Normandie Université

THESE

Pour obtenir le grade de Docteur

opéré par l'Université de Rouen

Spécialité Biologie

Caractérisation multi-omique des Lymphomes B Diffus à Grandes Cellules

Présentée et soutenue publiquement par
Sylvain MARESCHAL

Thèse soutenue publiquement le 11 décembre 2015 devant le jury composé de		
Mr Gilles SALLES	PU PH, Hospices Civils de Lyon Université Claude Bernard Lyon I	Président
Mme Thérèse COMMES	PU PH, Institut de Recherche en Biothérapie Université Montpellier I	Rapporteur
Mr Bertrand NADEL	DR, Centre d'Immunologie de Marseille-Luminy Université Aix Marseille	Rapporteur
Mr Emmanuel BARILLOT	DR, Institut Curie Mines ParisTech	Membre
Mr Jean-Philippe JAIS	MCU PH, Plateforme de bioinformatique Université Paris-Descartes	Membre
Mr Fabrice JARDIN	PU PH, INSERM 918, Centre Henri Becquerel Université de Rouen	Membre
Mr Thierry LECROQ	PR, LITIS équipe TIBS Université de Rouen	Membre

Thèse dirigée par Fabrice JARDIN et Thierry LECROQ

UNIVERSITÉ DE ROUEN

EdNBISE
Ecole doctorale Normande
Biologie Intégrative,
Santé, Environnement

CENTRE HENRI BECQUEREL
HAUTE-NORMANDIE

Résumé

Français

Caractérisation multi-omique des Lymphomes B Diffus à Grandes Cellules

Les lymphomes B diffus à grandes cellules (LDGCB) sont des prolifération cancéreuses agressives prenant leur origine dans le système lymphatique, et qui représentent 4 000 nouveaux cas par an en France. Si des progrès significatifs ont été réalisés ces dernières années concernant leur prise en charge, un peu plus d'un tiers des patients ne répondent pas aux immuno-chimiothérapies actuelles. Les travaux présentés dans ce mémoire visaient à caractériser ces tumeurs par plusieurs techniques à haut débit, afin de mettre en évidence des altérations somatiques susceptibles d'expliquer ou de faire suspecter dès le diagnostic ce phénotype réfractaire. Sur le plan transcriptomique, la sous-classification en lymphomes « Germinal Center B-cell like » et « Activated B-Cell like » grâce à un nouveau test simple et fiable permet désormais d'identifier les patients atteints de ce second sous-type de moins bon pronostic. Sur le plan génomique, le développement d'outils d'analyse bio-informatique pour les puces CGH a conduit à la confirmation de nombreux gains et pertes de copies dans les génomes de ces tumeurs, et à la caractérisation plus fine de la perte du gène *CDKN2A*. Sur le plan mutationnel enfin, le séquençage exomique de 14 LDGCB réfractaires a permis de mettre en évidence de nouveaux gènes dont la mutation pourrait expliquer la résistance au traitement. Chacun de ces trois volets a fait l'objet d'un état des lieux bibliographique, et de développements d'outils bio-informatiques qui pourront à l'avenir être également appliqués à d'autres types de cancers.

Mots-clés

Lymphome B, Bio-informatique

Test diagnostique, CGH, Séquençage de nouvelle génération

Anglais

Multi-omic characterization of Diffuse Large B-Cell Lymphoma

Diffuse large B-cell lymphomas (DLBCLs) are tumors originating in the lymphatic system, accounting for 4 000 new cases per year in France. While much progress has been made regarding their treatment, one of three patients still does not respond to current immuno-chemotherapies. The present work consisted in characterizing these cancers using various high-throughput technologies, in order to identify somatic alterations that could explain this refractoriness or allow it to be suspected at diagnosis. Gene expression profiling had previously identified two subtypes with distinct outcomes, termed « Activated B-Cell like » and « Germinal Center B-cell like », which we were able to identify using a new simple diagnostic test. The development of bioinformatics software to handle CGH array data confirmed several copy number gains and losses in DLBCLs, and led to a more precise characterization of *CDKN2A* loss. Finally the sequencing of 14 exomes of relapsed or refractory DLBCLs produced an interesting picture of somatic mutations associated with this phenotype, and highlighted several new leads. An extensive review of the bibliography is proposed on each of these three aspects of tumoral genome alteration, as well as several new bioinformatics methods that may be applied to distinct cancer types in a near future.

Keywords

B-cell lymphoma, Bioinformatics

Diagnostic test, CGH array, Next-generation sequencing

Laboratoire d’accueil

Unité INSERM 918, Génétique et clinique des prolifération lymphoïdes
Centre Henri Becquerel
Rue d’Amiens
76038 Rouen CEDEX 1

Remerciements

A mon directeur de thèse, le Pr Fabrice Jardin, pour la confiance qu'il m'a témoignée durant ces années de master et de thèse, et pour les nombreuses opportunités professionnelles qui m'ont été offertes. Pour sa disponibilité et son implication également, que je sais précieuses du fait de ses nombreuses obligations. J'étends ces remerciements au Pr Hervé Tilly, avec qui il a su maintenir une unité de recherche dynamique, sur un sujet stimulant.

A mon co-directeur de thèse, le Pr Thierry Lecroq, pour avoir toujours répondu « présent » à mes (trop rares) sollicitations sur des aspects bio-informatiques. Plus généralement à l'ensemble de l'équipe pédagogique du master de bio-informatique de Rouen, et tout spécialement au Dr Hélène Dauchel sans qui rien de tout ceci n'aurait été possible.

A mon jury de thèse, pour avoir accepté cette charge. Au Pr Gilles Salles et au Dr Emmanuel Barillot, pour l'expertise clinique et bio-informatique dont ils me font l'honneur de bénéficier par leur présence. Au Pr Thérèse Commes et au Dr Bertrand Nadel, non moins versés dans ces deux domaines, pour avoir accepté la lourde tâche de rapporteur sur cette thèse. Enfin au Dr Jean-Philippe Jaïs, pour sa présence dans ce jury et pour sa collaboration tout au long de ma thèse et de mon master.

Au Dr Philippe Ruminy, pour son encadrement au quotidien, sa bonne humeur et les nombreux projets qui ont pu nous occuper. Pour sa créativité et son pragmatisme, qui manquent trop souvent dans le domaine de la recherche et sont pourtant indispensables à toute recherche visant un progrès tangible pour la société. Aux Dr Philippe Bertrand et Florian Clatot, pour leur rigueur méthodologique et statistique à toute épreuve qui viennent compléter ce tableau, et le plaisir que j'ai pu avoir à travailler avec eux.

A Sydney, pour son indéfectible soutien moral, scientifique et linguistique (parfois même artistique). C'était un plaisir et un privilège que de partager ce bureau et ces années de thèse avec toi, je te souhaite toute la réussite que tu mérites et qui t'attend, j'en suis convaincu.

Plus généralement à l'ensemble des étudiants en médecine, pharmacie ou sciences passés dans ce bureau ou au sein du laboratoire durant ces années. A David et Thomas, pour la curiosité qu'ils ont pu témoigner vis à vis de R et de la bio-informatique en général, et avec qui cela a toujours été un plaisir de discuter et travailler. A Pierre-Julien, qui reprendra le flambeau avec brio j'en suis sûr. A Vincent, pour son inépuisable connaissance de la clinique des lymphomes. A Hélène, pour sa bonne humeur et son franc-parler. A Marion, pour les

bons moments passés au laboratoire et ailleurs, en dépit de nos différences. A Abdel, autre partenaire de galère dans cette aventure qu'a été la thèse. A tous les autres que je ne peux que nommer faute de place : Bruno, Fanny, Aurélie, Bérangère, Camille, Anne-Lise, Suzan, Laura, « petit » David, Marion « 2 », Florian, Emilie, Ophélie, Ismaël, Sophie, Jean-Yves...

A l'ensemble du personnel des plateaux techniques de recherche, cytogénétique et biologie moléculaire que j'ai pu côtoyer au quotidien. Au Dr Christian Bastard, Dominique et Pascaline qui veillent sur tout ce monde et m'ont souvent fait profiter de leur expertise. A Catherine, Françoise, Sandrine et Nathalie pour leur compétence et leur gentillesse. A Vinciane, Mathieu, Soizic, « la p'tite » Lætitia, Anne-Laure, Marie, Mélissa et Annie pour l'animation et la bonne humeur que leur présence suscite dans les laboratoires. A Élodie, technicienne sur le papier seulement qui sortira des merveilles du PGM, à n'en pas douter.

A tous ceux qui se sont impliqués dans la relecture de ce mémoire de thèse, chacun dans son domaine respectif. Tout spécialement à Clara, pour son goût immodéré des virgules, ses encouragements et son enthousiasme. A l'ensemble de mes amis, Rouennais intérimaires et Grenoblois de souche, et à ma famille pour m'avoir toujours soutenu dans ce choix professionnel qui m'a pourtant gardé longtemps loin d'eux.

Abréviations

ABC *Activated B-Cell like*

ACP Analyse en Composante Principale

ADN Acide DésoxyriboNucléique

ADNc ADN complémentaire

APOBEC *APOlipoprotein B mRNA Editing enzyme, Catalytic polypeptide-like*

ARN Acide RiboNucléique

BAC *Bacterial Artificial Chromosome*

BAM *Binary (sequence) Alignment Map*

BCNU Bis-ChloroethylNitrosoUrea

BCR *B-Cell Receptor*

BEAM BCNU, Étoposide, Aracytine, Melphalan

BQSR *Base Quality Score Recalibration*

BWA *Burrows-Wheeler Aligner*

CDR *Complementarity Determining Region*

CGH *Comparative Genomic Hybridization*

CHOP Cyclophosphamide, Hydroxyadriamycine (doxorubicine), Oncovin (vincristine) et Prednisone

CNA *Copy Number Alteration*

CNV *Copy Number Variation*

CNLOH *Copy-Neutral Loss of Heterozigosity*

COFRAC Comité français d'accréditation

CPA Cellules Présentatrices d'Antigènes

CpG Cytosine et Guanine adjacentes sur un même brin d'ADN (séparées par un phosphate)

CSH Cellule Souche Hématopoïétique

DA-EPOCH-R *Dose-Adjusted Rituximab, Etoposide, Prednisolone, Oncovin (vincristine), Cyclophosphamide, Hydroxydaunorubicin (doxorubicin) and Rituximab*

DASL *cDNA-mediated Annealing, Selection, extension, and Ligation*

DSBR *Double-Strand Break Repair*

EBV *Epstein-Barr Virus*

ECOG *Eastern Cooperative Oncology Group*

EFS *Event-Free Survival*

ESMO *European Society for Medical Oncology*

FDG FluoroDeoxyGlucose ^{18}F

FFPE *Formalin-Fixed Paraffin-Embedded*

FISH *Fluorescent In-Situ Hybridization*

GATK *Genome Analysis ToolKit*

GCB *Germinal Center B-cell like*

GEDI *Gene Expression and Dosage Integrator*

GELA Groupe d'Étude des Lymphomes de l'Adulte

HPST Hôpital, Patients, Santé et Territoires

HR+ *Hormone-Receptor positive*

IC_{95} Intervalle de Confiance à 95%

IF *Impact Factor*

IHC ImmunoHistoChimie

IPI *International Prognostic Index*

IPIaa IPI ajusté à l'âge

IVD *In-Vitro Diagnostics*

LAGC Lymphome Anaplasique à Grandes Cellules

LAL Leucémie Aiguë Lymphoblastique

LBPM Lymphome B Primitif du Médiastin

LCM Lymphome à Cellules du Manteau

LDGCB Lymphome Diffus à Grandes Cellules de phénotype B

LDH Lactate DésHydrogénase

LF Lymphome Folliculaire

LFt Lymphome Folliculaire transformé

LH Lymphome Hodgkinien

LINE *Long Interspersed Nuclear Elements*

LLC Leucémie Lymphoïde Chronique

LMC Leucémie Myéloïde Chronique

LNH Lymphome Non-Hodgkinien

LOWESS *LOcally WEighted Scatterplot Smoothing*

LPM Lymphome Primitif du Médiastin

LPS *Linear Predictor Score*

LYSA *LYmphoma Study Association*

MCR *Minimal Common Region*

MEM *(Super)Maximal Exact Match*

MHC *Major Histocompatibility Complex*

NHEJ *Non-Homologuous End Joining*

NK *Natural Killer*

NOS *Not Otherwise Specified*

OMS Organisation Mondiale de la Santé

PCNSL *Primary Central Nervous System Lymphoma*

PLC Progéniteur Lymphoïde Commun

POSIX *Portable Operating System Interface*

qNPA *Quantitative Nuclease Protection Assay*

qPCR *Quantitative Polymerase Chain Reaction*

REAL *Revised European American Lymphoma classification*

R-CHOEP Rituximab, Cyclophosphamide, Hydroxyadriamycine (doxorubicine), Oncovin (vincristine), Étoposide et Prednisone

R-CHOP CHOP + Rituximab

R-DHAP Rituximab, DexamétHasone, Aracytine, cisPlatine

R-GDP Rituximab, Gemcitabine, Dexaméthasone, cisPlatine

R-GEMOX Rituximab, GEMcitabine, OXaliplatine

R-ICE Rituximab, Ifosfamide, Carboplatine, Étoposide

RNA-seq *(whole) RiboNucleic Acid sequencing*

RT-MLPA *Reverse-Transcriptase Multiplex Ligation dependent Probe Amplification*

RSS *Recombination Signal Sequences*

SAM *Sequence Alignment Map*

SHM *Somatic Hyper-Mutation*

SINE *Short Interspersed Nuclear Elements*

SNC Système Nerveux Central

SNP *Single Nucleotide Polymorphism*

SNV *Single Nucleotide Variation*

TDM TomoDensitoMétrie

TdT *Terminal deoxynucleotidyl transferase*

TEP Tomographie par Émission de Positons

TLDA *Taqman Low-Density Array*

TLR *Toll-Like Receptor*

TMA *Tissue MicroArray*

UTR *UnTranslated Region*

VIH Virus de l'Immunodéficience Humaine

WES *Whole-Exome Sequencing*

WGS *Whole-Genome Sequencing*

Table des matières

Résumé **3**

Remerciements **5**

Abréviations **7**

Table des matières **12**

Table des figures **14**

1 Introduction **15**

1.1 Avant-propos 15
1.2 Lymphocytes B et immunité adaptative 15
1.3 Lymphopoïèse B 17
1.3.1 Cellules souches et progéniteurs 17
1.3.2 Stade pro-B 18
1.3.3 Stade pré-B 20
1.3.4 Stade immature 20
1.3.5 Stade mature 21
1.4 Immunopoïèse B 21
1.4.1 Follicules primaires 21
1.4.2 Rencontre avec l'antigène 23
1.4.3 Activation par les lymphocytes T 23
1.4.4 Formation du centre germinatif 25
1.4.5 Maturation d'affinité 26
1.4.6 Commutation isotypique 26
1.4.7 Différenciation terminale 28
1.5 Classification des hémopathies malignes 29
1.5.1 Systèmes de classification 29

1.5.2 Progéniteurs B . . . 29
1.5.3 Lymphocytes B naïfs . . . 31
1.5.4 Lymphocytes B activés par l'antigène . . . 31
1.5.5 Lymphocytes B différenciés . . . 33
1.5.6 Autres lignées hématopoïétiques . . . 34
1.6 Lymphomes B diffus à grandes cellules . . . 35
1.6.1 Présentation biologique . . . 35
1.6.2 Présentation clinique . . . 37

2 Caractères transcriptomiques 43
2.1 État de l'art . . . 43
2.1.1 Sous-types GCB et ABC . . . 43
2.1.2 Autres caractères transcriptomiques . . . 49
2.2 Objectifs et approche suivie . . . 52
2.3 Méthodes . . . 53
2.3.1 DASL . . . 53
2.3.2 RT-MLPA . . . 54
2.3.3 LPS . . . 55
2.4 Classification GCB-ABC par DASL . . . 56
2.4.1 Première approche : *clustering* non-supervisé . . . 56
2.4.2 Seconde approche : modèle LPS . . . 57
2.5 Classification GCB-ABC par RT-MLPA . . . 59
2.5.1 Choix des gènes . . . 59
2.5.2 Développement de l'outil . . . 60
2.5.3 Résultats et discussion . . . 60
2.5.4 Article publié (JMD 2015) . . . 61

3 Caractères génomiques 73
3.1 État de l'art . . . 73
3.1.1 Altérations du nombre de copies (CNA) . . . 73
3.1.2 Translocations . . . 84
3.2 Objectifs et approche suivie . . . 90
3.3 Méthodes . . . 91
3.3.1 Hybridation Génomique Comparative (CGH) . . . 91
3.4 Bibliothèque logicielle génomique : Rgb . . . 94
3.4.1 Choix technologiques . . . 94

3.4.2 Implémentation . . . 95
3.4.3 Résultats et discussion . . . 97
3.4.4 Article publié (Bioinformatics 2014) . . . 97
3.5 Bibliothèque logicielle d'analyse CGH : cghRA . . . 101
3.5.1 Vue d'ensemble . . . 101
3.5.2 Modélisation du nombre de copies . . . 103
3.5.3 Filtration des polymorphismes . . . 107
3.5.4 Priorisation des MCR . . . 109
3.6 Application à la série GHEDI . . . 111
3.6.1 Plan d'expérience . . . 111
3.6.2 Résultats et discussion . . . 111

4 Caractères mutationnels 115
4.1 État de l'art . . . 115
4.1.1 Mécanistique . . . 115
4.1.2 Séquençage à haut débit des LDGCB . . . 118
4.1.3 Mutations récurrentes dans les LDGCB . . . 124
4.2 Objectifs et approche suivie . . . 129
4.3 Méthodes . . . 130
4.3.1 Séquençage exomique . . . 130
4.4 Séquençage d'exomes de LDGCB réfractaires . . . 136
4.4.1 Pipeline d'analyse . . . 136
4.4.2 Résultats et discussion . . . 140
4.4.3 Article accepté (GCC) . . . 141

5 Bilan : oncogénèse des LDGCB 183
5.1 Événements précoces . . . 183
5.1.1 Dérégulations épigénétiques . . . 183
5.1.2 Translocations de *BCL2* . . . 185
5.1.3 Mutations de *CD79A* et *CD79B* . . . 186
5.2 Au cœur de l'oncogénèse . . . 186
5.2.1 La voie du BCR . . . 186
5.2.2 La voie des TLR . . . 189
5.2.3 La voie NFκB . . . 190
5.3 Un nouveau programme transcriptomique . . . 191
5.3.1 Survie cellulaire . . . 191
5.3.2 Blocage de la différenciation . . . 193
5.3.3 Évasion immunitaire . . . 195

