

VI.5.3.2. Diversité intra-clonale au sein du *subset #2*

Pour chaque séquence *subset #2* identifiée par notre stratégie et par ARRest/Assign subsets, nous avons regardé le nombre de CDR3 unique par patient (*cf.* Figure 42A).

Leur abondance relative est représentée sur la Figure 42B. Ceci nous a permis d'observer que la diversité intra-clonale au sein de ce *subset* était retrouvée pour chaque patient. Par exemple pour le patient #4, nous retrouvons 97 *reads* correspondant à 9 CDR3 uniques alors que pour le patient #14, les 41 *reads* identifiés sont repartis en 19 CDR3 uniques (*cf.* Figure 42A). Nous avons ensuite souhaité quantifier les différents sous clones pour chacun des patients. Certains clones apparaîtraient plus représentés (> 5 *reads*) que d'autres (*cf.* Figure 42B).

Cependant, compte tenu de l'absence d'UMI utilisable, nous ne pouvons pas éliminer à ce niveau qu'il s'agisse simplement de biais de PCR.

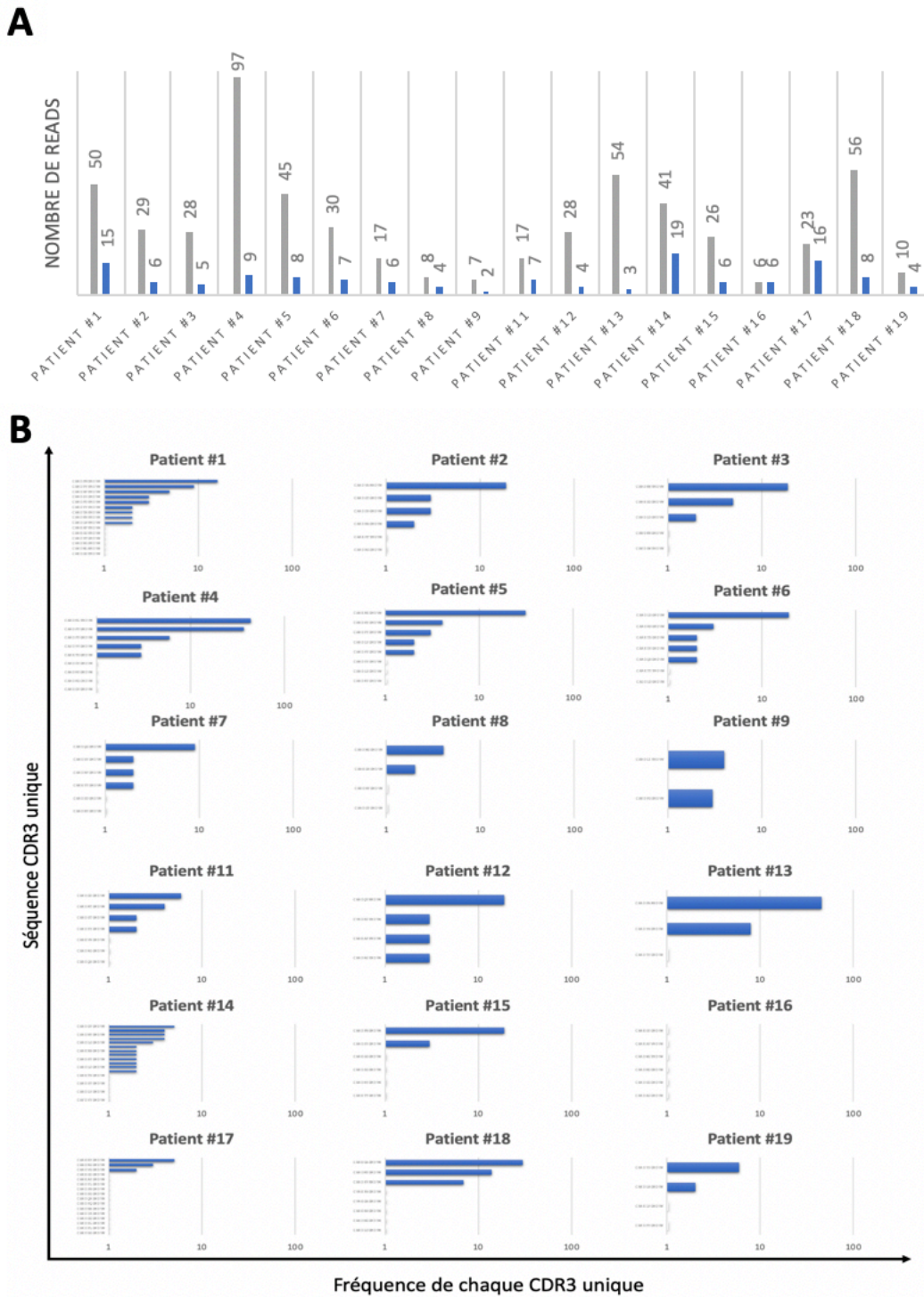


Figure 42. Répartition du nombre de reads de séquences stéréotypées pour le subset #2 et la fréquence des CDR3 uniques

En gris, le nombre de reads obtenus pour chaque échantillon. En bleu, le nombre de CDR3 unique.

VI.5.3.3. Diversité des séquences ayant la même séquence CDR3

Concernant les séquences ayant le même CDR3, nous avons voulu savoir s'il s'agissait de plusieurs clones distincts ou simplement de diversité intra-clonale. Pour simplifier, nous prendrons l'exemple du Patient #1. Sur la Figure 43, nous avons représenté un *cluster* de 16 séquences obtenues pour ce patient qui avaient un CDR3 identique.