6 Discussion **197**
6.1 Transfert de la classification GCB / ABC . . . 197
6.1.1 Contexte . . . 197
6.1.2 Solution *NanoString Technologies* . . . 197
6.1.3 Solution RT-MLPA . . . 199
6.1.4 Perspectives . . . 201
6.2 Quels outils pour détecter les CNA ? . . . 203
6.2.1 Contexte . . . 203
6.2.2 Puces SNP . . . 203
6.2.3 Séquençage à haut débit . . . 205
6.2.4 Techniques de transfert . . . 206
6.3 Identifier les tumeurs réfractaires . . . 207
6.3.1 Contexte . . . 207
6.3.2 Séquençage exomique . . . 207
6.3.3 Transfert à la pratique clinique . . . 209

A Annexes **213**
A.1 Revue publiée (Leuk. Lymphoma 2015) . . . 213

Bibliographie **225**

Table des figures

1.1 Lymphopoïèse B dans la moelle osseuse . . . 19
1.2 Organisation d'un ganglion lymphatique . . . 22
1.3 Déroulement de la réaction immune adaptative . . . 24
1.4 Commutation isotypique . . . 27
1.5 Incidence des principales hémopathies malignes . . . 30
1.6 Mise en évidence d'un lymphome par TDM et TEP . . . 41

2.1 *Clustering* GCB/ABC originel sur puces Lymphocip . . . 44
2.2 Principaux algorithmes IHC appliqués aux LDGCB . . . 46
2.3 Principe de la technologie DASL . . . 54
2.4 Principe de la RT-MLPA . . . 55
2.5 *Clustering* de 131 LDGCB basé sur 19 gènes GCB / ABC . . . 57
2.6 Classification GCB / ABC par LPS dans 189 LDGCB . . . 58
2.7 Expression différentielle GCB / ABC de *LMO2* et *TNFRSF13B* . . . 59

3.1 Anomalies de mitoses et altérations chromosomiques . . . 74
3.2 Modèle *Double-Strand Break Repair* . . . 75
3.3 Recombinaisons et altérations chromosomiques . . . 76
3.4 Distribution du nombre de chromosomes dans les LDGCB . . . 78
3.5 Bilan des CNA répertoriés dans les LDGCB (partie 1) . . . 80
3.6 Bilan des CNA répertoriés dans les LDGCB (partie 2) . . . 81
3.7 Principales translocations impliquant *IGH* . . . 86
3.8 Principe de la CGH . . . 92
3.9 Diagramme UML des principales classes de Rgb . . . 96
3.10 Exemples de représentations générées par Rgb . . . 98
3.11 Interface de cghRA . . . 102
3.12 Modèles de prédiction du nombre de copies . . . 105
3.13 Validation des prédictions du nombre de copies . . . 106

3.14 Validation des prédictions de polymorphismes 109
3.15 Pénétrance et MCR 110
3.16 Relation entre délétion de *CDKN2A* et expression génique 112
3.17 Impact des délétions de *CDKN2A* sur l'expression de ses cibles 112
3.18 Impact pronostique des délétions de *CDKN2A* 112

4.1 Principales signatures mutationnelles dans les cancers 116
4.2 Gènes d'intérêt dans les LDGCB pointés par 4 études NGS 120
4.3 Evolution clonale dans les syndromes de Richter 122
4.4 Étude de saturation pour 21 types de cancers 125
4.5 Bilan des SNV répertoriés dans les LDGCB 126
4.6 Principe de la capture exomique SureSelect™ 131
4.7 Principe du séquençage Illumina 133
4.8 Principe du séquençage *paired-ends* 135
4.9 Présentation détaillée du pipeline d'analyse d'exomes 137

5.1 Bilan des caractéristiques récurrentes des LDGCB ABC, GCB et LPM . . 184
5.2 Altérations récurrentes au sein des voies du BCR, des TLR et NFκB . . . 187

6.1 Principe de la technologie nCounter™ 198
6.2 Application de la RT-MLPA à la recherche de mutations 201
6.3 Exemple du profil RT-MLPA en développement 202
6.4 Exemple de profil génomique par puce SNP 204
6.5 Exemple de CNA détectés par séquençage exomique 205
6.6 Exemple de CNA détectés par reséquençage ciblé 206
6.7 Exemple de fiche de synthèse d'une RCP moléculaire 210
6.8 Gènes mutés dans les LDGCB cutanés primitifs de la jambe 211

1 Introduction

1.1 Avant-propos

Le présent mémoire vise à détailler une série de travaux réalisés sur les Lymphomes Diffus à Grandes Cellules de phénotype B (LDGCB), dans le but de mieux cerner les caractéristiques génomiques et transcriptomiques de ces tumeurs et leurs potentiels impacts cliniques. Afin de replacer ces résultats dans le contexte des connaissances actuelles, les mécanismes de la différenciation physiologique des lymphocytes B à l'origine de ces tumeurs seront rappelés en introduction (§1.2, §1.3, §1.4). Sauf indication contraire, les éclaircissements donnés dans ces trois sections sont basés sur la référence bibliographique [1]. Un panorama des différentes transformations malignes pouvant survenir au cours de cette différenciation sera ensuite établi. Il permettra de replacer l'entité des LDGCB au sein du groupe des hémopathies B (§1.5). Les aspects biologiques et cliniques spécifiques des LDGCB seront enfin présentés en détail avant d'entrer dans le vif du sujet (§1.6).

Les différents travaux réalisés seront présentés en trois parties, selon qu'ils visaient à déterminer des caractères transcriptomiques (§2), génomiques (§3) ou mutationnels (§4), chaque partie étant précédée d'un état de l'art bibliographique. Ces différents points de vue seront repris pour établir un bilan des mécanismes de l'oncogénèse des LDGCB (§5), et enfin une remise en perspective des trois principaux travaux de cette thèse sera proposée (§6).

1.2 Lymphocytes B et immunité adaptative

Les lymphocytes représentent 20% à 40% des « globules blancs » circulant dans la sang périphérique (leucocytes), ces derniers se distinguant des érythrocytes et thrombocytes fortement majoritaires par la présence d'un noyau (1-4 10^9, 4-6 10^{12} et 2-4 10^{11} par litre). Parmi les lymphocytes, 10% à 15% appartiennent à la lignée B,

80% à la lignée T et moins de 10% sont des lymphocytes NK. La lignée lymphocytaire B se caractérise par la production d'**immunoglobulines**, excrétées (anticorps) ou sous forme membranaire (BCR), qui constituent la base de l'**immunité adaptative**. A l'opposé de l'immunité innée qui peut être mobilisée en quelques heures en cas d'infection, les mécanismes de l'immunité adaptative sont plus lents (quelques jours) et reposent sur la reconnaissance spécifique du corps étranger (antigène, cellulaire ou moléculaire) par ces immunoglobulines, déclenchant une série de réactions visant à l'éradiquer.

Les lymphocytes B se présentent ainsi comme un pool de clones, chacun doté d'une immunoglobuline à même de lier un antigène différent, circulant dans l'organisme à la recherche de leur cible. La diversité de ces immunoglobulines est donc un élément clé de l'immunité adaptative, produit *a priori* selon des mécanismes aléatoires (§1.3) qui leur offrent un niveau de variabilité exceptionnel au sein du génome humain (on estime en effet qu'il existe environ 10^7 variantes fonctionnelles chez l'Homme). Cette grande diversité des antigènes couplée au faible nombre de lymphocytes B exprimant chacun un clonotype d'immunoglobuline rendant cette rencontre relativement improbable, cette dernière se trouve grandement facilitée par les cellules présentatrices d'antigènes (CPA ; monocytes, macrophages, cellules dendritiques ou lymphocytes B activés). Après avoir internalisé et digéré un corps étranger, ces CPA migrent en effet vers les organes lymphoïdes secondaires, où elles mettent à disposition des lymphocytes B et T qui y résident les antigènes obtenus.

Lorsqu'un lymphocyte B vient à rencontrer l'antigène qu'il est génétiquement programmé pour reconnaître, il se voit doté de capacités de prolifération et de différenciation en **plasmocytes**, cellules capables de produire en grande quantités les anticorps complémentaires de l'antigène à l'origine de la réaction. Ces anticorps contribuent à son élimination sur plusieurs plans :

— Mettant à profit leurs deux sites de fixation à l'antigène, ils permettent d'**agglutiner** leurs cibles en formant des ponts entre elles, limitant ainsi leur dissémination dans l'organisme (complexes immuns).
— En recouvrant une cible (**opsonisation**), ils facilitent son internalisation et sa digestion par les cellules phagocytaires (macrophages, cellules dendritiques).
— L'opsonisation est également due à la fixation aux anticorps du **complément**, un complexe protéique circulant par ailleurs capable de perforer la membrane cellulaire d'une cible lorsqu'il est activé (complexe d'attaque membranaire).
— Enfin un anticorps lié est un signal de **dégranulation** reconnu par les lymphocytes NK et les granulocytes, qui relargueront ainsi diverses molécules destinées à détruire la cellule cible ou mettre en place une réaction inflammatoire.

Lors de cette réaction immunitaire, une augmentation modérée et transitoire des lymphocytes circulants (hyperlymphocytose réactionnelle) pourra être observée. Elle s'effacera au bout de quelques semaines par la mort des lymphocytes produits, seuls quelques représentants ayant évolué en **lymphocytes B mémoires**. Ces derniers sont capables de survivre plusieurs années dans l'organisme, et d'initier une réaction plus forte et plus rapide si l'antigène venait à refaire surface (réponse secondaire).

1.3 Lymphopoïèse B

1.3.1 Cellules souches et progéniteurs

Comme l'ensemble des cellules sanguines, les lymphocytes B naissent d'une succession d'étapes de différenciation et de spécialisation débutant au sein de la moelle hématopoïétique localisée dans les os plats et l'épiphyse des os longs. Les **Cellules Souches Hématopoïétiques** (CSH, Figure 1.1), cellules multi-potentes qui prolifèrent dans cet environnement, formeront par différenciations successives les progéniteurs des différentes lignées hématopoïétiques, regroupées dans deux embranchements principaux :

— la *lignée myéloïde*, comprenant les lignées érythrocytaire, mégacaryocytaire, granuleuse et monocytaire, ainsi que les cellules dendritiques.
— la *lignée lymphoïde*, aboutissant aux lymphocytes B, T et NK.

Le maintien des CSH et le contrôle de leur différenciation sont étroitement liés au micro-environnement présent dans la moelle osseuse. Elles sont regroupées au sein de niches spécifiques où elles entretiennent, notamment, des relations étroites avec un ensemble de cellules dites « stromales » (cellules conjonctives, adipeuses, ostéoblastes et ostéoclastes). Parmi les nombreux facteurs susceptibles de retenir les CSH et leur descendance dans la moelle osseuse (*homing*), l'expression de la chimiokine CXCL12 par le stroma et du récepteur associé CXCR4 par les CSH constitue un mécanisme bien décrit, y compris hors de la sphère hématopoïétique [2]. L'engagement de ces CSH vers la lignée lymphoïde B à travers un réseau de progéniteurs intermédiaires est un processus complexe, impliquant notamment l'expression des gènes *FLT3*, *EBF1*, *TCF3* (E2A), *IKZF1* (IKAROS) et *PAX5* (voir pour revue Bryder et Sigvardsson 2010 [3]).

On retrouve parmi cette filiation le **Progéniteur Lymphoïde Commun** (PLC, Figure 1.1), caractérisé par l'expression des protéines membranaires CD34 (comme la quasi-totalité des progéniteurs hématopoïétiques), CD45RA, et CD10 [4]. L'apparition du récepteur IL7R à sa surface constitue l'un des premiers pas vers la lignée

lymphoïde B [4, 5], fournissant à sa descendance un signal de prolifération fort entretenu par l'interleukine 7 (IL7) produite par les cellules stromales.

1.3.2 Stade pro-B

Le stade **pro-B** qui s'en suit constitue l'une des premières étapes de différenciation, caractérisée par l'expression des marqueurs membranaires CD10 et CD19 [1] et la recombinaison du *locus* 14q32.33 codant les chaînes lourdes d'immunoglobulines (*IGH*, Figure 1.1A). Cette manipulation de l'ADN génomique du lymphocyte est l'un des principaux mécanismes à l'origine de la diversité des immunoglobulines, mettant en relation un segment « variable » (V), un segment de « diversité » (D) et un segment de jonction « J » parmi les nombreux présents dans ce *locus*.

Cette **recombinaison VDJ** a lieu en deux temps, fusionnant d'abord un segment D et un segment J, ajoutant ensuite un segment V à ce produit. Ces deux opérations similaires sont l'œuvre des enzymes RAG1 et RAG2, qui reconnaissent aux extrémités de ces segments des motifs nucléotidiques conservés appelés RSS. L'appariement de ces motifs en présence des RAG permet la formation d'une large boucle d'ADN qui sera excisée, chacune des deux cassures double-brins ainsi créée dans le *locus IGH* étant provisoirement réparée par la jonction en épingle à cheveux des deux brins complémentaires. Ces structures provisoires sont ensuite la cible de l'endonucléase Artemis (*DCLRE1C*), qui en ouvrant l'épingle de manière assymétrique est fréquemment la source d'une diversité appelée « P ». L'enzyme TdT est également recrutée sur ces sites, et ajoute de manière aléatoire jusqu'à 20 nucléotides aux extrémités ainsi libérées, ajoutant encore un degré de diversité appelé « N ». Le complexe protéique DNA-PK, membre du système NHEJ impliqué dans la réparation des cassures double-brins, assure enfin le raboutage des deux extrémités ainsi modifiées.

Outre la formation d'un exon à la séquence fortement aléatoire, ces deux réarrangements successifs permettent de rapprocher du promoteur propre à chaque segment V une séquence régulatrice de l'expression située en 3' des segments J, et donc d'augmenter l'expression du *locus* IGH. Si la recombinaison et les nombreux ajouts de nucléotides n'ont pas altéré le cadre de lecture ou conduit à l'apparition d'un codon STOP, la chaîne lourde ainsi produite pourra se combiner aux protéines V pré-B et λ5 pour former un **pré-BCR** à la surface de la cellule (Figure 1.1B). L'expression de ce pré-BCR conduira au repliement de la chromatine au niveau du second allèle non-réarrangé du *locus IGH*, qui sera ainsi conservé intact et évitera à la cellule d'exprimer deux immunoglobulines différentes (exclusion allélique). Dans le cas contraire, une nouvelle recombinaison VDJ sera tentée sur ce second allèle et fournira à la cellule une deuxième chance de passer ce point de contrôle indispensable à sa survie. La

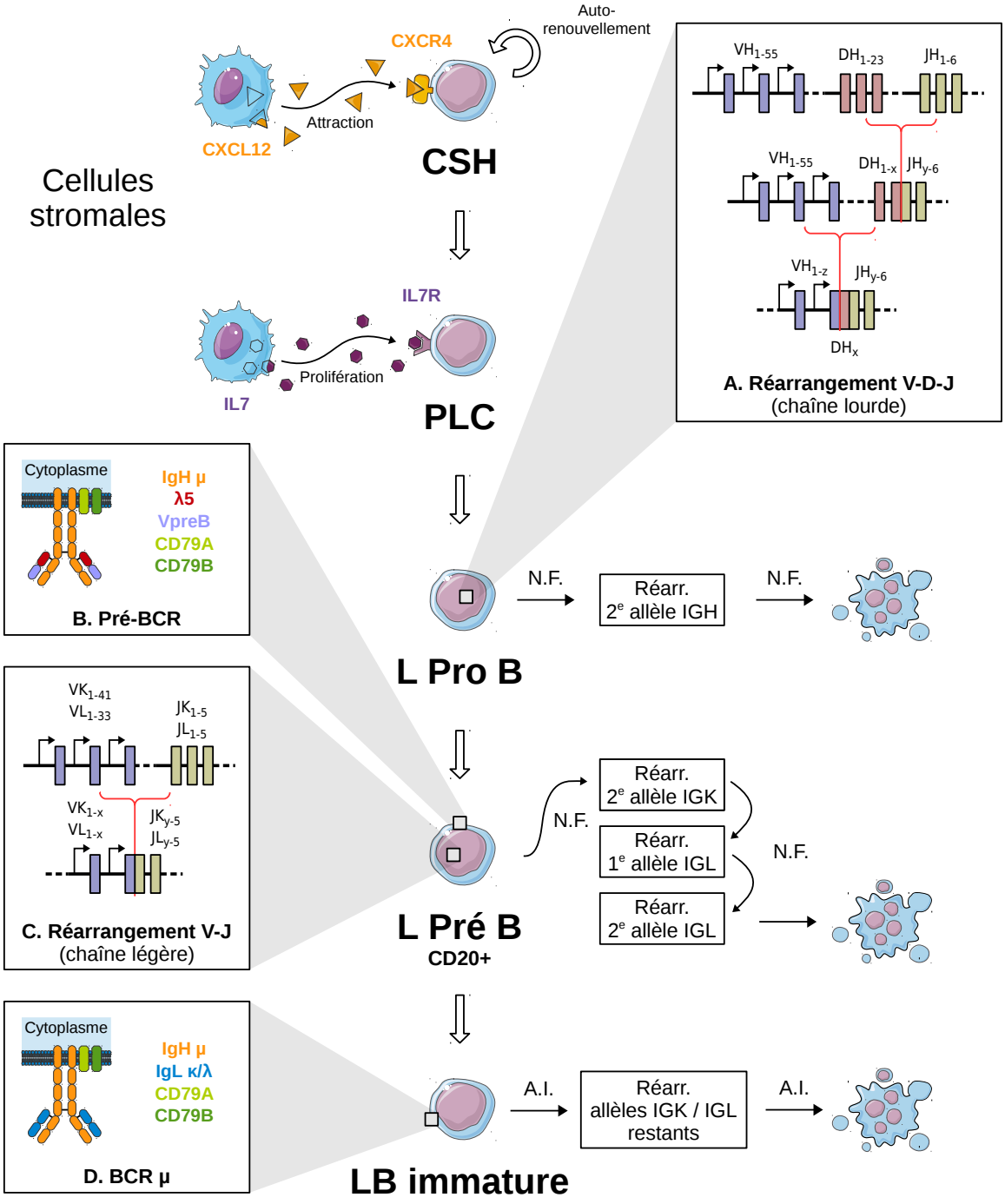


Figure 1.1 – Lymphopoïèse B dans la moelle osseuse

Les principales étapes de la différenciation des lymphocytes B dans la moelle osseuse sont présentées en vis-à-vis des mécanismes et produits conditionnant leur différenciation (panneaux A à D). Réarr. : Réarrangement, N.F. : en cas de réarrangement Non Fonctionnel, A.I. : en cas de récepteur Auto-Immun. Réalisé à partir du kit *Servier Medical Art*, sous licence Creative Common BY.

région variable issue du réarrangement VDJ sera exprimée avec la première région constante située en 3', conduisant à la production d'une chaîne lourde de type µ.

1.3.3 Stade pré-B

L'expression d'un pré-BCR fonctionnel marque le passage au stade **pré-B**, qui se caractérise par l'apparition du marqueur membranaire CD20 et par le réarrangement d'un ou plusieurs *loci* codant pour les chaînes légères d'immunoglobulines (*IGK* en 2p12, *IGL* en 22q11.2). Cette **recombinaison VJ** (1.1C) repose sur les mêmes mécanismes biologiques que la recombinaison VDJ des chaînes lourdes, mais conduit toutefois à une diversité plus modérée. En effet l'absence de segment D dans les *loci* codant les chaînes légères limite le nombre de combinaisons possibles, et l'enzyme TdT source de la N diversité nucléotidique n'est plus active à ce stade. L'existence d'un *locus* secondaire IGL offre également à la cellule deux réarrangements supplémentaires pouvant encore être tentés si les deux allèles du *locus* IGK sont recombinés de manière incorrecte. Si les recombinaisons successives des 4 allèles conduisent à l'expression d'une chaîne légère fonctionnelle, celle-ci pourra remplacer la pseudo chaîne légère du pré-BCR afin de fournir à la cellule un **BCR** complet (1.1D), et mettre fin au processus (exclusion isotypique). Les éventuels allèles non réarrangés resteront en configuration germinale et ne seront pas exprimés (exclusion allélique). Selon l'historique de la cellule, ce BCR pourra présenter des chaînes légères κ (2/3) ou λ (1/3), qui ne présentent pas de différence fonctionnelle majeure. Il s'agit là encore d'un point de contrôle déterminant le devenir de la cellule, les lymphocytes B ne parvenant pas à produire un BCR fonctionnel étant sujets à l'apoptose.

1.3.4 Stade immature

Ces lymphocytes B présentant un BCR fonctionnel correspondent au stade **immature**, dont la réactivité est mise à l'épreuve. En effet si ces lymphocytes viennent à rencontrer un antigène dans la moelle osseuse (vraisemblablement auto-immun), un phénomène d'édition du BCR s'apparentant à un retour au stade pré-B est engagé. Suite à la réactivation des enzymes RAG1 et RAG2, l'allèle fonctionnel peut subir un second réarrangement entre un V situé en 5' de la jonction VJ et un J en 3', offrant une nouvelle chance au lymphocyte d'exprimer une chaîne légère fonctionnelle et non-auto-immune. Comme au stade pré-B, les allèles non-réarrangés peuvent être engagés successivement si le réarrangement devait être non-fonctionnel, et l'apoptose constitue la dernière option envisagée. Les lymphocytes B immatures passant ce dernier point de contrôle cessent provisoirement d'exprimer le récepteur CXCR4 qui les confine dans la moelle osseuse, et gagnent la rate via la circulation sanguine.

1.3.5 Stade mature

La **rate** constitue un carrefour majeur dans la circulation sanguine, où se déroulent les réactions immunitaires aux corps étrangers véhiculés par le sang et l'élimination des cellules sanguines sénescentes. C'est également dans cet organe que les lymphocytes B immatures produits par la moelle osseuse achèveront leur maturation, en passant par plusieurs états transitoires [6]. Le stade T1 correspond ainsi aux lymphocytes arrivés dans la pulpe rouge de la rate, où ils rencontrent de nombreux macrophages à même d'éliminer les lymphocytes B bloqués dans leur différenciation par un réarrangement non-fonctionnel des immunoglobulines ou rendus anergiques par un auto-antigène. La migration de ces lymphocytes B T1 vers les follicules spléniques ou leurs zones marginales (pulpe blanche) s'accompagne d'un passage vers le stade T2, qui préfigurera leur spécialisation une fois matures (lymphocytes B folliculaires ou lymphocytes B de la zone marginale).

Cette transition se caractérise par l'expression pour les **lymphocytes B folliculaires** d'un second BCR de phénotype δ, en plus du BCR μ déjà présent. Cette dualité du BCR est permise par la proximité des parties constantes μ et δ dans le *locus IGH*, et la production d'un transcrit qui retiendra l'une ou l'autre des parties constantes par épissage alternatif, présentant ainsi la même spécificité pour l'antigène. En quittant la rate pour gagner la circulation sanguine, ces lymphocytes B folliculaires ayant survécu aux nombreux points de contrôle décrits sont dits **matures naïfs**, et sont prêts à initier une réaction immunitaire.

Le potentiel de migration de la minorité de **lymphocytes B de la zone marginale** produits reste quant à lui plus controversé [7]. Leur devenir et les mécanismes d'immunité adaptative indépendante des lymphocytes T auxquels ils participent [1] ne seront pas développés ici, les LDGCB étant issus des lymphocytes B folliculaires.

1.4 Immunopoïèse B

1.4.1 Follicules primaires

En quittant la rate, les lymphocytes B folliculaires matures naïfs pourront être recrutés au sein du système lymphatique, composé d'un ensemble de ganglions reliés en série par des canaux afférents et efférents (Figure 1.2). Le système lymphatique transporte la lymphe, qui est issue du drainage du liquide interstitiel de l'ensemble des tissus de l'organisme, piégeant de manière efficace d'éventuels agents infectieux. La vascularisation des ganglions lymphatiques comprend, entre autres, des

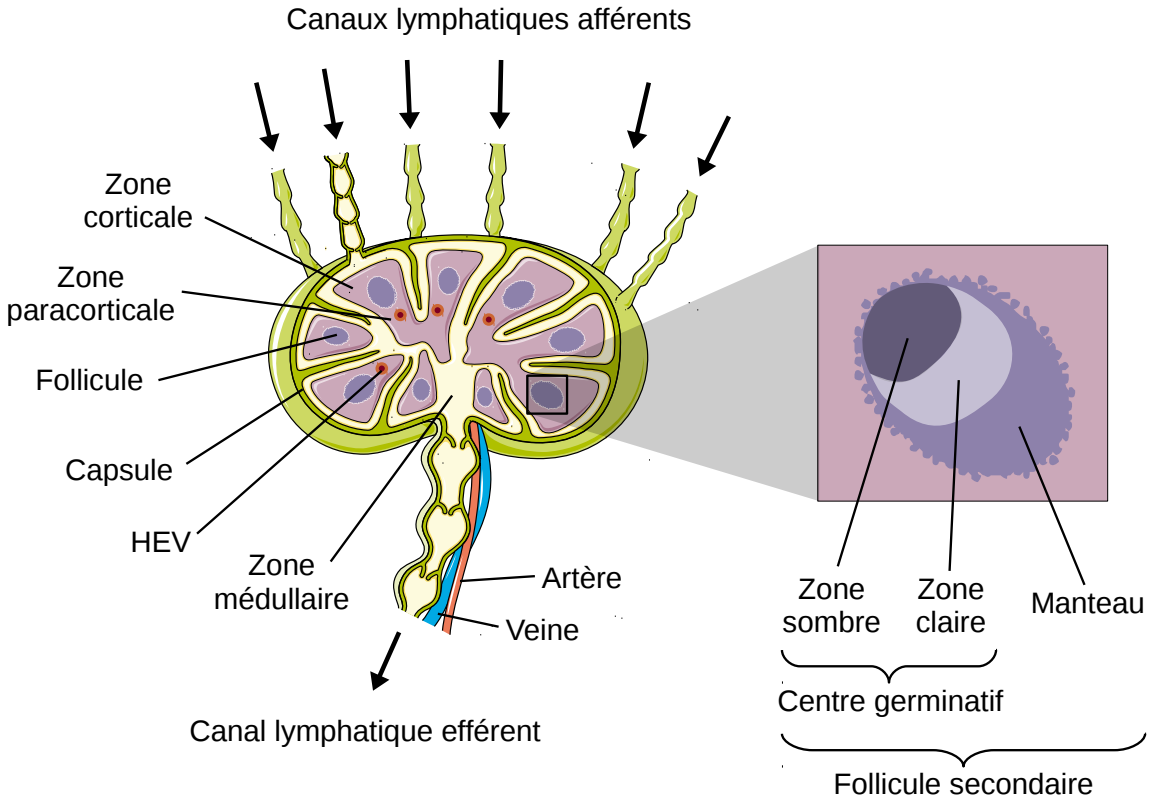


Figure 1.2 – Organisation d'un ganglion lymphatique

Réalisé à partir du kit *Servier Medical Art*, sous licence Creative Common BY.

veinules à endothélium épais (HEV). Cet endothélium particulier a la propriété d'attirer puis de fixer les lymphocytes B naïfs du sang périphérique, par l'entremise de sélectines (*SELL*) et intégrines (*LFA1*). Après transmigration à travers l'endothélium des HEV, les lymphocytes B folliculaires naïfs sont regroupés au sein de la zone corticale en amas cellulaires appelés **follicules primaires** (Figure 1.3A). Cette migration est permise par la reprise de l'expression du récepteur chimiotactique CXCR4 et de son parent CXCR5, permettant au lymphocyte de répondre aux chimiokines CXCL12 et CXCL13 produites par les cellules folliculaires dendritiques présentes sur place.

Cette position au cœur du système lymphatique maximise les chances de rencontre entre un lymphocyte B naïf et l'antigène correspondant. Ce dernier peut être rencontré sous sa forme soluble, drainée par la circulation lymphatique, ou présenté à la surface des cellules dendritiques ayant préalablement internalisé et digéré un corps étranger. Si les ganglions drainant les antigènes rencontrés au niveau de la peau constituent une grande majorité des organes lymphoïdes secondaires où se déroule cette réaction, des systèmes similaires existent au niveau des muqueuses (MALT) et de la circulation sanguine (follicules de la rate).

1.4.2 Rencontre avec l'antigène

La **reconnaissance de l'antigène** par plusieurs BCR membranaires d'un même lymphocyte B folliculaire naïf entraine leur fusion et l'activation subséquente d'une séries de réactions intra-cellulaires. Les BCR ayant réagi avec l'antigène sont internalisés par endocytose, conduisant à la digestion de l'antigène et sa présentation en surface via le MHC de classe II (Figure 1.3B). Les motifs ITAM des protéines CD79A et CD79B associées au BCR se trouvent phosphorylés par les tyrosine kinases de la famille Src (*LYN*, *SYN*, *BLK*), conduisant à l'activation des voies MAP kinase (via Ras) et NFκB (via Syk, Btk et PKC-β), entre autres. Ces différentes voies métaboliques contribuent à relancer la prolifération de ces cellules, décupler leurs capacités de survie et engager leur différenciation. Le récepteur chimiotactique CXCR5 est enfin remplacé par le récepteur CCR7, diminuant l'attraction par la chimiokine CXCL13 produite au sein du follicule pour attirer le lymphocyte B vers la région paracorticale riche en lymphocytes T. Par ailleurs l'antigène peut avoir été reconnu par des anticorps circulants et les protéines du **complément** durant son parcours à travers l'organisme, offrant ainsi aux lymphocytes B folliculaires un deuxième signal à reconnaitre : la protéine C3d. Cette fraction du complément est reconnue par le récepteur membranaire CR2, dont les partenaires CD19 et CD81 sont à même de potentialiser l'activation de la voie du BCR par l'antigène, via le recrutement de LYN sur CD79A/CD79B et via l'activation de la voie PI3 kinase.

L'antigène est reconnu indépendamment par les **lymphocytes T auxiliaires** (CD4+), dans le paracortex ganglionnaire (Figure 1.3C). De manière assez similaire au BCR des lymphocytes B, chacun de ces lymphocytes T exprime un TCR membranaire différent à même de reconnaitre un antigène bien spécifique, qui devra lui être présenté via le MHC de classe II des cellules dendritiques. En miroir de la migration des lymphocytes B vers la zone T, le lymphocyte T CD4+ reconnaissant son antigène cible est ensuite attiré vers le follicule B par le remplacement du récepteur chimiotactique CCR7 par CXCR5. Cette activation du lymphocyte T auxiliaire est potentialisée en cas de reconnaissance d'un pathogène par des mécanismes parallèles moins spécifiques (TLR), qui conduisent à l'expression par la cellule dendritique des protéines B7 (CD80, CD86) également reconnues par le lymphocyte T.

1.4.3 Activation par les lymphocytes T

La rencontre entre les lymphocytes B et T activés indépendamment par deux exemplaires d'un même antigène a lieu à proximité des follicules (Figure 1.3D). Un double pontage se met alors en place entre les deux lymphocytes, le récepteur CD40

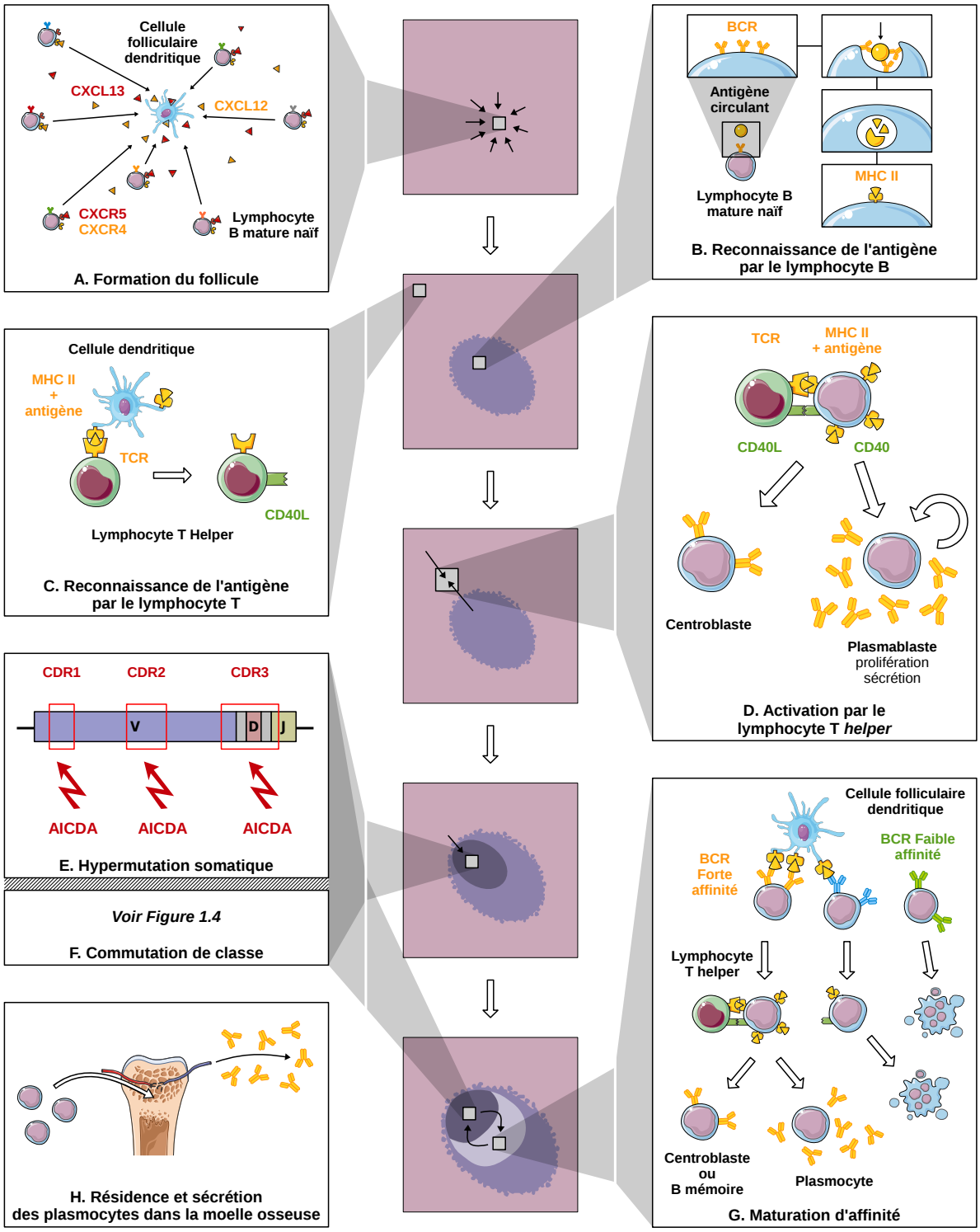


Figure 1.3 – Déroulement de la réaction immune adaptative

Les principales étapes de l'immunopoïèse dans un ganglion lymphatique sont présentées en vis-à-vis des principaux mécanismes impliqués (panneaux A à H). Réalisé à partir du kit *Servier Medical Art*, sous licence Creative Common BY.

du lymphocyte B interagissant avec son ligand CD40L exprimé à la surface du lymphocyte T, et l'antigène présenté par le MHC de classe II du lymphocyte B interagissant avec le TCR. Il en résulte l'activation de diverses cascades enzymatiques au sein du lymphocyte B, aboutissant à accentuer les phénomènes de prolifération, survie et différenciation initiés précédemment par le BCR (voies NFκB et AP-1, expression de *MYC*...). La production très localisée de cytokines par le lymphocyte T (IL2, IL4, IL21) contribue également à potentialiser ces diverses réactions.

La prolifération locale occasionnée par l'activation via le lymphocyte T conduit à la formation d'un petit groupe extra-folliculaire de **plasmablastes**, produisant des IgM circulantes de faible spécificité (réponse immune primaire). Ces plasmablastes iront se loger dans les extensions de la zone médullaire (Figure 1.2), et assureront pendant une courte période leur rôle de sécrétion. Les anticorps précoces qu'ils produisent assurent la complexation des antigènes circulants, qui pourront ainsi être captés par les cellules folliculaires dendritiques et présentés aux lymphocytes B durant les prochaines phases de leur différenciation.

1.4.4 Formation du centre germinatif

Après leur activation par un lymphocyte T auxiliaires correspondant, les lymphocytes B activés non-impliqués dans la différenciation plasmablastique seront de nouveau attirés au sein du follicule, pour former un **centre germinatif** (Figure 1.2). Les **centroblastes** issus de leur prolifération forment la « zone sombre » autour de son point de réentrée dans le follicule, tandis que la différenciation de ces centroblastes en centrocytes conduit à l'établissement d'une « zone claire » également riche en cellules folliculaires dendritiques et lymphocytes T auxiliaires. Les nombreux autres lymphocytes B naïfs qui formaient le follicule primaire se retrouvent expulsés en périphérie du follicule, formant la « zone du manteau » de ce qu'il convient désormais d'appeler un follicule secondaire. Ce centre germinatif sera le siège de deux mécanismes majeurs conduisant à la spécialisation des lymphocytes : l'hypermutation somatique et la commutation isotypique.