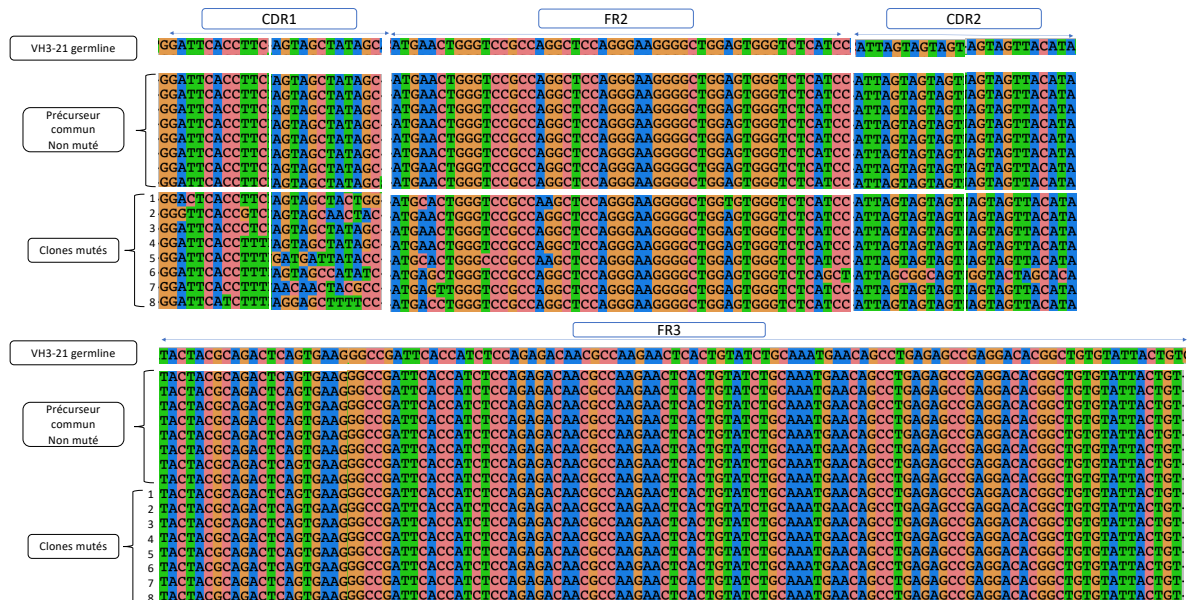


Figure 43. Séquences nucléotidiques des 16 séquences du Patient #1 ayant un CDR3 identique

Parmi elles, 8 séquences sont identiques à la séquence germinale VH3-21 et 8 autres présentent différentes mutations. Nous avons réalisé une analyse phylogénétique de ces différentes séquences (*cf.* Figure 44). On a identifié 2 « groupes » de clones pouvant dériver du clone majoritaire commun (VH3-21 non muté).

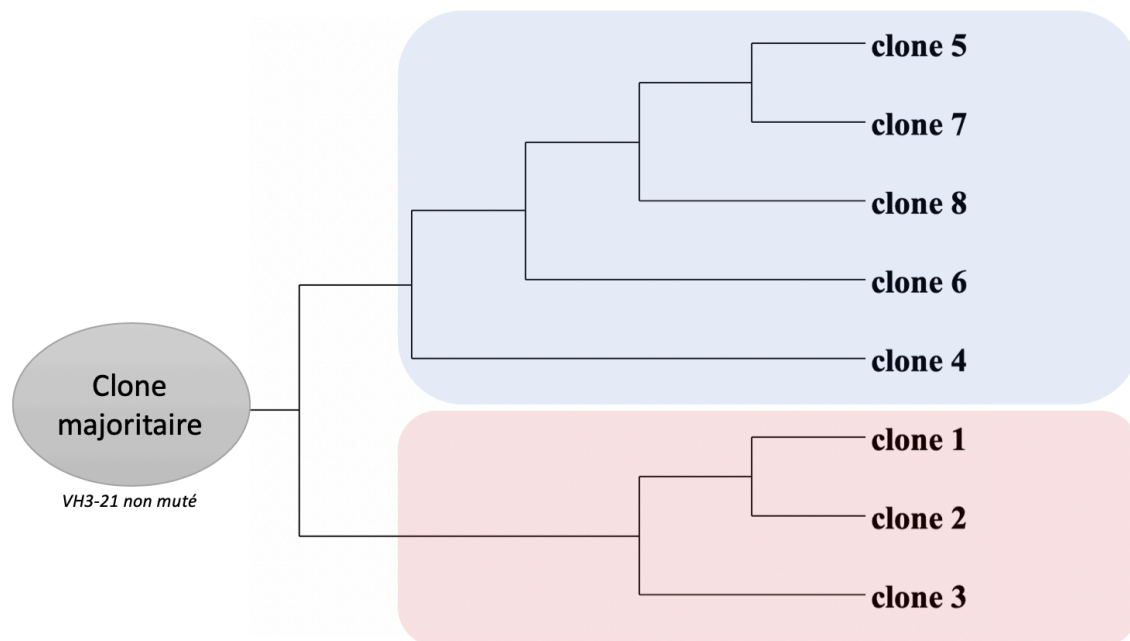


Figure 44. **Phylogénie moléculaire des clones ayant un CDR3 identique de subset #2**

Cladogramme représentant les distances phylogénétiques des 8 séquences du cluster des 16 séquences du patient #1 ayant le même CDR3.

VI.5.3.4. Regard critique sur l'analyse des données

A la suite de cette première analyse, nous avons voulu éprouver la spécificité des motifs recherchés. Nous avons donc modifié des positions *a priori* conservées dans le motif CDR3 et ensuite avons recherché ce nouveau motif au sein des séquences obtenues.

En clair, pour le motif du *subset #2*, au niveau de la position 3 l'acide aminé est annoncé dans la littérature comme étant un D (*cf.* Figure 45). Nous avons donc recherché toutes les séquences n'ayant pas de D comme acide aminé à la position 3. Les critères d'identification du gène V, du gène J ainsi que la taille de la séquence CDR3 restant inchangés.

Nous avons pu trouver des séquences chez tous les patients correspondant à ces nouveaux critères. Elles ont été soumises à ARRest/Assign subsets qui a assigné ces séquences comme étant des séquences stéréotypées du *subset #2*.

De la même façon, nous avons modifié l'acide aminé de la 8^{ème} position en regardant les séquences qui n'avaient pas de D. Nous avons trouvé une séquence chez le Patient #12 qui n'est pas identifié comme *subset #2* avec ARRest/Assign subsets.

La position 3 fait partie de la région N nucléotides ajoutée lors de la recombinaison des immunoglobulines (*cf.* Figure 45). Au contraire, la position 8 fait partie du gène J.

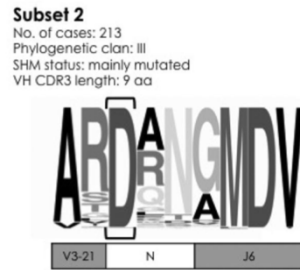


Figure 45. **Motif de la séquence CDR3 du subset #2**

[d'après Agathangelidis et al. 2012 (47)]

Cette analyse nous a permis d'avoir un regard critique sur notre stratégie de recherche des séquences stéréotypées. Les acides aminés présents au niveau des régions N et P nucléotides ont une probabilité de diversité plus importante que ceux présents au niveau des gènes V, D et J. Dans les motifs de CDR3 utilisés pour notre analyse, certains acides aminés dans ces régions ont été imposés. Cette constatation nous apporte la preuve d'un manque de sensibilité de notre technique d'analyse des données. Il serait donc nécessaire de refaire cette analyse en ne figeant pas d'acides aminés dans les régions N et P nucléotides, mais seulement dans les régions des gènes V, D et J.