L'expression du gène *BCL6* par les lymphocytes B et T est un point de contrôle majeur dans la formation du centre germinatif, initié par l'interaction CD40 / CD40L [8]. Elle conditionne l'entrée des lymphocytes activés dans le follicule, l'expression des récepteurs chimiotactiques nécessaires à l'établissement dans la zone sombre et la prolifération des centroblastes dans cette même zone. L'expression de nombreux gènes assurant la cohérence du centre germinatif contrôlent ou sont contrôlés par *BCL6* (*MEF2B*, *MEF2C*, *IRF4*...), définissant un réseau de régulation transcriptionnelle dont *BCL6* est le noyau. Des modifications épigénétiques, modulées par les protéines

EZH2 et EP300 entre autres, sont également connus pour limiter les effets transcriptionnels de *BCL6* à certaines de ses cibles [8].

1.4.5 Maturation d'affinité

La cascade enzymatique déclenchée via CD40 par un lymphocyte T auxiliaires induit également la mise en place d'un processus de mutation dirigée appelé **hypermutation somatique** (SHM), prenant place dans la zone sombre du centre germinatif (Figure 1.3E). **AICDA** catalyse en effet la désamination des cytosines en uraciles, bases de l'ARN non physiologiques dans l'ADN qui sont excisées par la glycosylase UDG pour donner des sites abasiques. Ces sites sont ensuite clivés par APEX1, formant des cassures simple-brins dont la réparation par une ADN polymérase peu fidèle conduit à leur remplacement par des bases aléatoires, et donc à des mutations ponctuelles. Ce processus touche préférentiellement les *loci* codant les chaînes lourdes et légères d'immunoglobulines, à un taux d'environ une mutation par cycle cellulaire.

La zone claire du centre germinatif constitue une arène où la descendance du lymphocyte B activé par l'antigène s'affronte, afin de sélectionner les lymphocytes les plus affins pour l'antigène (Figure 1.3G). Cette migration de la zone sombre vers la zone claire marque le passage du stade centroblaste au stade **centrocyte**, également caractérisé par la perte provisoire des capacités prolifératives du lymphocyte B. Les nombreuses cellules folliculaires dendritiques qui s'y trouvent présentent l'antigène, que les centrocytes devront reconnaître pour survivre. Les centrocytes les plus aptes à capter l'antigène sont ainsi ceux qui en présentent la plus grande densité à leur surface via le MHC de classe II, monopolisant les rares lymphocytes T auxiliaires présents dans le centre germinatif aptes à leur fournir via leur CD40L les signaux nécessaires à leur survie [8]. Ces signaux induisent pour le centrocyte un retour au stade centroblastique prolifératif dans la zone sombre, améliorant ainsi l'affinité moyenne du pool de centroblastes en transmettant son génome muté.

Cette sélection sur l'affinité de l'immunoglobuline à l'antigène explique par ailleurs que la SHM s'observe principalement dans les 3 régions CDR des immunoglobulines, en contact direct avec l'antigène, tout en laissant les régions *framework* relativement intactes. Après plusieurs aller-retours entre les zones claire et sombre, certains centrocytes choisiront de terminer leur différenciation, soit en plasmocytes sécrétant les anticorps, soit en lymphocytes B mémoires.

1.4.6 Commutation isotypique

Tout au long de leur parcours au sein du centre germinatif, les lymphocyte B peuvent être amenés à **commuter l'isotype** de leur BCR, qui n'incluait jusqu'à

présent que des chaînes lourdes μ ou δ. Ce second réarrangement du *locus IGH* est contrôlé par diverses cytokines exprimées par le micro-environnement, et permettent de spécialiser la réponse immunitaire en exprimant des anticorps légèrement différents. Les IgM produits par défaut sont ainsi tout particulièrement aptes à réagir aux infections bactériennes, en provoquant l'opsonisation via le système du complément et l'agrégation via la formation de structures pentamériques. La production d'INF-γ par les lymphocytes T_H1 en réponse à une infection virale conduit à la production d'IgG, activant les différents systèmes du complément et l'opsonisation. Les infections parasitaires sont à l'origine d'une commutation en IgE en réponse à l'IL4, ces anticorps étant plutôt spécialisés dans les processus de dégranulation et d'inflammation médiées par les granulocytes. L'activation d'un lymphocyte B dans les MALT conduit enfin de manière préférentielle à une commutation en IgA, dimères d'anticorps dont le principal mode d'action est la rétention physique des antigènes dans la muqueuse, limitant ainsi leur diffusion dans l'organisme.

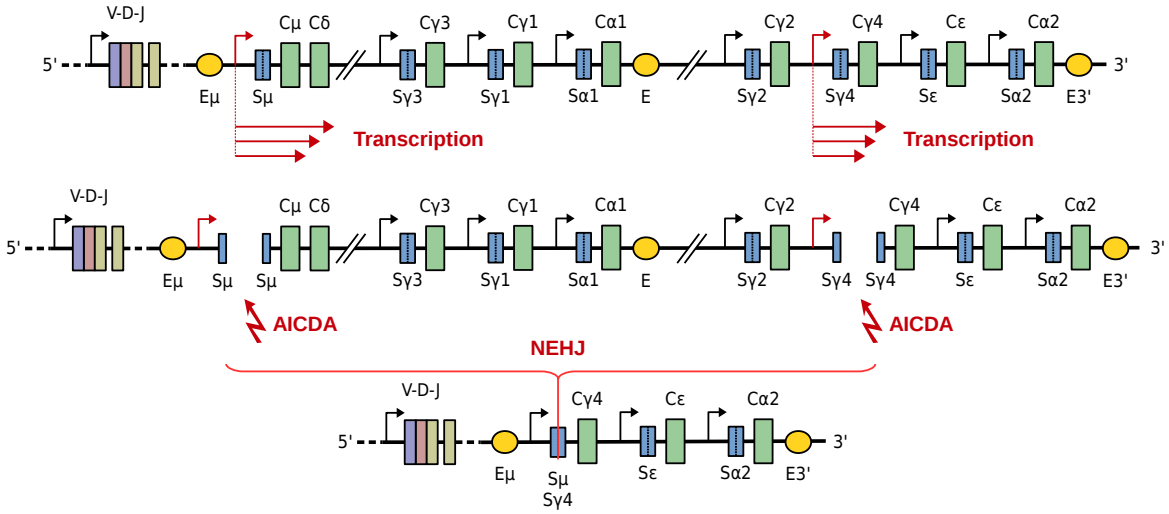


Figure 1.4 – Commutation isotypique

L'exemple de la commutation d'une IgM en IgG4 est donné afin d'illustrer la commutation isotypique. Le VDJ réarrangé est présenté en 5', sans insister sur les segments V et J non réarrangés toujours présents en 5' et 3' respectivement. Les différentes régions de *switch* (S) sont représentées en bleu, les parties constantes (C) elles-mêmes composées de plusieurs exons non représentés sont visibles en vert et les différentes régions régulatrices de la transcription (*enhancers*, E) en jaune. La transcription préférentielle de deux régions S+C favorise l'action d'AICDA, qui induit la formation de cassures double-brin dans le *locus IGH*. La réparation de ces cassures par le système de jonction des extrémités non-homologues (NHEJ) conduit au raboutage des deux segments restants, et à la mise en vis à vis du VDJ et d'une nouvelle constante pour l'immunoglobuline qui résultera de la transcription de ce *locus*.

D'un point de vue mécanistique, cette commutation isotypique est le fruit de la réparation de deux cassures double-brins causées dans le *locus IGH* causées là encore par par l'enzyme **AICDA** (Figure 1.4). Dans ce second modèle, les cassures simple-

brins ne sont pas réparées et évoluent en cassure double-brins. La spécificité de ce mécanisme est assurée par la nécessité pour AICDA d'agir sur un ADN simple-brin riche en GC présentant un certain motif nucléotidique, tel que dans les régions *switch*. Le désappariement des deux brins d'ADN est assuré par la transcription d'une des régions *switch*, qui conduit à l'hybridation du transcrit à l'un des deux brins d'ADN matrice tout en laissant le second brin d'ADN exposé à l'action d'AICDA (boucle R). La réponse à différentes cytokines extra-cellulaires permet ainsi d'influencer la commutation vers l'un ou l'autre des isotypes, en déclenchant la transcription du *switch* correspondant. Les deux cassures survenues dans deux régions *switch* sont enfin réparées par les mécanismes de réparation non-homologue (NHEJ), conduisant à la juxtaposition de l'exon VDJ et d'une constante autre que μ ou δ.

1.4.7 Différenciation terminale

Les différents signaux reçus dans le centre germinatif, notamment les cytokines IL2, IL4 et IL6, modifient progressivement la balance entre immunoglobulines membranaires et sécrétées, faisant des centrocytes à dominante membranaire des **plasmocytes** à dominante sécrétée. Cette variation s'opère par modification de l'épissage alternatif des ARN produits lors de la transcription du *locus IGH*, privilégiant un codon STOP et un site de polyadénylation plus précoces pour former un transcrit n'incluant pas les domaines transmembranaires et cytoplasmiques codés par les deux derniers exons. Cette différenciation passe par l'expression au sein du lymphocyte B des facteurs de transcription *PRDM1* et *IRF4*, et l'inhibition de l'expression du *BCL6* caractéristique du centre germinatif qui en découle. Les plasmocytes seront enfin relocalisés dans la moelle osseuse, où ils se fixeront définitivement pour assurer leur rôle sécrétoire d'anticorps circulants pendant plusieurs mois (Figure 1.3H).

Une partie des centrocytes sélectionnés pour leur affinité à l'antigène se différencient en **lymphocytes B mémoires**, dont le rôle s'apparente aux lymphocytes B matures naïfs. Contrairement aux plasmocytes, ces lymphocytes activés ne produisent que très peu d'anticorps et continuent leur circulation entre les différents organes lymphoïdes secondaires pendant des années, en vue d'établir une nouvelle réaction immunitaire. Ayant subi commutation isotypique et maturation d'affinité, ces lymphocytes sont extrêmement compétents à reconnaitre l'antigène s'il devait refaire surface dans l'organisme, ce qui peut expliquer l'intensité et l'efficacité de la réponse immune secondaire.

1.5 Classification des hémopathies malignes

1.5.1 Systèmes de classification

La grande variété des proliférations pathologiques d'origine hématopoïétique a donné lieu à de nombreuses classifications au fil des années, généralement adossées à des visions très personnelles et souvent contradictoires des processus en jeu. Établir une classification exhaustive et consensuelle de ces hémopathies est toutefois un objectif autrement plus ambitieux, que plusieurs consortiums se sont assigné au fil du temps. L'établissement de la *working formulation* en 1982 [9] représente la première initiative de ce type ayant eut un impact notable, faisant suite à la classification de Kiel jusqu'alors relativement répandue en Europe [10]. Si ces premières classifications se bornaient à des critères purement morphologiques, l'arrivée des marquages immunologiques et des techniques de biologie moléculaire a permis d'améliorer sensiblement ces classifications, menant à la révision de classification de Kiel [11] et à l'initiative européo-américaine REAL [12]. Cette tâche a finalement été reprise par l'OMS, organisme le plus à même de fournir un consensus international, qui publie une édition révisée de sa classification sur une base régulière depuis 2000 [13, 14]. Si la classification OMS offre une définition précise pour de nombreuses entités (dont certaines encore provisionnelles), nous nous contenterons d'évoquer ici les hémopathies les plus courantes en terme d'incidence (Figure 1.5).

Les hémopathies malignes sont traditionnellement distinguées par la lignée cellulaire dont la cellule tumorale est issue. On distingue ainsi les hémopathies d'origine **lymphoïde** (lymphomes, syndromes lymphoprolifératifs et leucémies aiguës lymphoblastiques), des hémopathies **myéloïdes** (leucémies aiguës myéloïdes, syndromes myéloprolifératifs ou myélodysplasiques). Le versant lymphoïde, sur lequel nous nous attarderons dans cette introduction, regroupe une grande variété de tumeurs caractérisées par leur ressemblance à des lymphocytes normaux plus ou moins différenciés (§1.3, §1.4).

1.5.2 Progéniteurs B

La **LAL** correspond ainsi au stade cellulaire le plus précoce, impliquant des lymphocytes immatures tels que trouvés dans la moelle osseuse (lymphoblaste). Si le stade de différenciation peut varier d'un patient à l'autre, la présence d'un réarrangement DJ du *locus IGH* dans une majorité des cas suggère généralement un stade pro-B ou postérieur [14]. Ce blocage dans la différenciation est associé à un fort potentiel de prolifération, les cellules tumorales remplaçant rapidement la moelle hématopoïétique normale. En conséquence, ces hémopathies se caractérisent par une

Pathologie	Acronyme	Nombre de cas en France (2012)	
Lymphomes non hodgkiniens	**LNH**	**21 175**	**62,9%**
Myélome Multiple & plasmocytome	MM	4 888	23,1%
Leucémie Lymphoïde Chronique Lymphome Lymphocytique	LLC	4 464	21,1%
Lymphome Diffus à Grandes Cellules B	LDGCB	4 096	19,3%
Lymphome Folliculaire	LF	2 530	11,9%
Lymphome de la Zone Marginale	LZM	1 772	8,4%
Lymphome T/NK à cellules matures	LT / LNK	1 419	6,7%
Lymphome Lymphoplasmocytaire Macroglobulinémie de Waldenström	MW	1 247	5,9%
Lymphome à Cellules du Manteau	LCM	659	3,1%
Lymphome de Burkitt	LB	~100	0,5%
Syndromes myéloprolifératifs		**5 756**	**17,1%**
Leucémie Aiguë Myéloïde	LAM	2 984	51,8%
Leucémie Myéloïde Chronique	LMC	807	14,0%
Autres syndromes myéloprolifératifs chroniques	-	1 965	34,1%
Syndromes myélodysplasiques	**SMD**	**4 059**	**12,1%**
Lymphome de Hodgkin classique	**LH**	**1 880**	**5,6%**
Hémopathies à cellules immatures		**810**	**2,4%**
Leucémie Aiguë Lymphoblastique Lymphome Lymphoblastique (B, T ou SAI*)	LAL LLB / LLT	810	100,0%
Tous cancers confondus		***389 034***	
Prostate		***56 841***	
Sein		***48 763***	
Côlon / Rectum		***42 152***	
Poumon		***39 495***	
Hémopathies malignes		***33 680***	

Figure 1.5 – Incidence des principales hémopathies malignes

L'incidence de chaque hémopathie est présentée en nombre de nouveaux cas répertoriés en France en 2012 [15], et en proportion de sa catégorie (fond gris). Pour ces catégories, la proportion présentée est calculée par rapport au total des hémopathies malignes. Les cancers solides avec une incidence plus élevée sont également présentés à titre de comparaison.

raréfaction des cellules normales du sang périphérique (anémie, neutropénie, thrombopénie) et par la présence de cellules immatures tumorales circulantes, appelées « blastes » (> 20%). Les signes cliniques sont en rapport avec l'insuffisance médullaire et peuvent associer un syndrome anémique (pâleur, fatigue essoufflement, tachycardie), des infections ou un syndrome hémorragique. Ces hémopathies peuvent parfois se manifester cliniquement sous une forme tumorale (généralement une adénopathie progressive) alors que les paramètres de l'hémogramme sont conservés. Ces présentations cliniques particulières sont appelées « lymphomes lymphoblastiques ». En terme de prévalence, ces deux formes touchent préférentiellement des patients jeunes (75% < 6 ans), et restent toutefois rares par rapports aux autres formes de leucémies (LLC, LAM, LMC). D'un point de vue génétique, diverses translocations définissent des sous-types plus ou moins fréquents selon les classes d'âges : t(9;22) *BCR-ABL1* chez l'adulte (25%), t(12;21) *ETV6-RUNX1* chez l'enfant, réarrangements de *MLL* chez les enfants en bas âge...

1.5.3 Lymphocytes B naïfs

Les **LCM** sont issus de la prolifération de cellules du manteau, lymphocytes B matures naïfs situés autour des follicules ganglionnaires. Ils présentent à ce titre un taux élevé d'immunoglobulines de surface monoclonales, dont les mutations somatiques sont généralement absentes, parfois modérées (15% à 40% des cas). La translocation t(11;14) impliquant *IGH* et *CCND1* est considérée comme l'événement déclencheur de l'oncogénèse [14], menant à une sur-expression de *CCND1* typique de ces lymphomes. Cette anomalie est également supplémenté d'une liste conséquente d'événements secondaires récurrents (gains en 3q26 et 8q24, pertes en 6q, 9p21, 11q22-q23, 13q, 17p...). Il s'agit d'une pathologie généralement incurable, avec une médiane de survie comprise entre 3 et 5 ans.

1.5.4 Lymphocytes B activés par l'antigène

Les **lymphomes folliculaires** (LF) impliquent indistinctement des centrocytes et centroblastes (exprimant entre autres CD20, CD10 et *BCL6*), désagrégeant la structuration du centre germinatif normal. Ce blocage dans le centre germinatif implique l'expression d'un BCR continuellement la cible de la SHM, présentant des mutations importantes. Il s'agit d'un lymphome indolent, pouvant s'étendre de manière asymptomatique à de nombreux ganglions, à la rate et à la moelle osseuse avant d'être découvert. Si les méthodes de stratification pronostique sont encore sujettes à débat, il est possible de définir 3 grades anatomopathologiques de LF en fonction de la densité de centroblastes observés dans les follicules tumoraux, dont le stade ultime

(IIIB) s'apparente au LDGCB [14]. Si une rémission durable est rarement obtenue, il s'agit néanmoins d'une maladie indolente dont la survie médiane est supérieure à 8 ans quel que soit le grade. La translocation t(14;18) entre *IGH* et *BCL2* est retrouvée dans plus de 80% de ces lymphomes, caractéristique mais non suffisante à l'établissement de l'oncogénèse. La transformation en LDGCB (préférentiellement GCB) est un phénomène relativement courant, de mauvais pronostic, qui reste l'objet de nombreuses études.