VII. Discussion

VII.1. Mise en place du séquençage haut débit des gènes des immunoglobulines

VII.1.1. Aspects qualitatifs de la méthode

Le séquençage IGHV à partir d'une amplification au niveau du locus FR1 fonctionne avec la stratégie mise en place dans ce travail (dérivation du protocole BIOMED-2 (71)). La technique semble être stable car elle a pu être réalisée sur deux séquençages indépendants avec des échantillons différents. En effet, sur ces deux expérimentations, nous avons atteints les rendements théoriques établis par le fournisseur en termes de nombre de *reads* notamment. Les qualités des séquences obtenues étaient très satisfaisantes au regard des scores de qualité et du pourcentage de séquences productives. En revanche, nous avons invalidé la stratégie de barcodage moléculaire utilisée (UMI). En effet, pour utiliser les UMIs de manière optimale, il est nécessaire de les ajouter aux séquences d'intérêts par ligation et non par amplification, comme proposé par divers kits commerciaux (79).

L'amplification étant réalisée à partir du locus FR1, nous avons un manque de sensibilité au niveau du pourcentage d'homologie par rapport à la région V amplifiée. Le groupe ERIC (« *European Research Initiative on Chronic lymphocytic leukemia* ») recommande, pour déterminer le statut mutationnel de façon optimale, la réalisation d'une amplification des gènes des immunoglobulines à partir de la région Leader même si l'amplification au niveau du locus FR1 est acceptable (80). En effet, l'obtention d'une séquence IGHV complète est nécessaire pour obtenir avec précision le statut mutationnel du clone pour des valeurs proches de 2%. Comme nous l'avons décrit, la décision thérapeutique en dépend (chimiothérapie vs thérapie ciblée). Ainsi, dans l'hypothèse d'une implémentation en routine pour le diagnostic des LLC, nous devons poursuivre les mises au point d'amplification à partir de la région Leader.

VII.1.2. Aspects quantitatifs de la méthode

En réalisant une gamme de dilution d'un pool de 14 clones de LLC différents, nous avons détecté leurs abondances relatives de $5 \cdot 10^{-2}$ à $1 \cdot 10^{-4}$. Un phénomène de compétition entre les différents clones a été observé lorsqu'ils sont chacun présents en quantité trop importante ($5 \cdot 10^{-2}$). Ce phénomène est atténué voir annulé lorsque les différents clones sont présents à des concentrations plus faibles. Ainsi, avec cette technique et les profondeurs de séquençage obtenues, c'est-à-dire une moyenne de 900 000 *reads* par échantillon, nous avons pu obtenir une quantification linéaire des clones testés jusqu'à une abondance de $1 \cdot 10^{-4}$. Notre méthode semble donc suffisamment sensible pour le suivi de maladie résiduelle de la LLC et/ou d'autres lymphomes B de bas grade dont l'architecture clonale est relativement simple (au moins du point de vue des immunoglobulines).

L'étude de la maladie résiduelle dans la LLC a commencé depuis quelques années afin d'évaluer la réponse aux traitements des patients (81). Elle prend son essor depuis l'introduction des nouvelles thérapeutiques ciblées. Elle est utilisée dans les protocoles cliniques comme étant un marqueur prédictif de la réponse aux traitements et de la survie sans progression des patients (82–84). Depuis le début de la mise en place de l'étude de la MRD dans la LLC, la cytométrie en flux a été la technique de choix, développée au départ avec un panel de 4 couleurs, puis 6 et enfin standardisée entre 8 et 10 couleurs actuellement. Ceci a été possible car les cellules clonales de LLC ont un phénotype particulier qui leur permet d'être

facilement isolables des lymphocytes normaux (82). Au niveau moléculaire, la mise en place peine à se démocratiser contrairement à la cytométrie en flux, mais tend à évoluer, surtout depuis la mise en place du séquençage haut débit (82,84,85). Le seuil consensuel de détection des cellules clonales utilisé est à 1.10^{-4} d'après les recommandations du groupe ERIC (80). Dans tous les protocoles cliniques, la détection des cellules clonales a été fixée au seuil de 1.10^{-4} . Ce seuil établi a permis de pouvoir comparer facilement les différentes études entre elles pour pouvoir évaluer de manière optimale la réponse aux différents traitements (63). En effet, de nouveaux traitements ciblés émergent comme l'ibrutinib, l'idélalisib et le vénétoclax entre autres. Afin d'établir les stratégies de traitements les plus adaptées aux différents patients, il est nécessaire d'avoir un outil fiable, robuste et comparable entre les différentes études pour évaluer la réponse à ces traitements. La MRD dans les LLC n'a pas pour objectif à l'heure actuelle d'être utilisée pour la totalité des patients mais reste encore réservée aux protocoles cliniques afin d'évaluer son impact dans les stratégies thérapeutiques (86). Après son rôle probable dans une détection précoce des patients non répondeurs aux différents traitements, elle pourrait aussi être utilisée dans le but d'envisager l'arrêt de traitement administré au long cours (82,86).

La suite de notre travail va être de s'assurer que notre technique pourrait être utilisée et répondre de manière fiable à la question de la MRD dans la LLC. Pour cela, il est nécessaire et prévu de réaliser une gamme de dilution à partir d'un échantillon de cellules de LLC triées. Le but étant d'obtenir un échantillon pur de cellules clonales que l'on pourra diluer dans des échantillons d'ADN polyclonaux. Nous aurons ainsi une approximation de la quantité de cellules clonales optimale. La gamme de dilution va être réalisée jusqu'à obtenir l'indélectabilité du clone dans le but d'obtenir le seuil de sensibilité de la technique. Ces expérimentations devront être réalisées plusieurs fois pour tester en parallèle la reproductibilité de la technique.

Ces résultats ainsi validés nous permettrons de proposer un outil d'évaluation de la maladie résiduelle chez les patients LLC. De plus, cette technique peut être applicable à toutes les lymphopathies B ayant une séquence V(D)J identifiée au diagnostic. En effet, l'avantage de cette technique est qu'elle peut être réalisée sur ADN comme sur ARN. Cependant, avec l'utilisation d'ARN, les aspects quantitatifs sont plus difficiles à évaluer *stricto sensu*.

VII.2. Recherche de séquences stéréotypées dans une cohorte de 19 patients exempts d'hémopathie

Dans ce travail, nous avons pu voir que certains *subsets* (#2, #3, #5 et #59 principalement) étaient présents chez des patients exempts d'hémopathie. Pour les identifier, nous avons utilisé une approche par recherche d'identité de séquences stéréotypées (motifs dans les séquences CDR3) décrits dans la littérature (47,48). Pour rappel, les critères indispensables à l'identification des séquences stéréotypées sont l'identification du gène V, du gène J, la taille de la séquence CDR3, 50% d'identité avec la séquence CDR3 et 70% de similarité pour cette séquence (*cf.* Figure 12A) (48). Tous ces critères ont été identifiés par ces auteurs grâce à l'étude d'une cohorte de plus de 20 000 patients LLC. Ils ont pu développer un outil d'assignation de ces séquences stéréotypées basée sur une statistique bayésienne qui permet de prédire l'appartenance de chaque séquence à un *subset* majeur (72).