Les lymphomes de **Hodgkin** (LH) représentent une entité à part, caractérisée par de larges cellules tumorales polynucléées disséminées (cellules de Reed-Sternberg). Majoritairement issus de la lignée lymphoïde B, ces tumeurs présentent une empreinte mutationnelle figée évocatrice du centre germinatif [14]. Ces tumeurs se caractérisent par l'expression des marqueurs CD30 et CD15, et par des amplifications des bras 2p, 9p (particulièrement 9p23-p24) et 12q. D'un point de vue clinique, le LH se caractérise par un tumeur de localisation préférentiellement cervicale (75%) ou médiastinale, avec un pronostic plutôt encourageant (survie à 5 ans > 90%). L'infection par le VIH et l'EBV semblent prédisposer à cette pathologie, qui touche majoritairement des patients de 15 à 35 ans ou les personnes âgées.

Le lymphome de **Burkitt** (LB) est une tumeur issue de la prolifération de lymphocytes B ayant au moins atteint le centre germinatif, comme la présence d'une immunoglobuline clonale et mutée peut le suggérer [14]. Elle se caractérise par une vitesse de développement impressionnante (doublement de la population tumorale en 1 à 2 jours), aisément discernable en histologie par la présence de nombreuses cellules en division (Ki67) et en apoptose. On distingue les formes endémiques, particulièrement fréquentes chez les enfants de 4 à 7 ans dans plusieurs régions d'Afrique, des rares formes sporadiques retrouvées en Europe et ailleurs. Ces deux formes se traduisent par l'apparition d'une masse tumorale importante et évolutive, préférentiellement au niveau du visage pour la forme endémique (50% des cas) et de l'abdomen pour la forme sporadique. L'étiologie de ces tumeurs reste à préciser, mais des liens ont pu être établis avec des virus tels que l'EBV et le VIH, ainsi qu'avec le paludisme [14]. La translocation t(8;14) impliquant *MYC* et *IGH* est emblématique de ces tumeurs, avec la particularité d'impliquer des mécanismes de cassure différents entre formes endémiques et sporadiques (SHM et *switch* respectivement, §3.1.2). La division rapide de ces cellules tumorales les rendent particulièrement sensibles aux chimiothérapies, faisant de ce lymphome une maladie agressive mais curable, pour peu que le syndrome de lyse tumorale soit maitrisé.

Les **LZM**, dont les lymphomes des **MALT** sont une sous-catégorie, correspondent également à un stade postérieur au centre germinatif, non plus pour des lymphocytes folliculaires mais pour des lymphocytes de la zone marginale. Ce sont des lymphomes

indolents (survies à 5 ans de l'ordre de 60% à 80%) pouvant évoluer en LDGCB plus agressifs. Ils peuvent présenter une différenciation plasmocytaire tumorale marquée rendant le diagnostic parfois délicat, et présentent peu de spécificités génétiques à l'exception des translocations impliquant *MALT1* (18q21) dans les lymphomes des MALT. Si les LZM se développent principalement dans les ganglions lymphatiques, les lymphomes des MALT sont restreints aux régions anatomiques riches en muqueuses (système gastro-intestinal, poumons, sphère ORL et contours de l'œil). Ces derniers présentent une interaction forte avec diverses infections chroniques, par des bactéries telles que *H. pylori* dans les MALT gastriques.

La **LLC** consiste en une multiplication anormalement élevée lymphocytes B matures activés relativement normaux d'un point de vue morphologique, envahissant le sang périphérique, la moelle, la rate et les ganglions. Ces cellules tumorales présentent un niveau modeste d'immunoglobulines de surface, caractérisées par un réarrangement préférentiel de certains segments VH d'une part, et par l'empreinte mutationnelle de la SHM dans 50% à 60% des cas d'autre part [14]. La LLC est la plus fréquente des leucémies de l'adulte, avec un âge moyen au diagnostic de 65 ans. Si les symptômes sont généralement relativement modestes, 2% à 8% des LLC évoluent en LDGCB, avec des implications cliniques plus lourdes (syndrome de Richter). Les anomalies génétiques les plus fréquentes sont la délétion 13q14.3 (50%), la trisomie 12 (20%), et les délétions situées en 11q et 17p (<10%).

1.5.5 Lymphocytes B différenciés

Les lymphomes lymphoplasmocytaires, dont la **macroglobulinémie de Waldenström** (MW) est la variante la plus répandue, correspondent à un stade de différenciation intermédiaire entre le centre germinatif et le plasmocyte [14]. Ils se caractérisent par une population tumorale hétérogène plus ou moins proche du plasmocyte, résidant dans la moelle osseuse, qui secrète dans la circulation sanguine une quantité importante d'immunoglobulines (IgM dans le cas de la MW). Cette hypergammaglobulinémie se traduit par une hyperviscovisité sanguine dans 30% des cas, et par des symptômes neurologiques dus à une réaction immunitaire vis-à-vis de la myéline couvrant les neurones et à la formation de dépôts d'anticorps circulants. D'un point de vue génétique, peu d'anomalies visibles en caryotypage conventionnel sont caractéristiques de la MW, si ce n'est la trisomie 4 pouvant toucher jusqu'à 20% des cas. L'application des techniques de séquençage à haut débit a récemment permis d'identifier la mutation L265P de *MYD88* comme présente dans plus de 90% des cas, bien que d'autres pathologies (notamment les LDGCB ABC) présentent cette mutation à des fréquences élevées [16]. Ils s'agit d'une pathologie indolente, avec une survie médiane de 5 et 10 ans, qui peut occasionnellement évoluer en LDGCB.

Le **myélome multiple** (MM) se développe à l'un des stades terminaux de la différenciation des lymphocytes B, le plasmocyte. Ces tumeurs se trouvent généralement dans la moelle osseuse, et provoquent divers symptômes liés à la fragilisation des os (douleurs osseuses, hypercalcémie, fractures) et à la perturbation de l'hématopoïèse qui s'y déroule en temps normal (anémies principalement). S'agissant de plasmocytes, ces tumeurs sécrètent une quantité importante d'immunoglobulines, souvent incomplètes (chaines légères libres et protéines « M », issues dans 50% des cas d'une IgG), dont l'accumulation peut occasionner une insuffisance rénale. Si le MM présente parfois des caractéristiques morphologiques facilitant le diagnostic, des spécificités génétiques peuvent également être identifiées. Des translocations du *locus IGH* sont ainsi retrouvées dans 55% à 70% des cas, avec des partenaires tels que *FGFR3*, *CCND1*, *CCND3*, *MAF*... Si l'hyperdiploïdie est fréquente, l'hypodiploïdie est un facteur de mauvais pronostic dans ces tumeurs, de même que certains partenaires de la translocation d'*IGH* (*FGFR3*, *MAF*) et la délétion de *TP53*. Il s'agit d'une pathologie considérée comme incurable, avec une survie médiane de 3 à 4 ans [14].

1.5.6 Autres lignées hématopoïétiques

Si les prolifération d'origine lymphoïde sont les plus fréquentes, les lignées myéloïdes peuvent également faire l'objet de transformations oncogéniques. Il existe ainsi une grande diversité de syndromes dits **myéloprolifératifs** ou **myélodysplasiques**, auxquels ajouter une part significative de **LAM**. Cette dernière se caractérise par l'envahissement de la moelle osseuse par des précurseurs myéloïdes (blastes), tandis que les syndromes myéloprolifératifs impliquent un ou plusieurs types de cellules myéloïdes normalement différenciées. La LMC en est le modèle le mieux caractérisé, dont la translocation caractéristique t(9;22) *BCR-ABL1* et les traitements ciblés associés sont fréquemment cités en exemple des progrès réalisés en oncologie grâce à la biologie moléculaire. Les syndromes myélodysplasiques se caractérisent enfin par l'envahissement de la moelle par des cellules hématopoïétiques dysfonctionnelles, et donc une ou plusieurs cytopénie périphérique (insuffisance médullaire). Enfin les cellules dendritiques et macrophages peuvent également donner lieu à des hémopathies, mais avec une incidence anecdotique.

1.6 Lymphomes B diffus à grandes cellules

1.6.1 Présentation biologique

LDGCB, NOS

Comme son nom le suggère, le Lymphome Diffus à Grandes Cellules de phénotype B (LDGCB) se caractérise au niveau cytologique par des cellules tumorales anormalement larges (de l'ordre de grandeur d'un noyau de macrophage, ou du double de la taille d'un lymphocyte normal [14]). On le qualifie de « diffus » car les lymphocytes B ganglionnaires ne suivent plus une structuration normale en follicules compacts, mais diffusent au sein de la région tumorale du ganglion. La morphologie des cellules tumorales définit 3 sous-types majoritaires, progressivement abandonnés au profit des sous-types moléculaires identifiables de manière plus objective. Le sous-type morphologique le plus fréquent est ainsi dénommé **centroblastique**, et se compose principalement de centroblastes (§1.4.4). Une composition majoritaire en cellules plus proches du stade plasmocytique (90%) définit par opposition le sous-type **immunoblastique**. Enfin le type **anaplasique** se rapproche morphologiquement des lymphomes T anaplasiques et du lymphome de Hodgkin, avec des cellules géantes et éventuellement cohésives.

D'un point de vue immunophénotypique, le LDGCB exprime les marqueurs typiques de la lignée B CD19, CD20, CD22 et CD79A [14]. Il exprime également un Ki67 élevé (40% à 90%), témoignant de ses divisions rapides et de son agressivité. Là encore deux sous-types peuvent être différenciés selon l'expression de plusieurs marqueurs propres au centre germinatif (*CD10*, *BCL6*) ou aux stades plus tardifs (*IRF4*). Cette distinction entre LDGCB similaires ou non au centre germinatif (**GCB / non-GCB** en IHC, **GCB / ABC** en biologie moléculaire) est discutée en détail dans la section 2.1.1. Les sous-types morphologiques centroblastiques et immunoblastiques peuvent logiquement être mis en parallèle des sous-types GCB et ABC respectivement [17], même si le manque de reproductibilité de leur caractérisation rend la comparaison difficile [14]. Le groupe anaplasique se caractériserait plutôt en IHC par l'expression de CD30, et l'expression de CD5 par 5% à 10% des LDGCB environ définit encore un sous-type au pronostic assez sombre [14, 18].

Le LDGCB peut apparaitre *de novo*, ou à la suite de la transformation d'un syndrome lymphoprolifératif chronique de bas grade (LF, LZM, MW, LLC...), selon des mécanismes encore étudiés aujourd'hui.

Sous-entités reconnues par l'OMS

L'OMS définit 4 sous-types de LDGCB, présentant des histiologies distinctes.

Les **LDGCB riches en lymphocytes T** constituent le premier de ces sous-types, et se caractérisent par des cellules tumorales éparpillées au sein d'un fort contingent de lymphocytes T et d'histiocytes réactionnels, similairement à ce qui peut être observé dans le lymphome de Hodgkin (§1.5.4). Cette configuration rend les analyses moléculaires délicates dans la mesure où la majorité des cellules prélevées ont soit un génome normal, soit un profil transcriptionnel non-B, bien que le phénomène soit moindre que dans le cas du LH. Ce sous-type est associé à un pronostic particulièrement sombre, et concerne moins de 10% des LDGCB [14].

Les **LDGCB primitif du SNC** (PCNSL) définissent un second sous-type de mauvais pronostic, qui se caractérise par une tumeur primaire confinée au système nerveux central (encéphale, moelle épinière) et aux yeux. Cette entité doit ainsi être distinguée des métastases cérébrales pouvant provenir d'un LDGCB survenu ailleurs dans l'organisme (notamment au niveau des testicules). Là encore les analyses moléculaires sont difficiles, dans la mesure où la localisation n'autorise pas toujours une biopsie de taille significative. Les marqueurs IHC typiques du LDGCB sont retrouvés (CD20, CD22, CD79A), en plus des marqueurs propres au sous-type ganglionnaire ABC (expression forte d'*IRF4* dans plus de 90% des cas, de *CD10* dans 10% à 20% des cas). De nombreux événements génomiques récurrents tendent à confirmer cette ressemblance : délétions 6q et 9p21, gains 18q21 (§3.1.1), mutations de *MYD88* et *CD79B* (§4) [19]...

Les **LDGCB cutanés primitif de la jambe** représentent un second sous-type défini par une localisation bien spécifique, et une ressemblance marquée avec le sous-type ganglionnaire ABC (mutations et CNA). Ces lymphomes se développent originellement au niveau de la peau, sur les parties inférieures des jambes (85% des cas), et ce préférentiellement chez la femme agée [14].

Les **LDGCB EBV+ du sujet âgé** constituent le dernier sous-type, qui souligne le rôle que peut jouer l'EBV dans l'étiologie des lymphomes [14]. Ce sous-type se caractérise par une survenue chez un patient infecté par l'EBV hors de tout contexte d'immuno-déficience (SIDA, traitement, anomalie génétique ou autre hémopathie) hormis l'immunosénescence (médiane de 71 ans). Il peut représenter jusqu'à 10% des LDGCB dans les pays asiatiques, et il est associé à un pronostic défavorable (survie médiane de deux ans environ). Si l'IHC révèle là encore un profil plutôt ABC (*CD10* et *BCL6* négatifs, *IRF4* positif), les caractéristiques génomiques ne sont cette fois-ci pas particulièrement en adéquation avec cette ressemblance [20].

Lymphomes apparentés

En plus de ces sous-types de LDGCB, de nombreux lymphomes B à grandes cellules présentent de fortes similarités avec les LDGCB et peuvent y être assimilés. On peut ainsi citer les entités suivantes [21] :

- LDGCB primitif du médiastin
- LDGCB intravasculaire
- LDGCB associé à une inflammation chronique
- Granulomatose lymphomatoïde
- LDGCB ALK+
- Lymphome plasmablastique
- LDGCB associés à la maladie de Castleman (HHV-8 multicentrique)
- LDGCB associé aux épanchements

Parmi ces entités notables, les **Lymphomes primitifs du médiastin** représentent un sous-type parfois mis en relation comme un troisième membre de la classification moléculaire GCB / ABC [22]. Ils se caractérisent par une survenue originelle au sein du médiastin (vraisemblablement dans le thymus), région thoracique difficile d'accès située entre les deux poumons, et ce préférentiellement chez la femme jeune (âge médian au diagnostic de 35 ans environ). Leurs caractéristiques génétiques les rapprochent davantage des lymphomes de Hodgkin que des LDGCB, et peuvent aider à différencier la survenue primitive dans le médiastin d'une atteinte médiastinale subséquente à un LDGCB ganglionnaire. De nombreux sites ganglionnaires et extra-ganglionnaires peuvent être envahis par la suite, notamment les reins ou le SNC en cas de progression [14]. L'expression du marqueur CD30 typique du lymphome de Hodgkin peut également être observée dans les LPM, avec une hétérogénéité toutefois suffisante pour faire la distinction, de même qu'une signature transcriptionnelle proche du sous-type GCB. Le pronostic de ces tumeurs est particulièrement bon, une survie à 5 ans de 97% ayant été rapportée en phase II aux État-Unis sous DA-EPOCH-R [23].

1.6.2 Présentation clinique

Diagnostic

Le LDGCB se présente cliniquement comme une tumeur d'évolution rapide, impliquant une ou plusieurs structures ganglionnaires voire plusieurs sites extra-ganglionnaires (appareil digestif majoritairement) [14]. Le diagnostic de LDGCB repose sur l'analyse anatomopathologique d'une biopsie-exérèse chirurgicale d'une ou plusieurs adénopathies tumorales. L'**analyse histologique** permet de confirmer les

différentes spécificités morphologiques du ganglion tumoral et des cellules qui le peuplent, ainsi que la présence ou l'absence de marqueurs cellulaires mis en évidence par immunohistochimie (§1.6.1) [21]. L'ADN peut également être extrait de la biopsie, afin de consolider le diagnostic par la mise en évidence d'une **clonalité B** (sur-représentation d'un réarrangement V(D)J, suggérant l'expansion clonale d'un lymphocyte B mature) [24].

En terme d'incidence, les LDGCB représentent jusqu'à 30% des LNH, soit un peu plus de 4 000 nouveaux cas par an en France (Figure 1.5). Il se retrouve plus fréquemment chez les hommes (sexe-ratio de 1,2) et à un âge avancé (médiane au diagnostic de 64 ans), bien qu'il soit retrouvé également chez le jeune adulte [14, 25].

Évaluation

L'évaluation du stade de développement de la maladie est réalisée selon les critères d'**Ann-Arbor**, graduant l'envahissement de l'organisme en 4 niveaux principaux de gravité croissante [24] :

Stade I Un ganglion tumoral (I), envahissement éventuel des tissus voisins (IE).

Stade II Plusieurs sites tumoraux, tous du même coté du diaphragme (II). On distingue là encore l'atteinte des tissus voisins, dans un site au maximum (IIE).

Stade III Plusieurs sites ganglionnaires de part et d'autre du diaphragme.

Stade IV Une ou plusieurs atteintes viscérales (extra-ganglionnaires).

L'évaluation de l'état général du patient, qui conditionne l'intensité du traitement qu'il est à même de supporter, est également basée sur un score de performance de gravité croissante, établi cette fois-ci par l'**ECOG** [21] :

ECOG 0 Activité normale

ECOG 1 Symptomatique, capable de réaliser de petits travaux

ECOG 2 Incapable de travailler, alité moins de 50% de la journée

ECOG 3 Confiné au lit ou dans un fauteuil plus de 50% de la journée

ECOG 4 Confiné au lit, incapable de s'occuper de lui-même

Ces différentes échelles sont rassemblées, avec d'autres critères biologiques, au sein d'un unique score pronostique : l'**IPI**. Si cet outil de stratification a été développé avant l'avènement du Rituximab [26], son intérêt pronostique a depuis été confirmé [27], même si des alternatives encore peu appliquées en clinique ont été proposées [28]. Chacun des 5 facteurs de risque suivants contribue un point à ce score, qui stratifie ainsi les patients en 6 groupes de pronostic décroissant (IPI 0 à 5).

— Dosage sérique de la LDH supérieur à la normale
— Stade de Ann-Harbor de III ou IV
— ECOG de 2 ou plus
— Age supérieur à 60 ans
— Plus d'un site tumoral extra-ganglionnaire

Les trois premiers critères de l'IPI définissent chez les patients de moins de 60 ans une variante appelée **IPIaa**, répondant aux mêmes principes (IPIaa 0 à 3).

Traitement

Les recommandations de traitement dépendent principalement de l'âge et de l'IPIaa, et sont basées sur les résultats d'essais cliniques de grande ampleur mettant en évidence la meilleure option thérapeutique pour chaque groupe de patients [24]. De nombreux facteurs doivent toutefois être pris en compte dans le choix du traitement, qui est dans la pratique décidé lors de réunions de concertations pluridisciplinaires impliquant cliniciens, anatomopathologistes et biologistes.

Les patients de **moins de 60 ans sans facteur de risque** (IPIaa à 0) et sans masse tumorale importante reçoivent ainsi le traitement de référence, basé sur 6 cures de **R-CHOP** espacées de 21 jours (R-CHOP21). Ce traitement consiste en une polychimiothérapie composée de cyclophosphamide (agent alkylant, endommageant l'ADN), doxorubicine (anthracycline, intercalant de l'ADN), vincristine (poison du fuseau mitotique) et prednisone (corticoïde), particulièrement dommageable aux cellules en division. L'addition de l'anticorps monoclonal Rituximab permet de compléter ces effets par une destruction ciblée des lymphocytes B, normaux ou tumoraux, en ciblant le marqueur CD20 exprimé à leur surface depuis le stade Pré-B (Figure 1.1). La survie à 3 ans dans cette catégorie est alors supérieure à 90% [21].

Les patients de **moins de 60 ans présentant un facteur de risque** (IPIaa à 1 ou masse tumorale importante) peuvent recevoir une chimiothérapie plus agressive, se déroulant en 3 étapes : « induction » par 4 cures de **R-ACVBP** espacées de 14 jours, suivies de 2 cures de méthotrexate à forte dose et d'une « consolidation séquentielle » incluant Rituximab, étoposide, ifosfamide et aracytine [21]. La toxicité importante de cette chimiothérapie ne permet son application qu'à des patients suffisamment jeunes pour y résister, et atteints d'une maladie suffisamment agressive pour équilibrer bénéfices et risques. Là encore une survie à 3 ans de 87% peut être atteinte [21]. Alternativement, l'addition d'une radiothérapie localisée de la masse tumorale au standard R-CHOP21 peut également être appliqué à ces patients.

Le traitement des patients de **moins de 60 ans avec plusieurs facteurs de risque** (IPIaa à 2 ou 3) est moins consensuel, et plusieurs thérapies lourdes sont actuellement considérées. Le R-CHOP14, consistant en des cures plus rapprochées que

dans le protocole standard, est une solution qui tend à disparaître par manque de gain significatif en terme de survie globale vis à vis du R-CHOP21 [24]. Les chimiothérapies lourdes telles que R-ACVBP ou R-CHOEP sont également possibles, mais leur supériorité par rapport au R-CHOP21 reste à démontrer dans le cadre d'un essai clinique. L'**autogreffe** de cellules souches hématopoïétiques peut également être envisagée en première ligne dans ce groupe de patients, bien que là encore les bénéfices soient modestes. Les CSH sont recceuillies par cytaphérèse avant myélosuppression, via une chimiothérapie particulièrement intensive (BEAM), puis réinjectées afin de rétablir une production de cellules hématopoïétiques normale. Si les risques de rejet sont nuls compte tenu de la provenance du greffon, ce traitement entraine une aplasie requérant une hospitalisation de plusieurs semaines en chambre stérile, et n'est envisagée que si les bénéfices attendus sont supérieurs aux alternatives.

Entre **60 et 80 ans**, les chimiothérapies lourdes ne sont généralement plus envisagées et le traitement curatif consiste généralement en 6 à 8 cures de R-CHOP21. **Après 80 ans**, une variante du R-CHOP21 basée sur un dosage moindre (**R-miniCHOP**) est préférée, le traitement à visée curative étant conditionné par une évaluation gériatrique préalable. La doxorubicine présentant une toxicité cardiaque importante, et peut éventuellement être remplacée par des produits équivalents chez les patients âgés et à risque (R-C.OP).

Suivi

L'évaluation de la réponse au traitement est généralement réalisée grâce à des techniques d'imagerie médicale telles que la tomodensitométrie (TDM) et la TEP au FDG, couplées en un unique examen aujourd'hui référence dans les LDGCB. Si la TDM offre une vision précise des organes internes en mesurant l'absorption des rayons X (Figure 1.6A), la TEP permet de localiser un traceur radioactif injecté au préalable. Le traceur utilisé s'apparentant au glucose, il va être capté préférentiellement par les cellules à fort métabolisme (dont les cellules tumorales qui se divisent rapidement), et donc entrainer une hyper-fixation dans ces régions (Figure 1.6B).

Plusieurs foyers présentent un métabolisme du glucose physiologiquement élevé (médiastin, foie, cerveau...), ou sont impliqués dans son élimination (reins). L'échelle de **Deauville** propose ainsi une analyse comparative à ces foyers :

1. Pas de fixation
2. Fixation d'intensité inférieure ou égale au médiastin
3. Fixation d'intensité comprise entre celle du médiastin et du foie sain
4. Fixation modérément supérieure au foie sain
5. Fixation fortement supérieure au foie sain

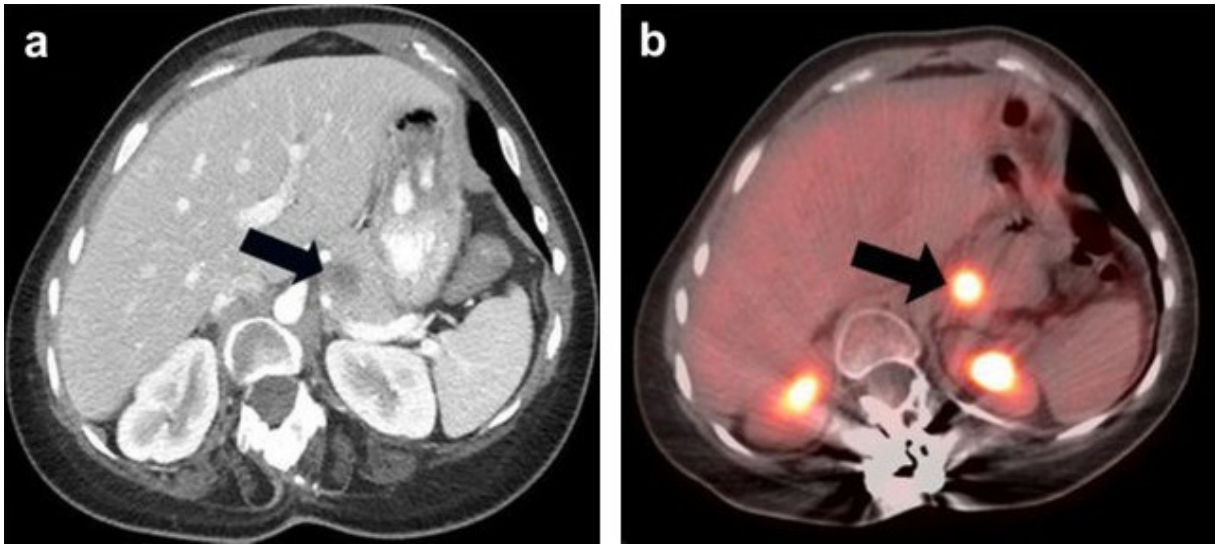


Figure 1.6 – Mise en évidence d'un lymphome par TDM et TEP

Le panneau **a** présente en coupe axiale une tumeur localisée dans le pancréas (flèche noire). Le panneau **b** présente une coupe similaire en combinant TDM et TEP, mettant en exergue un foyer métabolique tumoral (marqué d'une flèche noire) et l'activité métabolique des deux reins. Figure adaptée de Nguyen *et al*, JOP 2011 [29].

L'observation des échelons 1 à 3 après traitement est ainsi considérée selon les critères de Cheson (2014) comme une **réponse métabolique complète**, la tumeur n'étant plus détectable. Les échelons 4 et 5 sont interprétés par comparaison aux images obtenues au diagnostic, définissant des **réponses partielles**, **maladies stables** ou **progressives** [30]. Cette évaluation est généralement appliquée en fin de traitement, et éventuellement à mi-traitement (3 ou 4 cures) [24].

En l'absence de rechute, un suivi régulier par un médecin est conseillé (3, 6, 9, 12, 18, 24, 30 et 36 mois, puis une fois par an). La survie des patients après deux ans sans événement est jugée similaire à celle de la population générale [24, 31].

Rechute

Considérant le traitement de référence (R-CHOP), environ 30% des LDGCB sont amenés à rechuter [24]. Selon une étude portant sur 765 LDGCB traités par R-CHOP, environ 11% des patients sont réfractaires (maladie stable ou progressive sous traitement), 4% présentent une rechute rapide (moins de 6 mois) et 5% décèdent durant le traitement [32]. Les rechutes tardives peuvent expliquer les 10% restants.

Les outils diagnostiques et d'évaluation sont similaires à ceux utilisés lors du diagnostic initial, et impliquent si la première ligne avait une visée curative un traitement de rattrapage, dont d'éventuelles thérapies ciblées si la tumeur présente des caractéristiques qui le permettent. Différentes combinaisons de chimiothérapies peuvent

être utilisées en secondes lignes et suivantes, telles que R-DHAP, R-ICE, R-GEMOX ou R-GDP [24]. L'autogreffe est également privilégiée en traitement de rattrapage, et peut être elle même rattrapée par une **allogreffe**. Dans ce dernier cas, les CSH greffées proviennent d'un donneur géno-identique (frères et sœurs) ou phéno-identique (compatibilité HLA de 10/10) et offrent un potentiel effet curatif (*graft versus tumor*).

2 Caractères transcriptomiques

2.1 État de l'art

2.1.1 Sous-types GCB et ABC

Découverte grâce aux puces transcriptomiques

L'avènement des puces transcriptomiques à la fin des années 1990 [33–35] a permis une avancée majeure dans la compréhension des LDGCB, en révélant deux sous-types très différents sur les plans clinique et biologique.

En appliquant un algorithme de *clustering* non-supervisé à l'expression de 17 856 ARN mesurés au moyen d'une puce « à façon » [37], Alizadeh *et al* [36] ont pu fournir une comparaison sans précédent des principales lymphopathies B (LDGCB, LF, LCM, LLC). Observant au sein des LDGCB deux sous-ensembles se distinguant par la sur-expression ou non des ARN du centre germinatif, ils se sont attachés à affiner cette dichotomie en sélectionnant les ARN sur-exprimés par chacun de ces deux sous-types (Figure 2.1). La comparaison à différentes populations de lymphocytes normaux les a conduit à identifier ces deux sous-groupes comme « similaires aux lymphocytes B du centre germinatif » (GCB) ou « similaires aux lymphocytes B périphériques activés » (ABC), une stratification qui s'est avérée pronostique parmi les 40 patients étudiés. Appliquant une classification similaire basée sur les 100 ARN les plus discriminants dans la série d'Alizadeh *et al*, Rosenwald *et al* ont pu confirmer le pronostic plus favorable des LDGCB GCB par rapport aux ABC dans une série de 240 patients traités par CHOP (survie à 5 ans respectives de 60% et 35%), en notant toutefois l'apparition d'un « type 3 » intermédiaire et hétérogène [38]. Face à la difficulté de se procurer les puces « Lymphochip » nécessaires à cette classification et à l'instabilité inhérente aux techniques de *clustering* hiérarchique [39], Wright *et al* développeront l'année suivante un test basé sur les puces pan-transcriptomiques commercialisées par Affymetrix qui reste encore à l'heure actuelle le « *gold standard* » [40]. Ce facteur

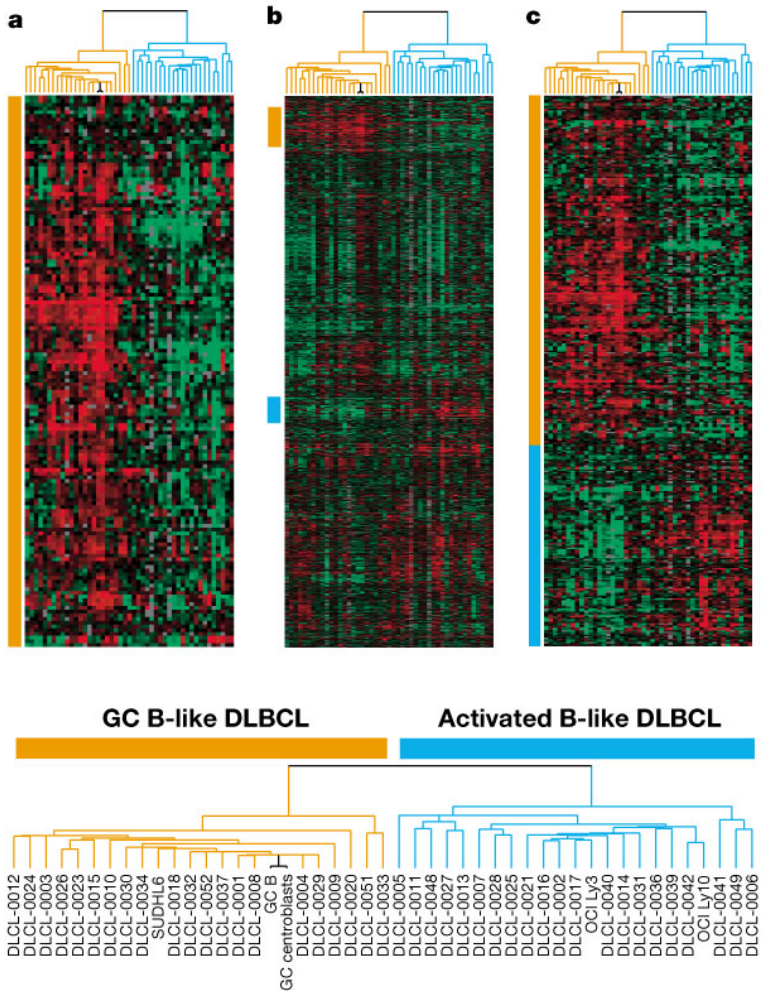


Figure 2.1 – *Clustering* GCB/ABC originel sur puces Lymphocip

Figure extraite de Alizadeh *et al*, Nature 2000 [36]

pronostique se verra une fois de plus confirmé à l'ère du Rituximab, avec l'observation de survies à 5 ans de 80% et 50% pour les LDGCB GCB et ABC parmi une cohorte de 233 patients traités par R-CHOP [41].

Outre cet aspect pronostique, ces deux sous-types s'avèrent très dissemblables sur le plan biologique, comme nous le soulignerons à travers les différents chapitres de ce mémoire. En accord avec le principe de la classification actuelle des hémopathies [14], qui tend à attribuer un « type cellulaire d'origine » à chaque classe de lymphome [42], cette disparité biologique pourrait s'expliquer par une origine cellulaire différente. Le sous-type GCB correspondrait ainsi à la transformation de centroblastes, bloqués à ce stade primitif dans un centre germinatif dont l'expansion deviendra tumorale. L'observation d'un phénomène d'hyper-mutation somatique particulièrement intense dans ce sous-type tendrait à corroborer cette hypothèse [43], de même que l'expression de gènes caractéristiques de la formation du centre germinatif tels que *BCL6*, *MME* ou *CD38* [36]. Le sous-type ABC correspondrait quand à

lui à un type cellulaire plus tardif, proche du plasmocyte quoique moins différencié. Ce type cellulaire se retrouve dans l'immunopoïèse en aval de l'activation classique par l'antigène (dite « chronique »), qui s'avère dispensable dans environ 10% des ABC (activité basale ou « tonique ») [42]. Le maintien à long terme de cette activation, mise en évidence entre autres par la sur-expression d'*IRF4* [44], contribue dans ce sous-type à une activation de la voie NFκB [45] qui fait l'objet de nombreux développements thérapeutiques [46–48].

Transposition en algorithmes IHC

L'utilisation de puces transcriptomiques nécessitant un appareillage et un savoir-faire peu commun dans les laboratoires de diagnostic, la nécessité d'élaborer un outil plus accessible est rapidement apparue. L'immunohistochimie constituait pour cela un outil de choix, cette classification s'intégrant parfaitement dans la routine anatomopathologique actuelle qui impliquait déjà des observations microscopiques sous différents marquages.

Hans *et al* [49] se sont ainsi attelés à développer un algorithme IHC capable d'identifier avec une précision de 80% les LDGCB de type GCB, regroupant les ABC et indéterminés dans une classe « non-GCB » (Figure 2.2). La validité de cet algorithme fut fréquemment remise en question [50–53], et de nombreuses alternatives furent proposées au fil des années (Figure 2.2). Si l'algorithme originel de Hans visait clairement la classification GCB/ABC découverte grâce aux *Lymphochips*, il faut signaler que plusieurs équipes s'en servirent plutôt de base pour créer des outils pronostiques plus complexes ou performants, d'où découle une certaine confusion dans la littérature [53, 54].

Parmi les nombreux algorithmes IHC applicables aux LDGCB (Figure 2.2), seuls les algorithmes **Hans** [49], **Choi** [55], **Choi*** [56], **Tally** [56] et **Visco-Young** [57] furent développés dans des séries de patients également caractérisés par les techniques transcriptomiques de référence, avec l'objectif d'obtenir la même classification à moindre coût. **Choi** *et al* [55], profitant du perfectionnement des anticorps, établirent ainsi un algorithme plus complexe affichant une concordance de 93% avec le classifieur de Wright *et al* appliqué sur puces Affymetrix U133+2. Consciente de la variabilité inter-observateurs propre aux techniques IHC, cette équipe a également pu démontrer que la variabilité habituellement observée de ±10% sur chaque marqueur n'avait pas d'impact significatif sur le résultat (IC_{95} de 87% à 93%). *BCL6* étant considéré comme particulièrement difficile à interpréter en IHC [58], Meyer *et al* [56] s'attachèrent à démontrer que s'en passer (**Choi***) conduit à des résultats comparables. Ils constatèrent également que l'ordre des marquages n'avait que peu d'importance, et mirent au point un algorithme se contentant de « compter les points »

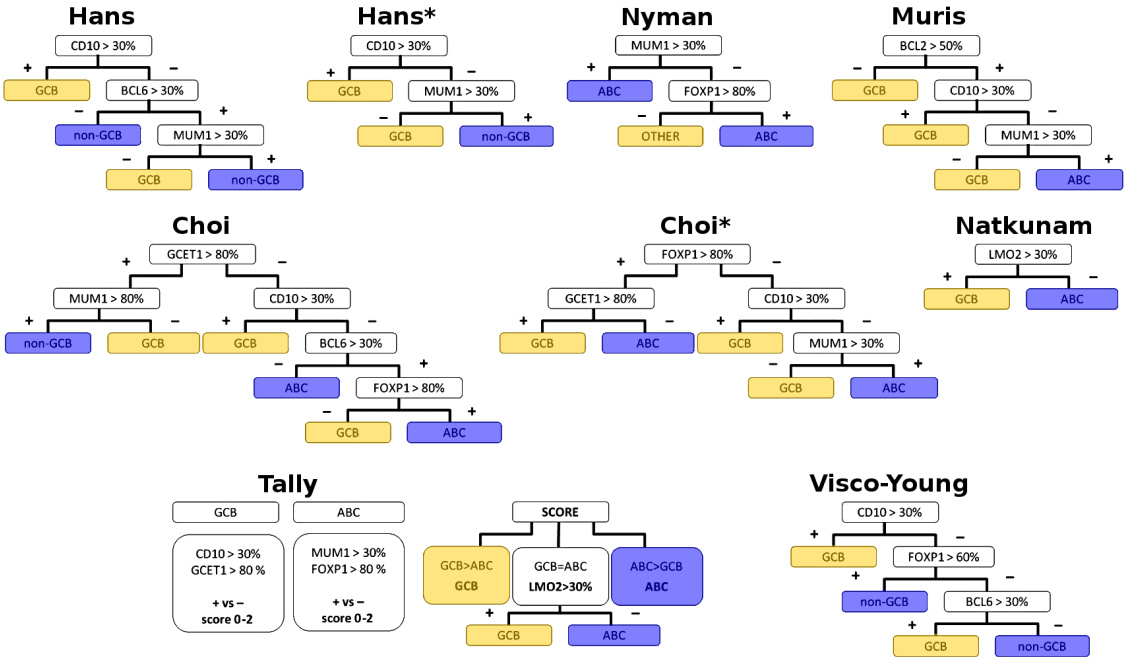


Figure 2.2 – Principaux algorithmes IHC appliqués aux LDGCB

Figure adaptée de Coutinho *et al*, CCR 2013 [54]

(**Tally**), algorithme qui démontra un niveau de concordance légèrement supérieur aux autres dans leur série de 192 cas classés par puces Affymetrix (93%). Visco *et al* [57], appliquant une méthodologie mathématiquement plus rigoureuse dans une série de 475 LDGCB, démontra enfin une concordance similaire en revenant à un algorithme plus simple (**Visco-Young**), basé sur 3 anticorps successifs incluant le controversé *BCL6*.