Les séquences retrouvées comme appartenant à un *subset* par notre stratégie d'analyse ont été soumises à l'assignation avec cet outil pour confirmer notre identification. Pour certains

subsets, la confirmation par cet outil a été faite dans 100% des cas notamment le *subset #2* et le *subset #3*. Cette confirmation dans 100% des cas pour ces deux *subsets* nous montre que les critères utilisés dans notre analyse ont une bonne spécificité. En revanche, nous avons pu voir que celle-ci est variable selon les *subsets* recherchés. Pour certains *subsets*, notamment le #202, nous avons identifié de nombreux *reads* pour chacun des patients comme étant du *subset #202*. L'assignation de ces séquences n'a pas été confirmée par ARRest/Assign subsets. Pour ce *subset*, les critères que nous utilisons pour sa désignation ne sont pas assez discriminants pour l'identifier avec certitude. Ce phénomène a été reproduit avec le *subset #31*. En revanche, ce manque de spécificité nous a permis de voir qu'il existe chez les patients de la cohorte une production de séquences proches de ces *subsets*.

Secondairement, nous avons pu mettre en évidence un manque de sensibilité dans notre analyse. En effet, dans notre stratégie d'identification des séquences stéréotypées, nous avons limité le choix en acides aminés à certaines positions des séquences CDR3 présentes dans les régions N et P nucléotides. Cependant, celles-ci ont pour rôle d'apporter une diversité dans les réarrangements des gènes des immunoglobulines, leur séquence n'est donc pas définie contrairement aux séquences germinales des gènes V, D et J. Il serait donc nécessaire de revoir nos motifs de séquences CDR3 pour chacun des *subsets* recherchés en ne limitant pas le choix des acides aminés possibles au niveau des régions N et P nucléotides. De cette façon, nous espérons identifier plus de séquences stéréotypées, sans pour autant perdre en spécificité.

Malgré ce manque de sensibilité, nous avons pu identifier des séquences stéréotypées chez tous les patients de la cohorte étudiée. Ceci a pu être mis en évidence aussi par *Muggen et al.* très récemment (74). Cette équipe a réalisé le séquençage de 155 sujets sains selon quatre classes d'âge et en triant les différentes sous populations de lymphocytes B. Ils ont également mis en évidence la présence de séquences stéréotypées notamment du *subset #2* et du *subset #5* à une fréquence variant de 0,1 à 0,7%. De plus, ils ont pu voir que la proportion de séquences stéréotypées augmente avec l'âge après 50 ans. Tout comme cette équipe, nous avons retrouvés les mêmes *subsets* majoritaires dans notre cohorte, c'est-à-dire le *subset #2* et le #5 notamment. En revanche, les quantifications relatives que nous avons obtenues ne sont pas du même ordre notamment pour le *subset #2*. Nous obtenons des fréquences de l'ordre de 0,01% par patient pour les fréquences les plus élevées (maximum 100 *reads* pour 700 000 *reads* productifs obtenus en moyenne) en comparaison au 0,1% obtenus chez leurs 155 patients séquencés (pour les fréquences les plus faibles). Ceci peut être expliqué au moins en partie par le manque de sensibilité observé au cours de l'analyse bio-informatique. En effet, *Muggen et al.* ont utilisé seulement ARRest/Assign subsets pour l'identification de leurs séquences stéréotypées contrairement à nous.

Au total, outre les difficultés analytiques rencontrées, nous avons répondu à notre question principale qui était de savoir si nous pouvions trouver des séquences stéréotypées de type LLC dans une population exempte d'hémopathie. Ce qui est intéressant, c'est qu'il semble que seulement certains BCRs stéréotypés soit produits de manière ubiquitaire. La proportion de BCRs de type « LLC » dans le répertoire normal semble augmenter avec l'âge (74). Cependant, il a été montré que les différentes sous populations de lymphocytes B tendent à diminuer avec l'âge (87). Nous pouvons émettre ici deux hypothèses liées au vieillissement : d'un côté une augmentation de la production de BCRs stéréotypés ou au contraire une diminution de la production des séquences non stéréotypées (augmentation relative de leur fréquence) (cf. Figure 46).

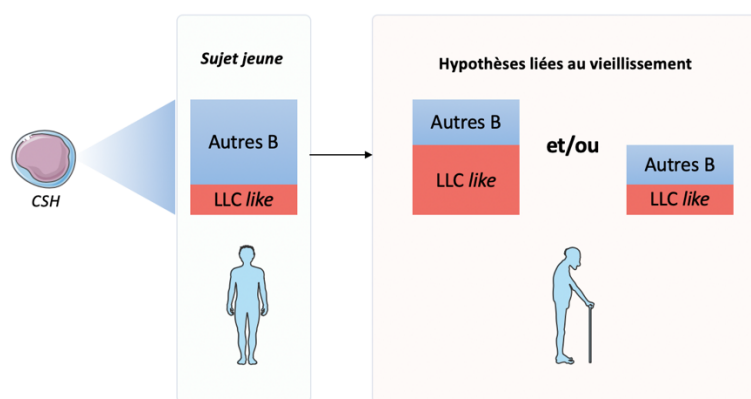


Figure 46. **Hypothèses de l'évolution du répertoire B liée au vieillissement**

La question est donc totalement ouverte de savoir si la présence et surtout l'augmentation de la production de ces séquences stéréotypées au cours du temps favoriserait l'émergence de clone LLC.

Muggen et al. ont aussi montré que le répertoire IGHV des lymphocytes B devenait de plus en plus restreint au cours du temps et se rapproche du répertoire des IGHV des LLC (74). Cependant, malgré la présence de ces séquences stéréotypées chez des sujets « sains » qui pour beaucoup ne développeront pas d'hémopathie de type LLC oriente plutôt vers une préexistence minoritaire de ces séquences de manière physiologique. Est-ce que cette « prédisposition » pourrait favoriser l'acquisition ou la sélection d'évènements secondaires comme certaines mutations de *SF3B1* par exemple, qui est impliqué dans la résistance à l'apoptose ?

Concernant les *subsets* « ubiquitaires » (tel que le #2 retrouvé chez 95% des témoins), l'acquisition/sélection d'évènements oncogéniques nécessaires à la transformation leucémique pourrait survenir dès le stade CSH comme démontré dans certaines études, soit plus tard au cours de la vie du LB. Dans le premier cas, ces mutations devraient être retrouvées dans les différents lignages hématopoïétiques et dans le second cas, elles devraient être restreintes au clone de LLC.

En revanche, pour les *subsets* non retrouvés (*subset* #4, #6 et #8) dans notre travail et dans l'étude de *Muggen et al.* il se pourrait qu'ils soient peu ou pas produits voir délétés au cours de la lymphopoïèse B (LB auto-réactif ?). On pourrait imaginer que la survie de ces lymphocytes nécessite un évènement oncogénique très tôt lors de la lymphopoïèse B (délétion 13q ?) leur permettant de résister à l'apoptose. Ces derniers seraient alors plus avancés dans le processus de lymphomagenèse. En effet, la délétion 13q est présente chez près de 50% des patients et presque toujours clonale (*i.e.* : présente dans 100% de la masse tumorale).

Pour y répondre complètement, nous pourrions rechercher la présence de mutations d'oncogènes dans le compartiment B des témoins porteurs de BCRs stéréotypés d'une part et dans les autres lignages d'autre part. Nous avons conservé les fractions non-B après tri cellulaire pour le faire.

Enfin, pour explorer l'architecture clonale de ces populations B et non-B, nous pourrions imaginer la réalisation d'un RNAseq *single cell* qui permettrait par exemple d'associer l'expression d'un BCR *subset* #2, la présence de mutations *SF3B1* et la mise en évidence d'anomalies de l'épissage des ARNs.

Conclusion et perspectives

Dans ce travail, nous avons pu développer un nouvel outil de séquençage des gènes des immunoglobulines. En effet, cette technique avait déjà été mise au point par notre équipe de recherche « Contrôle des réponses immunes B et des lymphoproliférations » (CRIBL) – UMR 7276/INSERM 1262 dirigée par le Pr Michel Cogné, à partir d'ARN (88). C'était donc naturel de vouloir poursuivre en adaptant cette technique à une application de diagnostic notamment pour la recherche des mutations somatiques des patients LLC. Malgré les difficultés techniques rencontrées, notamment dues à l'utilisation d'une PCR multiplex, nous avons réussi à valider une méthode efficace de préparation des librairies et un *pipeline* simple d'analyse pour identifier les séquences clonales et leur statut mutationnel.

De plus, cet outil permet d'étudier la maladie résiduelle dans cette pathologie au seuil consensuel utilisé (1.10^{-4}). Cet aspect du travail réalisé va ouvrir sur une collaboration avec le CHU de Poitiers pour étudier la réponse aux nouvelles thérapies ciblées dans la LLC et pouvoir comparer l'étude de la MRD par cytométrie en flux vs la MRD par séquençage haut débit.

Enfin, la question de la production de séquences stéréotypées favorisant la sélection clonale dans la LLC n'est pas clairement élucidée. Au-delà d'études moléculaires plus avancées évoquées, un suivi au long cours d'une cohorte de sujets sains pourrait être intéressant.

Références bibliographiques

1. Lefranc M, Lefranc G. Génétique moléculaire des immunoglobulines. *IMGT Int ImMunoGeneTics Inf Syst*. 2002;
2. Lefranc M-P. IMGT, the international immunogenetics database. *Nucleic Acids Res*. 2003;31(1):307-10.
3. Schroeder HW, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S41-52.
4. Batteux F, Garraud O, Prin L, Renaudineau Y, Vallat L. Lymphocytes B : diversité, ontogénèse, différenciation et activation [Internet]. Disponible sur: http://www.assim.refer.org/raisil/raisil/L02_files/page82-7.-lymphocytes-b.pdf
5. Nagasawa T. Microenvironmental niches in the bone marrow required for B-cell development. *Nat Rev Immunol*. 2006;6(2):107-16.
6. Dekoter R, Singh H. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science*. 2000;288(5470):1439-41.
7. Meffre E, Casellas R, Nussenzweig MC. Antibody regulation of B cell development. *Nat Immunol*. 2000;1(5):379-85.
8. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol*. 2006;24:541-70.
9. Nishana M, Raghavan SC. Role of recombination activating genes in the generation of antigen receptor diversity and beyond. *Immunol Suppl*. 2012;137(4):271-81.
10. Helmink BA, Sleckman BP. The response to and repair of RAG-mediated DNA double-strand breaks. *Annu Rev Immunol*. 2012;30:175-202.
11. Darlow JM, Stott DI. VH replacement in rearranged immunoglobulin genes. *Immunol Suppl*. 2005;114(2):155-65.
12. Lee G, Neiditch M, Salus S, Roth D. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell*. 2004;117(2):171-84.
13. Lieber M. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*. 2010;79(1):181-211.
14. Raess P, Bagg A. The role of molecular pathology in the diagnosis of cutaneous lymphomas. *Pathol Res Int*. 2012;2012:913523.
15. Herzog S, Hug E, Meixlsperger S, Paik J-H, DePinho RA, Reth M, *et al*. SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway. *Nat Immunol*. 2008;9(6):623-31.
16. Herzog S, Reth M, Jumaa H. Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nat Rev Immunol*. 2009;9(3):195-205.
17. Immunologie : organes, cellules et molécules [Internet]. [cité 8 juill 2019]. Disponible sur: <http://pst.chez-alice.fr/ts2tp.htm>
18. Klein U, Dalla-Favera R. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol*. 2008;8(1):22-33.
19. Cerutti A, Cols M, Puga I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat Rev Immunol*. 2013;13(2):118-32.
20. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, *et al*. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375-90.
21. Haute Autorité de Santé. Tumeur maligne, affection maligne du tissu lymphatique ou hématopoïétique Leucémie lymphoïde chronique [Internet]. 2011. Disponible sur: https://www.has-sante.fr/upload/docs/application/pdf/2011-09/ald_30_lap_llc_web_2vf.pdf
22. Heinivaara O. Smudge cells in lymphocytic leukemia. *Ann Med Intern Fenn*.

1959;48:69-75.