Par opposition, les algorithmes **Hans*** [59], **Muris** [59], **Natkunam** [60] et **Nyman** [61] (Figure 2.2) visaient une classification biologiquement similaire, mais n'ont pas été développés comme substituts directs aux classifieurs transcriptomiques. Muris *et al* [59], concédant un pouvoir prédictif de la survie globale à l'algorithme de Hans (p=0,004), constatèrent que ce pouvoir pouvait être conservé en retirant d'une part *BCL6* de l'équation (**Hans***) et en y incluant *BCL2* (**Muris**), dont la seule expression fut déjà démontrée comme pronostique [62]. *LMO2* faisant partie des gènes les plus différentiellement exprimés entre LDGCB GCB et ABC [36], sa seule expression en fait un facteur pronostique fortement lié à cette classification, ce que Natkunam *et al* [60] ont pu démontrer avec leur algorithme basé uniquement sur ce gène (**Natkunam**). Constatant que l'algorithme de Hans ne permettait plus une stratification pronostique des patients à l'ère du Rituximab [63], un constat qui sera variablement confirmé ou infirmé par d'autres études, Nyman *et al* [61] s'attachèrent finalement à développer un nouvel algorithme de stratification pronostique (**Nyman**) en incluant des marqueurs liés au phénotype ABC (*IRF4* et *FOXP1*).

Chacun de ces algorithme fut occasionnellement validé et fréquemment rejeté par des études indépendantes, entretenant dans ce domaine une certaine défiance à l'égard de l'IHC en général. Une grande variabilité à la fois dans l'exécution et dans l'interprétation de ces marquages fut ainsi démontrée par de Jong *et al* [58], qui observèrent un taux de concordance deux-à-deux entre 9 laboratoires de 40% à 90% selon les marqueurs (la concordance étant inversement proportionnelle au nombre de classes d'intensités à attribuer au marquage). Des résultats similaires (50% à 80% de concordance), quoique jugés cette fois-ci acceptables, furent retrouvés par une autre équipe [64]. Une grande disparité entre les prédictions de ces différents algorithmes fut également notée [54, 58], bien que de toute manière peu de ces prédictions soient considérées comme significativement liées au pronostic [51, 54, 58, 65, 66]. Culpin *et al* [65] proposent de nombreuses explications plausibles à ces disparités : l'utilisation de TMA plutôt que de biopsies complètes, qui réduisent l'hétérogénéité d'une tumeur et d'un marquage à un unique cœur de moins d'un mm de diamètre [67], l'uniformité des populations et thérapeutiques qui est difficile à assurer en dehors d'essais cliniques dédiés, ou encore la grande variabilité des anticorps, protocoles de marquage et techniques de *scoring* utilisés. Malgré les progrès réalisés en matière d'aide informatique à l'interprétation de marquages [68], la plupart sont encore aujourd'hui directement évalués au microscope, donnant à ces algorithmes une composante subjective de moins en moins acceptée à mesure que les outils informatiques et statistiques se développent.

En dépit des nombreuses études signalant la supériorité des techniques transcriptomiques sur les techniques IHC dans la classification GCB/ABC des LDGCB [51, 53], les algorithmes IHC (et plus particulièrement l'algorithme de Hans) constituent aujourd'hui le standard *de facto*, en diagnostic comme en recherche clinique [48].

Transposition en techniques moléculaires

En parallèle de ces développements IHC, de nombreuses équipes conscientes de ces faiblesses s'attachèrent à développer des outils transcriptomiques visant les mêmes objectifs. La cryoconservation des échantillons étant plutôt l'apanage des laboratoires de recherche, plusieurs développements visèrent à appliquer les techniques de mesure pan-transcriptomiques aux échantillons conservés en paraffine (FFPE), plus répandus mais moins propices aux analyses de biologie moléculaire [69].

Williams *et al* [70] démontrèrent ainsi qu'une technique d'amplification spécifique (Ribo-SPIA [71]) permettait l'analyse d'échantillons FFPE sur puces Affymetrix, avec une qualité inférieure au congelé mais néanmoins suffisante pour obtenir une classification fiable. Parmi 59 paires d'échantillons analysés à la fois en congelé avec le protocole classique et en FFPE avec Ribo-SPIA, excluant 11 échantillons indéterminés

en congelé, 90% furent classés correctement et 8% ne le furent pas du tout faute de certitude statistique (« indéterminés »). Cette technique fut appliquée dans une série de 475 TMA par Visco *et al* [57], qui constatèrent des niveaux de corrélation attendus avec les algorithmes IHC de Hans et de Choi, ainsi qu'avec leur propre algorithme (87%, 90% et 93% respectivement).

Barrans *et al* [72] firent de même avec une adaptation de la technologie DASL étendue au transcriptome complet [73], rapportant un taux d'échec expérimental de 17% et une concordance avec l'algorithme IHC de Hans de 78% dans une série de 206 biopsies FFPE. Le modèle statistique employé pour la classification (DAC, *DLBCL Automatic Classifier*) diffère toutefois du modèle bayésien développé par Wright *et al* et repris par la plupart des travaux ultérieurs. En l'absence de données collectées sur échantillons congelés, cette équipe démontra la validité de son approche par de nombreuses preuves indirectes, telles que la validation du modèle statistique seul dans plusieurs séries, la confirmation d'un pronostic plus sombre pour le groupe ABC et l'enrichissement de plusieurs signatures transcriptomiques liées au sous-type.

Dans la mesure où Wright *et al* ont pu mettre au point un classifieur reposant sur quelques dizaines de gènes, la nécessité et la complexité de ces techniques pan-transcriptomiques ont été fréquemment remises en question par des équipes proposant des outils plus ciblés et moins coûteux. Jaïs *et al* [74] appliquèrent par exemple un *clustering* hiérarchique à l'expression de 17 gènes de la signature de Wright *et al* mesurés en TLDA sur échantillons congelés pour obtenir cette classification. L'absence de comparaison directe avec des puces Lymphochip ou Affymetrix fut là encore comblée par des preuves indirectes (impact pronostique).

Coupler la simplicité d'un test ciblé à la portabilité d'un test compatible FFPE nécessite toutefois des techniques plus exotiques. Rimsza *et al* [75] utilisèrent à cet effet un test soustractif basé sur la protection des 14 ADNc ciblés contre la nucléase S1 (ArrayPlate qNPA [76]), dont les performances furent démontrées dans une série de 121 LDGCB classés également par puces Affymetrix sur échantillons congelés (95% de concordance en excluant les indéterminés). Collie *et al* [77] proposèrent l'utilisation d'une plateforme de qPCR multiplexe développée par PrimeraDx, dont l'utilisation IVD a été certifiée par la FDA et la communauté européenne (ICEPlex™), fournissant une preuve de concept encourageante basée sur l'expression de 14 gènes. Hall *et al* [78] parièrent sur la technologie QuantiGene Plex™ [79], reposant sur des constructions branchées d'oligonucléotides capables d'amplifier et rendre détectables des signaux très faibles (ADN viraux initialement). Bien qu'ayant validé leur modèle bayésien dans des cohortes analysées par d'autres techniques et démontré la supériorité de leur test vis-à-vis de la qPCR, ces derniers auteurs ne disposaient que de 40 échantillons classés en Affymetrix pour valider leur classification.

L'application de la technologie nCounter™ commercialisée par *NanoString Technologies* fut enfin mise en œuvre en parallèle par plusieurs équipes [80–82], présentant des résultats satisfaisants dans des cohortes distinctes et avec des modèles distincts. Malgré le coût et la distribution encore marginale de l'appareillage nécessaire à son application, cette technologie semble aujourd'hui la mieux placée pour devenir le nouveau standard en pratique clinique [83–85] (§6.1).

2.1.2 Autres caractères transcriptomiques

Profils transcriptomiques complexes

Malgré l'engouement suscité par la dichotomie GCB/ABC mise en évidence par Alizadeh *et al*, de nombreuses équipes tentèrent de reprendre à zéro la classification des LDGCB avec des puces plus récentes, et surtout des cohortes plus importantes (40 LDGCB seulement dans la série originelle).

Bien qu'ayant confirmé la validité pronostique des groupes GCB/ABC, Rosenwald *et al* [38] cherchèrent également à construire d'autres outils pronostiques dans leur série de 240 patients traités par CHOP. Ils sélectionnèrent pour cela les 670 gènes dont l'expression était liée au pronostic (modèles de Cox univariés), qu'ils regroupèrent ensuite par *clustering* hiérarchique. Comme Alizadeh *et al* avant eux, ils mirent ainsi en évidence une signature de bon pronostique propre au centre germinatif (GC), à laquelle s'ajoutèrent des signatures caractéristiques des ganglions lymphathiques réactionnels (LN, bon pronostic), de la perte du MHC de classe II (valeur pronostique non validée) et des cellules en division (P, mauvais pronostic). Ils simplifièrent ensuite leur modèle en choisissant 16 gènes représentatifs de cette classification, dont la combinaison linéaire des expressions prédisait la survie indépendamment de l'IPI. De tous les gènes absents des 4 signatures étudiées, seul *BMP6* améliorait de façon significative ce modèle et y fut donc ajouté.

Lenz *et al* [41] reprirent cette classification en combinant à 181 patients traités par CHOP de la série de Rosenwald *et al* 233 nouveaux patients traités par R-CHOP, tous analysés sur puces Affymetrix U133+2. Cette série sera l'occasion de valider l'impact pronostique de la classification GCB/ABC à l'ère du Rituximab, mais également des signatures GC/LN/P (la signature MHC II étant là encore non significative). Parmi la classification établie par Rosenwald *et al*, ils ne retinrent que les composantes GC et LN, renommant cette dernière « stromal 1 ». Ils y ajoutèrent une troisième signature « stromal 2 », composée de gènes choisis spécifiquement pour leur apport significatif à ce duo (modèles de Cox multivariés). Le modèle tri-varié ainsi obtenu fut démontré comme particulièrement prédictif de la survie, et ce indépendamment de l'IPI, dans

les cohortes CHOP et R-CHOP, ainsi que dans une cohorte extérieure de 177 LDGCB traités par CHOP.

Monti *et al* [86] établirent une classification parallèle dans 176 LDGCB traités par CHOP, en utilisant des puces Affymetrix U133A/B ne présentant pas le biais d'enrichissement en gènes d'intérêt hématologique des Lymphochips. Ils apportèrent un soin tout particulier à produire des signatures robustes, combinant les résultats issus de 3 techniques de *clustering* non-supervisé distinctes appliquées à 200 sous-échantillons aléatoires de leur série. Ils identifièrent les groupes de tumeurs ainsi obtenus en appliquant une méthode d'enrichissement en signatures moléculaires connues, se concentrant sur les gènes dont les niveaux d'expression étaient les plus représentatifs de chacun de ces groupes. Les 3 *clusters* les plus robustes furent ainsi identifiés comme « OxPhos » (production d'ATP par phosphorylation oxydative dans les mitochrondries), « BCR/proliferation » et « Host-Response » (lymphocytes T, tissu conjonctif et inflammation), mais ne présentèrent pas de valeur pronostique.

Combinant 3 cohortes de LDGCB publiées [87–89] à un nouvel outil statistique (CGSA), Rosolowski *et al* [90] mirent au point une quatrième classification. Les 50 groupes de gènes fortement corrélés observés au sein de leur jeu de données furent ainsi combinés en 3 groupes, respectivement nommés HiGA-PRO (prolifération), HiGA-SIR (stroma et réponse immune) et LoGA (groupes de gènes divers peu exprimés). Comme son homologue HR chez Monti *et al*, le sous-type majoritaire (52%) HiGA-SIR fut démontré comme associé à des patients plus jeunes, de meilleur pronostic et avec des tumeurs à plus faible densité de cellules tumorales par rapport à HiGA-PRO. Les auteurs peineront toutefois à décrire le nouveau sous-type LoGA qu'ils rapportent, l'hypothèse d'un artefact d'analyse n'étant pas à exclure compte tenu de ses caractéristiques inhabituelles.

Un certain recoupement peut être observé entre ces 4 études, qui ajoutent à la composante « centre germinatif » originellement observée par Alizadeh *et al* deux grandes composantes, l'une plutôt centrée sur le micro-environnement (stromal, HR, LN, HiGA-SIR) et l'autre sur la prolifération cellulaire (P, HiGA-PRO, BCR/proliferation). Dans la mesure où la plupart de ces signatures furent identifiées subjectivement par l'observation de quelques gènes jugés représentatifs parmi de nombreux autres souvent inconnus, il est toutefois nécessaire d'interpréter ces résultats avec précaution et de se méfier de la pollinisation intellectuelle croisée. Afin d'éviter cet écueil, Dybkær *et al* [91] choisirent récemment de définir des groupes *a priori* à partir de cellules saines triées, reprenant à sa source l'hypothèse d'un type cellulaire d'origine distinct sous-jacente à la classification GCB/ABC. Ils séparèrent ainsi les 1139 LDGCB collectés à travers diverses cohortes publiées en 5 « BAGS », selon leur ressemblance aux centroblastes, centrocytes, lymphocytes B mémoires ou plasmablastes,

une classe « indéterminée » venant là encore compléter le tableau. Bien que pronostique dans la cohorte globale (centrocytes et centroblastes étant associés à de particulièrement bons et mauvais pronostics respectivement), cette classification ne semble pas totalement indépendante des groupes GCB/ABC. Comparant ces profils d'expression géniques à ceux observés dans 26 lignées cellulaires B de sensibilité variable à différentes chimiothérapies, ces auteurs apportèrent également comme explication à cette différence de pronostic une plus grande résistance des centroblastes à la doxorubicine et la vincristine que les centrocytes, une constatation aux applications cliniques évidentes.

Expression de gènes isolés

Face à la complexité des puces pan-transcriptomiques, de nombreuses équipes tentèrent de mettre en évidence des biomarqueurs plus simples, se résumant à l'expression de quelques gènes facilement quantifiables en clinique, par IHC ou qPCR.

Le ratio entre nombre de gènes et nombre de patients étudiés dans les études pan-transcriptomiques étant propice à l'*over-fitting*, l'analyse de panels plus restreints choisis *a priori* sur des critères bibliographiques constituait une bonne approche pour résoudre ce problème. Lossos *et al* [92] quantifièrent ainsi l'expression de 36 gènes par qPCR dans 66 LDGCB traités par CHOP, et mirent en évidence les pouvoirs prédictifs de la survie des gènes *BCL6*, *LMO2*, *CCND2* et *BCL2*. Combinant ces gènes au sein d'un modèle multivarié, ils démontrèrent une stratification significative des patients confirmée dans les séries de puces disponibles à l'époque [38, 93]. Le prognostic défavorable de la sous-expression de *BCL6* fut observé par plusieurs équipes [94–96], mais semble effacé par le Rituximab [95, 97].

Jaïs *et al* [74] choisirent tout de même d'établir leur panel de gène à partir de 53 puces Affymetrix U133A mêlant patients traités par CHOP et R-CHOP, cherchant à la fois des gènes dont le pouvoir pronostique était lié (33) ou non (32) au Rituximab. Étudiant ces gènes en TLDA dans une cohorte étendue de 44 patients R-CHOP, ils confirmèrent 9 gènes pronostiques indépendants du traitement (*KANK1*, *BCL7A*, *MME*, *RFTN1*, *SYPL1*, *LMO2*, *MYBL1*, *LPP*, *FOXP1*) et 3 gènes pronostiques uniquement sous Rituximab (*APOBEC3G*, *JAK2*, *RAB33B*), et constatèrent que seuls ces derniers n'étaient pas redondants avec la classification GCB/ABC. Ce recoupement avec la classification GCB/ABC peut être observé dans de nombreux travaux, notamment par le biais des gènes *LMO2* [60, 98] et *FOXP1* [99, 100] qui font partie des plus représentatifs des sous-types GCB et ABC respectivement.

D'autres équipes choisirent de s'intéresser à des gènes peu spécifiques du lymphome, mais plutôt des cancers en général. *TP53* en est l'exemple emblématique

[101], les LDGCB ne dérogeant pas à la règle d'une expression anormalement élevée et d'un lien avec un pronostic plus sombre [102], y compris sous Rituximab [103]. La protéine codée par *MKI67*, marqueur de prolifération suivi en IHC dans de nombreuses pathologies [104], fut également mise en évidence comme un marqueur pronostique de choix dans les LDGCB, avant et après Rituximab [105]. *CXCR4*, bien qu'originellement étudié dans les lymphocytes, apparait également comme un acteur de dissémination tumorale non spécifique en oncologie [106], dont la sur-expression annonce des rechutes plus fréquentes dans les LDGCB également [107].