23. Cymbalist F. LLC: physiopathologie, diagnostic et approche thérapeutique. *Rev Francoph Lab.* 2013;2013(452):61-71.
24. Matutes E, Owusu-Ankomah K, Morilla R, Garcia Marco J, Houlihan A, Que TH, *et al.* The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia.* 1994;8(10):1640-5.
25. Moreau EJ, Matutes E, A'Hern RP, Morilla AM, Morilla RM, Owusu-Ankomah KA, *et al.* Improvement of the chronic lymphocytic leukemia scoring system with the monoclonal antibody SN8 (CD79b). *Am J Clin Pathol.* 1997;108(4):378-82.
26. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood.* 1975;46(2):219-34.
27. Binet JL, Lepage M, Dighiero G, Charron D, D'Athis P, Vaugier G, *et al.* A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer.* 1977;40(2):855-64.
28. Zengin N, Kars A, Kansu E, Özdemir O, Barişta İ, Güllü İ, *et al.* Comparison of Rai and Binet Classifications in Chronic Lymphocytic Leukemia. *Hematology.* 1997;2(2):125-9.
29. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med.* 2000;343(26):1910-6.
30. Letestu R, Lévy V, Eclache V, Baran-Marszak F, Vaur D, Naguib D, *et al.* Prognosis of Binet stage A chronic lymphocytic leukemia patients: the strength of routine parameters. *Blood.* 2010;116(22):4588-90.
31. Rossi D, Rasi S, Spina V, Brusca A, Monti S, Ciardullo C, *et al.* Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood.* 2013;121(8):1403-12.
32. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med.* 2005;352(8):804-15.
33. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature.* 2015;526(7574):525-30.
34. Nguyen-Khac F, Borie C, Callet-Bauchu E, Eclache V, Struski S. Cytogenetics in the management of chronic lymphocytic leukemia: an update by the Groupe francophone de cytogénétique hématologique (GFCH). *Ann Biol Clin (Paris).* 2016;74(5):561-7.
35. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell.* 2013;152(4):714-26.
36. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, *et al.* SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N Engl J Med.* 2011;365(26):2497-506.
37. Rossi D, Gaidano G. The clinical implications of gene mutations in chronic lymphocytic leukaemia. *Br J Cancer.* 2016;114(8):849-54.
38. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* 1999;94(6):1848-54.
39. Damle RN, Wasi T, Fais F, Ghiotto F, Valetto A, Allen SL, *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* 1999;94(6):1840-7.
40. Rosenquist R, Ghia P, Hadzidimitriou A, Sutton L-A, Agathangelidis A, Baliakas P, *et al.* Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia.* 2017;31(7):1477-81.
41. Schwartz RS, Stollar BD. Heavy-chain directed B-cell maturation: continuous clonal

selection beginning at the pre-B cell stage. *Immunology*. 1994;15(1):27-32.

42. Rao SP, Riggs JM, Friedman DF, Scully MS, LeBien TW, Silberstein LE. Biased VH gene usage in early lineage human B cells: evidence for preferential Ig gene rearrangement in the absence of selection. *J Immunol*. 1999;163(5):2732-40.
43. Chiorazzi N, Hatzi K, Albesiano E. B-cell chronic lymphocytic leukemia, a clonal disease of B lymphocytes with receptors that vary in specificity for (auto)antigens. *Ann N Acad Sci*. 2005;1062:1-12.
44. Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, *et al*. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *JCI Insight*. 1998;102(8):1515-25.
45. Messmer BT, Albesiano E, Efremov DG, Ghiotto F, Allen SL, Kolitz J, *et al*. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med*. 2004;200(4):519-25.
46. Stamatopoulos K. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood*. 2007;
47. Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan X-J, *et al*. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119(19):4467-75.
48. Darzentas N, Stamatopoulos K. Stereotyped B cell receptors in B cell leukemias and lymphomas. *Methods Mol Biol*. 2013;971:135-48.
49. Stamatopoulos K, Agathangelidis A, Rosenquist R, Ghia P. Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia*. 2017;31(2):282-91.
50. Granziero L, Ghia P, Circosta P, Gottardi D, Strola G, Geuna M, *et al*. Survivin is expressed on CD40 stimulation and interfaces proliferation and apoptosis in B-cell chronic lymphocytic leukemia. *Blood*. 2001;97(9):2777-83.
51. Puente XS, Jares P, Campo E. Chronic lymphocytic leukemia and mantle cell lymphoma: crossroads of genetic and microenvironment interactions. *Blood*. 2018;131(21):2283-96.
52. Soma LA, Craig FE, Swerdlow SH. The proliferation center microenvironment and prognostic markers in chronic lymphocytic leukemia/small lymphocytic lymphoma. *Hum Pathol*. 2006;37(2):152-9.
53. Herishanu Y, Pérez-Galán P, Liu D, Biancotto A, Pittaluga S, Vire B, *et al*. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood*. 2011;117(2):563-74.
54. Chiorazzi N. Implications of new prognostic markers in chronic lymphocytic leukemia. *Hematol Am Soc Hematol Educ Program*. 2012;2012:76-87.
55. Lanemo Myhrinder A, Hellqvist E, Sidorova E, Söderberg A, Baxendale H, Dahle C, *et al*. A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. *Blood*. 2008;111(7):3838-48.
56. Darwiche W, Gubler B, Marolleau J-P, Ghamlouch H. Chronic Lymphocytic Leukemia B-Cell Normal Cellular Counterpart: Clues From a Functional Perspective. *Front Immunol*. 2018;9.
57. Oppezio P, Magnac C, Bianchi S, Vuillier F, Tiscornia A, Dumas G, *et al*. Do CLL B cells correspond to naïve or memory B-lymphocytes? Evidence for an active Ig switch unrelated to phenotype expression and Ig mutational pattern in B-CLL cells. *Leukemia*. 2002;16(12):2438-46.
58. Toellner K-M, Jenkinson WE, Taylor DR, Khan M, Sze DM-Y, Sansom DM, *et al*. Low-level hypermutation in T cell-independent germinal centers compared with high mutation rates associated with T cell-dependent germinal centers. *J Exp Med*. 2002;195(3):383-9.
59. Gahn B, Schäfer C, Neef J, Troff C, Feuring-Buske M, Hiddemann W, *et al*. Detection

- of trisomy 12 and Rb-deletion in CD34+ cells of patients with B-cell chronic lymphocytic leukemia. *Blood*. 1997;89(12):4275-81.
60. Gahn B, Wendenburg B, Troff C, Neef J, Grove D, Haferlach T, *et al.* Analysis of progenitor cell involvement in B-CLL by simultaneous immunophenotypic and genotypic analysis at the single cell level. *Br J Haematol*. 1999;105(4):955-9.
 61. Di Ianni M, Baldoni S, Del Papa B, Aureli P, Dorillo E, De Falco F, *et al.* NOTCH1 Is Aberrantly Activated in Chronic Lymphocytic Leukemia Hematopoietic Stem Cells. *Front Oncol*. 2018;8.
 62. Quijada-Álamo M, Hernández-Sánchez M, Robledo C, Hernández-Sánchez J-M, Benito R, Montaña A, *et al.* Next-generation sequencing and FISH studies reveal the appearance of gene mutations and chromosomal abnormalities in hematopoietic progenitors in chronic lymphocytic leukemia. *J Hematol Oncol* *J Hematol Oncol*. 2017;10(1):83.
 63. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, *et al.* iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-60.
 64. Parikh SA. Chronic lymphocytic leukemia treatment algorithm 2018. *Blood Cancer J*. 2018;8(10).
 65. Rai KR, Jain P. Chronic lymphocytic leukemia (CLL)-Then and now. *Am J Hematol*. 2016;91(3):330-40.
 66. Quinquenel A, Fornecker L-M, Letestu R, Ysebaert L, Fleury C, Lazarian G, *et al.* Prevalence of BTK and PLCG2 mutations in a real-life CLL cohort still on ibrutinib after 3 years: a FILO group study. *Blood*. 2019;134(7):641-4.
 67. Kanagal-Shamanna R, Jain P, Patel KP, Routbort M, Bueso-Ramos C, Alhalouli T, *et al.* Targeted multigene deep sequencing of Bruton tyrosine kinase inhibitor-resistant chronic lymphocytic leukemia with disease progression and Richter transformation. *Cancer*. 2019;125(4):559-74.
 68. Blombery P, Anderson MA, Gong J-N, Thijssen R, Birkinshaw RW, Thompson ER, *et al.* Acquisition of the Recurrent Gly101Val Mutation in BCL2 Confers Resistance to Venetoclax in Patients with Progressive Chronic Lymphocytic Leukemia. *Cancer Discov*. 2019;9(3):342-53.
 69. Thangavadivel S, Byrd JC. Gly101Val BCL2 Mutation: One Step Closer to Understanding Venetoclax Resistance in CLL. *Cancer Discov*. 2019;9(3):320-2.
 70. Marsilio S, Khiabani H, Fabbri G, Vergani S, Scuoppo C, Montserrat E, *et al.* Somatic CLL mutations occur at multiple distinct hematopoietic maturation stages: documentation and cautionary note regarding cell fraction purity. *Leukemia*. 2018;32(4):1041-4.
 71. Van Dongen JJM, Langerak AW, Brüggemann M, Evans P a. S, Hummel M, Lavender FL, *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. 2003;17(12):2257-317.
 72. Bystry V, Agathangelidis A, Bikos V, Sutton LA, Baliakas P, Hadzidimitriou A, *et al.* ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. *Bioinformatics*. 2015;31(23):3844-6.
 73. Longo NS, Rogosch T, Zemlin M, Zouali M, Lipsky PE. Mechanisms That Shape Human Antibody Repertoire Development in Mice Transgenic for Human Ig H and L Chain Loci. *J Immunol*. 2017;198(10):3963-77.
 74. Muggen AF, de Jong M, Wolvers-Tettero ILM, Kallemeijn MJ, Teodósio C, Darzentas N, *et al.* The presence of CLL-associated stereotypic B cell receptors in the normal BCR repertoire from healthy individuals increases with age. *Immun Ageing A*. 2019;16.

75. Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, *et al.* Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol.* 2010;184(12):6986-92.
76. Wu J, Jia S, Wang C, Zhang W, Liu S, Zeng X, *et al.* Minimal Residual Disease Detection and Evolved IGH Clones Analysis in Acute B Lymphoblastic Leukemia Using IGH Deep Sequencing. *Front Immunol.* 2016;7.
77. Kotrova M, Trka J, Kneba M, Brüggemann M. Is Next-Generation Sequencing the way to go for Residual Disease Monitoring in Acute Lymphoblastic Leukemia? *Mol Diagn Ther.* 2017;21(5):481-92.
78. Pulsipher MA, Carlson C, Langholz B, Wall DA, Schultz KR, Bunin N, *et al.* IgH-V(D)J NGS-MRD measurement pre- and early post-allotransplant defines very low- and very high-risk ALL patients. *Blood.* 2015;125(22):3501-8.
79. Peng Q, Xu C, Kim D, Lewis M, DiCarlo J, Wang Y. Targeted Single Primer Enrichment Sequencing with Single End Duplex-UMI. *Sci Rep.* 2019;9(1):1-10.
80. Zalberg I, D'Andrea MG, Monteiro L, Pimenta G, Xisto B. Multidisciplinary diagnostics of chronic lymphocytic leukemia: European Research Initiative on CLL - ERIC recommendations. *Hematol Transfus Cell Ther [Internet].* 2019 [cité 22 sept 2019]; Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S2531137919301324>
81. Vuillier F, Claisse JF, Vandenvelde C, Travade P, Magnac C, Chevret S, *et al.* Evaluation of residual disease in B-cell chronic lymphocytic leukemia patients in clinical and bone-marrow remission using CD5-CD19 markers and PCR study of gene rearrangements. *Leuk Lymphoma.* 1992;7(3):195-204.
82. Letestu R, Garff-Tavernier ML. Determination of residual disease by flow cytometry in chronic lymphocytic leukemia: state of play of technical practices and prospects. *Hématologie.* 2019;25(2):24-35.
83. Dimier N, Delmar P, Ward C, Morariu-Zamfir R, Fingerle-Rowson G, Bahlo J, *et al.* A model for predicting effect of treatment on progression-free survival using MRD as a surrogate end point in CLL. *Blood.* 2018;131(9):955-62.
84. Böttcher S, Ritgen M, Fischer K, Stilgenbauer S, Busch RM, Fingerle-Rowson G, *et al.* Minimal residual disease quantification is an independent predictor of progression-free and overall survival in chronic lymphocytic leukemia: a multivariate analysis from the randomized GCLLSG CLL8 trial. *J Clin Oncol.* 2012;30(9):980-8.
85. Rawstron AC, Fazi C, Agathangelidis A, Villamor N, Letestu R, Nomdedeu J, *et al.* A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. *Leukemia.* 2016;30(4):929-36.
86. Quinquenel A. Relevance of MRD in CLL: recent data and role of combinations. *Hématologie.* 2019;25(2):17-23.
87. Morbach H, Eichhorn EM, Liese JG, Girschick HJ. Reference values for B cell subpopulations from infancy to adulthood. *Clin Exp Immunol.* 2010;162(2):271-9.
88. Ghazzaoui N, Issaoui H, Saintamand A, Oblet C, Carrion C, Denizot Y. The immunoglobulin heavy chain 3' regulatory region superenhancer controls mouse B1 B-cell fate and late VDJ repertoire diversity. *Blood Adv.* 2018;2(3):252-62.

Annexes

Annexe 1. Protocole de CMF	88
Annexe 2. Amorces leader	89
Annexe 3. Tableau de séquences des LLC du pool	90

Annexe 1. Protocole de CMF

Le protocole utilisé pour le dépistage de clone B dans le sang total.

- Prélever 200µL de sang total
- Réaliser 3 lavages successifs avec du PBS (centrifugation 5 min à 300g)
- Prélevé 50µL de sang « lavé » et ajouter 4µL de chaque Ac
- Incuber 15min à température ambiante à l'abri de la lumière
- Ajouter 500µL de Lysis Buffer®
- Incuber 15min à température ambiante à l'abri de la lumière
- Faire un lavage avec du PBS (centrifugation 5min à 300g)
- Éliminer le surnageant
- Ajouter 300µL de PBS
- Passer au cytomètre

Annexe 2. Amorces leader

5' – 3'	Adaptateur Illumina	UMI	Amorce spécifique
VH1.69	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GGACTGGACCTGGAGGTTCCCT
VH1.18	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GGACTGGACCTGGAGCATCC
VH1.46	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	TCTGCTTGGCTGGCTGTAGCTC
VH1C	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CTCTTCTTGGTGGCAGCAGC
VH2.70	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	ATACTTTGTTCCACGCTCCTGCT
VH2.26	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	ACACTTTGCTACACACTCCTGCTG
VH2.5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CACTTTGCTCCACGCTCCTG
VH3.9	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	ACTGAGCTGGATTTTCCTTTTGG
VH3E	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CTGAGCTGGGTTTTCTCTGTTG
VH3G	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GGAGTTTGGGCTGAGCTGG
VH4A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GAAACATCTGTGGTTCTTCCTTCTC
VH4B	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GAAGCACCTGTGGTTCTTCCTCC
VH4C	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GAAACACCTGTGGTTCTTCCTCC
VH5A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GTCAACCGCCATCCTCGC
VH5B	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	CAGGCTCCCCTCCACAGTG
VH6B	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	TCTGTCTCCTTCCTCATCTTCCTC
VH7A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	NNNNNNNNNN	GTGCCCACTCCCAGGTGC
JH	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG		CTTACCTGAGGAGACGGTGACC

Annexe 3. Tableau de séquences des LLC du pool

Échantillon	% mutation IGHV	V – gène	J – gène	Séquence CDR3	Taille CDR3
LLC #1	2,2	V3-21	J6	ARDANGMDV	9
LLC #2	14,4	V3-64D	J4	VSLIFRDRIFDY	12
LLC #3	0	V1-69	J4	ARAGDFWSGHNY	12
LLC #4	6,8	V1-3	J4	ASRPGMAVAVFDY	13
LLC #5	0	V5-10-1	J4	ARQQWLGGYYFDY	13
LLC #6	7	V4-34	J3	ARGFVTTQFADFID	15
LLC #7	0	V4-39	J4	ARHDSSSWWTRYWGY	15
LLC #8	6,6	V1-2	J6	ARDSGYGDYVYYGMDV	16
LLC #9	6,6	V2-5	J4	AHRLSGYNWNDADFIDY	16
LLC #10	1,35	V1-46	J4	AKQSYDSSGSSLGLDY	17
LLC #11	2,6	V3-9	J6	AKSTATGDRAYYYYGMDV	18
LLC #12	0	V4-59	J5	ARHGCTNGVCYLSDFWDP	19
LLC #13	0	V1-18	J6	ARGGSSTSFERVPYYYMDV	20
LLC #14	0	V5-51	J4	ARPAYFDDFWSGYHLPGYFDY	21

Serment De Galien

Je jure en présence de mes Maîtres de la Faculté et de mes condisciples :

- d'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement ;
- d'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement ;
- de ne jamais oublier ma responsabilité, mes devoirs envers le malade et sa dignité humaine, de respecter le secret professionnel.

En aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser les actes criminels.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert d'opprobre et méprisé de mes confrères, si j'y manque.

Mise au point du séquençage haut débit des gènes des immunoglobulines et application à la recherche de BCRs stéréotypés de type « LLC » chez des patients exempts d'hémopathie

Introduction : La Leucémie Lymphoïde Chronique (LLC) est une pathologie d'évolution lente qui se manifeste cliniquement après 60 ans. Les clones de LLC ont un répertoire de séquences V(D)J particulier et dans 30% des cas leurs BCRs (*B cell receptor*) sont « stéréotypés » malgré la diversité des réarrangements possibles (10^{12}). L'objectif principal de ce travail a été de rechercher ses séquences particulières au sein d'une population « saine » pouvant être le témoin d'une prédisposition de la sélection de ces clones. Pour répondre à cette question, nous avons développé une méthode de séquençage haut débit des gènes des immunoglobulines (NGS). **Méthode :** Nous avons mis au point une méthode NGS à partir d'ADN et développé un *pipeline* bio-informatique simple permettant le traitement de ces données générées. Une recherche de BCRs stéréotypés de type « LLC » au sein d'une cohorte de 19 patients exempts d'hémopathie a été réalisée. **Résultats :** La qualité des séquences obtenues par notre approche est très satisfaisante. Au niveau quantitatif, nous avons pu détecter des abondances relatives de différents clones au seuil de 1.10^{-4} . Avec l'analyse bio-informatique mise en place, nous avons pu mettre en évidence la présence de BCRs stéréotypés au sein d'une cohorte de 19 patients exempts d'hémopathie. **Discussion :** D'une part, la détection de différents clones ayant une abondance à 1.10^{-4} démontre la grande sensibilité de notre technique et permet d'envisager cet outil dans l'étude de la maladie résiduelle dans la LLC notamment. D'autre part, la présence de BCRs stéréotypés dans une cohorte de 19 patients « sains » prouve une production physiologique, ubiquitaire de lymphocytes B type « LLC ».

Mots-clés : Leucémie lymphoïde chronique, BCR stéréotypé, séquençage haut débit, immunoglobulines

Development of high throughput sequencing of immunoglobulin genes and application to the search for stereotyped BCR of the "CLL" type in patients free of hematopoietic malignancy

Introduction : Chronic Lymphoid Leukemia (CLL) is an indolent disease that manifests clinically after 60 years old. CLL clones have a particular V(D)J sequences repertoire and in 30% of cases their BCRs (B cell receptor) are "stereotyped" despite the diversity of the possible rearrangements (10^{12}). The main objective of this work was to research these particular sequences in a "healthy" population to explore the propensity of our immune system to produce "CLL-type" B cells. To answer this question, we have developed high throughput sequencing method of immunoglobulins genes (NGS). **Method :** We performed a PCR-based method on DNA and a simple bioinformatics pipeline for processing the generated data. A search for stereotyped "CLL-type" BCRs in a cohort of 19 patients free of hematopoietic malignancy was conducted. **Results :** The quality of the sequences obtained by our approach is very satisfactory. At the quantitative level, we were able to detect relative abundances of different clones at the 1.10^{-4} threshold. With the bioinformatics analysis implemented, we were able to highlight the presence of stereotyped BCRs in a cohort of 19 "healthy" patients. **Discussion :** First, the detection of different clones with an abundance of 1.10^{-4} demonstrates the great sensitivity of our technique and allows us to consider this tool in the study of minimal residual disease in CLL notably. Finally, the presence of stereotyped BCRs in a cohort of 19 "healthy" patients proves a physiological, ubiquitous production of "CLL-type" B lymphocytes.

Keywords : Chronic Lymphoid Leukemia, stereotyped BCR, high throughput sequencing, immunoglobulins