Plus près des lymphomes, on retrouve également comme facteur pronostique les sur-expressions de *MYC* et *BCL2* [108], qui peuvent être directement reliées aux translocations t(8;14) et t(14;18) impliquant une substitution du promoteur par celui des chaînes lourdes d'immunoglobuline (§3.1.2). L'antigène CD30, jusqu'ici plutôt considéré comme un marqueur de lymphomes hodgkiniens, a également été démontré comme sur-exprimé par environ un quart des LDGCB de moins bon pronostic [109], ouvrant la voie à de nombreux essais cliniques basés sur l'anti-CD30 Brentuximab Vedotin [110, 111]. Enfin l'antigène CD5 traditionnellement associé aux lymphocytes T a pu être détecté comme sur-exprimé dans 5% à 10% des LDGCB, mettant en évidence un sous-groupe restreint de mauvais pronostic potentiellement caractérisé par un micro-environnement inhabituel [14, 18].

De manière plus anecdotique, plusieurs équipes se sont intéressées à l'expression des micro-ARN, attirés par leur rôle de régulation de la transcription dans des groupes de gènes parfois conséquents. Les niveaux d'expressions de miR-18a [112], miR-23a [113], miR-129 [114], miR-155 [115], miR-181a [112], miR-200 [116], miR-222 [112] et miR-224 [117] ont pu être associés tour à tour au pronostic dans les LDGCB, tandis que des travaux sur une lignée cellulaire suggèrent miR-21 comme un facteur de résistance au CHOP [118]. Bien qu'un consensus soit encore difficile à obtenir sur la reproductibilité de chacun de ces impacts [119], ces marqueurs s'annoncent prometteurs dans la mesure où leur détection dans le sérum à des fins pronostiques a déjà pu être démontrée [120].

2.2 Objectifs et approche suivie

La classification des LDGCB en sous-types GCB et ABC constituant une stratification à la fois biologique et pronostique reconnue (§2.1.1), elle nous est apparue comme une base indispensable aux nombreux projets de recherche menés par l'unité dans le but de découvrir de nouveaux facteurs pronostics. Face au manque de reproductibilité des techniques d'immunohistochimie [51, 58, 121] et à la complexité de mise en œuvre des puces pan-transcriptomiques considérées à l'heure actuelle

comme la technique de référence [40], il nous est paru nécessaire de développer un outil capable d'analyser à moindre coût des séries de patients conséquentes.

C'est dans ce but que l'unité s'est dotée en 2010 de la technologie DASL (§2.3.1), avec l'idée de mesurer l'expression des gènes inclus dans le classifieur de référence [40]. Cette classification, dont l'élaboration et la mise en œuvre dans plus de 200 tumeurs ont constitué une partie de mes travaux de Master, a servi de base à de nombreuses publications (§2.4).

Suite à l'abandon de cette technique par la société Illumina en 2013, le docteur Philippe Ruminy a entrepris de simplifier cet outil de manière à le rendre indépendant des réactifs et appareillages propriétaires, en revenant aux principes de base de la RT-MLPA (§2.3.2). La sélection des gènes inclus dans ce test, la construction du modèle de prédiction et sa validation statistique dans plusieurs cohortes indépendantes ont constitué mes contributions à cet outil, aujourd'hui publié et breveté (§2.5).

2.3 Méthodes

2.3.1 DASL

Les kits DASL™(*cDNA-mediated Annealing, Selection, extension, and Ligation*) commercialisés par Illumina jusqu'en 2013 proposaient de mesurer l'expression de 96 gènes dans 96 échantillons, en se basant sur les technologies Illumina GoldenGate™et VeraCode™.

La chimie **GoldenGate™**, développée initialement pour le génotypage, repose sur l'hybridation multiplexe de couples d'oligonucléotides synthétiques propres à chaque cible (Figure 2.3). Ces deux oligonucléotides contiennent des parties complémentaires à la cible d'une vingtaine de nucléotides chacune (vert), qui permettront leur hybridation côte à côte puis leur ligation après une courte phase d'extension (une base). Cette jonction autorise une amplification par un couple d'amorces commun marqué d'un fluorochrome (P1 et P2), l'utilisation de plusieurs amorces P1 composées de séquences complémentaires et de fluorochromes différents permettant de quantifier chaque génotype. L'application de ce principe à la quantification d'ARN requiert simplement une phase préliminaire de rétro-transcription non-spécifique des ARN totaux, et l'utilisation d'une unique amorce P1. En ciblant une jonction d'exons lors de leur élaboration, on s'assure de plus une insensibilité à l'ADN génomique contaminant qui peut subsister après extraction.

La technologie **VeraCode™** permet de mesurer simultanément l'ensemble de ces signaux tout en identifiant à chaque fois la cible mesurée. Pour cela, une séquence

nucléotidique d'identification du gène (illu) est introduite dans l'oligonucléotide 3', permettant l'hybridation du brin d'ADN sur le micro-cylindre de verre correspondant (28 µm de diamètre) gravé d'un code-barre. L'appareil de mesure associé (BeadXpress™) lira conjointement les intensités de fluorescence et les code-barres, de manière à associer chaque signal à un gène. Le multiplexage des échantillons s'obtient en les répartissant sur une plaque de 96 puits, le BeadXpress™ réalisant séparément les étapes d'hybridation sur cylindre et de lecture pour chaque puits.

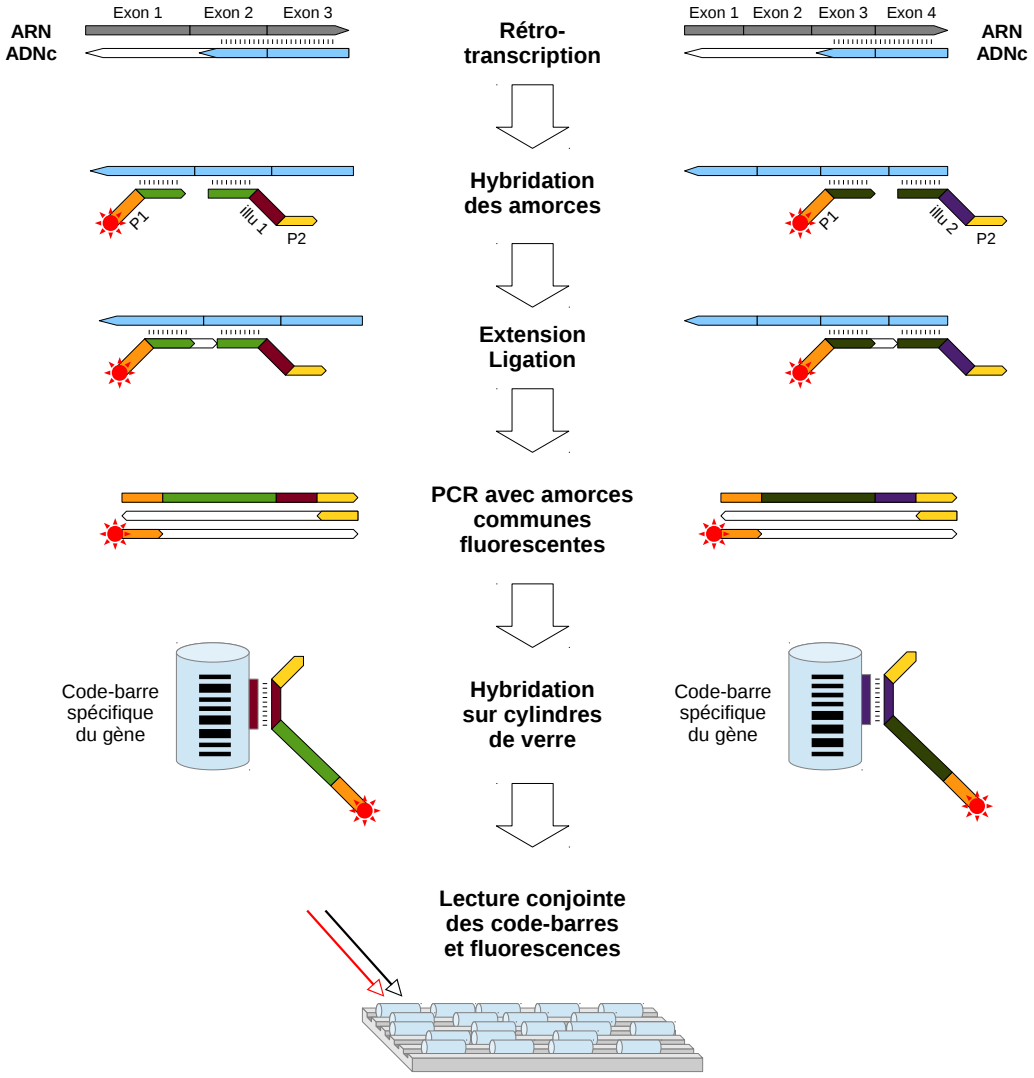


Figure 2.3 – Principe de la technologie DASL

2.3.2 RT-MLPA

La RT-MLPA (*Reverse-Transcriptase Multiplex Ligation dependent Probe Amplification*) permet de mesurer l'expression de plusieurs dizaines de gènes dans un échantillon, selon un principe qui a grandement inspiré le DASL d'Illumina (§2.3.1).

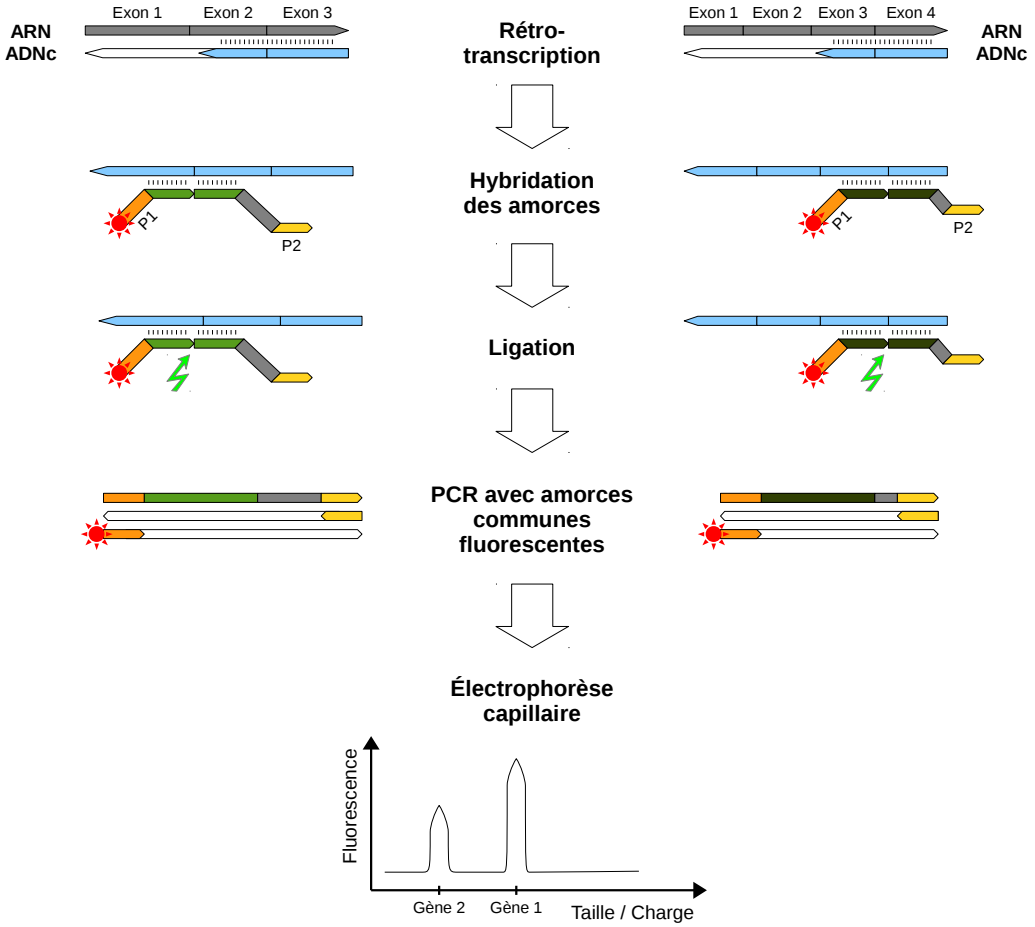


Figure 2.4 – Principe de la RT-MLPA

On retrouve le même couple d'oligonucléotides propre à chaque gène cible, placé au niveau d'une jonction exonique (Figures 2.3 et 2.4). L'étape d'extension n'est cependant pas nécessaire, le choix d'oligonucléotides adjacents permettant une ligation directe. L'association du signal fluorescent à un gène se fait selon la taille du produit de ligation, contrôlé par une séquence leste dont la taille diffère pour chaque couple d'oligonucléotides (gris). La lecture des signaux est finalement réalisée par électrophorèse capillaire, chaque cible produisant un pic migrant à une distance donnée dont la hauteur est proportionnelle à la quantité d'ARN présent dans l'échantillon.

2.3.3 LPS

Le *Linear Predictor Score*, développé par Wright *et al* [40] à partir de travaux précédents [122], est un modèle de classification binaire supervisée reposant sur un ensemble de variables continues (ici l'expression $expr_g$ de plusieurs gènes g). Sa mise en œuvre débute par l'application pour chaque gène d'un test de Student comparant son expression entre les deux groupes d'intérêt (ici GCB et ABC), dans une série d'échantillons pour laquelle cette classification est déjà connue (apprentissage).

Les statistiques de test qui en résultent (t) sont ensuite utilisées comme coefficients du modèle, permettant de calculer pour n'importe quel échantillon un score selon l'équation 2.1.

$$LPS = \sum_{g \in genes} t_g \times expr_g \tag{2.1}$$

Ces statistiques étant signées positivement lorsqu'un des deux groupes présente l'expression moyenne la plus forte et négativement dans le cas contraire, cette somme se verra fortement positive ou négative pour un échantillon qui présente un ensemble de caractéristiques propre à l'un ou l'autre des deux groupes. Considérant le théorème central limite, la distribution de ces scores au sein d'un groupe peut être modélisée par une loi normale dont les paramètres μ et σ^2 seront estimés dans la série d'apprentissage, et directement utilisés pour la prédiction. En effet, on calculera la probabilité pour un échantillon X d'appartenir à un sous-type en comparant les probabilités d'observer son score dans chacune de ces deux distributions (Équation 2.2), la probabilité d'appartenir à l'autre sous-type étant son complémentaire à 1.

$$P(X \in ABC) = \frac{\phi(LPS_X, \hat{\mu}_{ABC}, \hat{\sigma^2}_{ABC})}{\phi(LPS_X, \hat{\mu}_{ABC}, \hat{\sigma^2}_{ABC}) + \phi(LPS_X, \hat{\mu}_{GCB}, \hat{\sigma^2}_{GCB})} \tag{2.2}$$

2.4 Classification GCB-ABC par DASL

2.4.1 Première approche : *clustering* non-supervisé

Parmi les 96 gènes dont nous avions choisi de mesurer l'expression en DASL se trouvaient 19 des gènes sélectionnés par Wright *et al* comme les plus discriminants entre LDGCB GCB et ABC [40]. Ces gènes se répartissent en 12 gènes sur-exprimés dans les ABC (*FOXP1*, *SH3BP5*, *IRF4*, *IL16*, *BLNK*, *ENTPD1*, *ETV6*, *CCND2*, *PIM1*, *FUT8*, *IGHM* et *PTPN1*) et 7 sur-exprimés dans les GCB (*MME*, *LRMP*, *NEK6*, *BCL6*, *LMO2*, *ITPKB* et *MYBL1*).

Ces sous-types ayant originellement été mis en évidence par *clustering* hiérarchique non-supervisé [36], nous avons en première intention fait le choix d'appliquer la même méthode [123, 124], bien que le nombre de gènes soit réduit (plusieurs centaines de clones pour les puces Lymphochip de la publication d'origine contre 19 gènes ici). Comme les études précédentes [36, 74], nous avons retenu l'opposé du coefficient de corrélation de Pearson comme distance, et la méthode d'agglomération de Ward pour la construction des arbres [125]. Les *clusters* ont pu être identifiés *a*

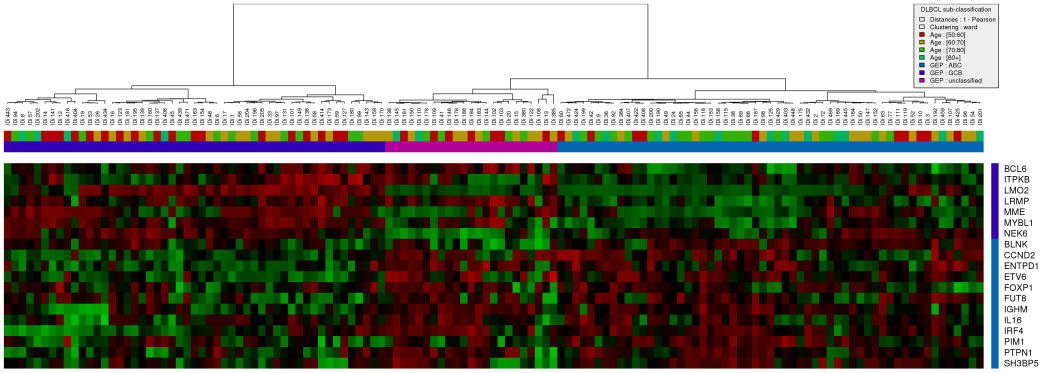


Figure 2.5 – *Clustering* de 131 LDGCB basé sur 19 gènes GCB / ABC

Figure extraite de Mareschal *et al*, Haematologica 2011 [123]

posteriori en fonction du niveau d'expression moyen des gènes GCB et ABC observés pour chacun d'eux, en plus d'un troisième *cluster* indéterminé (Figure 2.5).

L'application de cette méthodologie à 205 LDGCB, issus de la tumorothèque du centre et répartis sur 5 runs DASL, a contribué à de nombreuses découvertes :

- L'augmentation progressive de la proportion de LDGCB de type ABC après 50 ans et à mesure que l'âge au diagnostic augmente [123], une observation confirmée par nos soins et par une équipe indépendante [126] dans deux autres séries.
- L'expression d'un BCR d'isotype primaire (IgM ou IgD) chez les LDGCB ABC et d'isotype secondaire (IgG ou IgA) chez les GCB, en dépit d'une commutation isotypique de l'allèle non-exprimé chez les ABC [127].
- La valeur pronostique du TEP à mi-traitement dans les LDGCB, indépendante du sous-type GCB / ABC [124].
- L'inactivation de *CDKN2A* par co-occurrence de délétions et méthylations du promoteur chez les LDGCB ABC et dans une moindre mesure par méthylation seule chez les GCB [128].
- La répression transcriptionnelle de *CD70* par méthylation du promoteur chez les LDGCB ABC et par délétion chez les GCB, faisant du niveau d'expression de ce gène un facteur pronostique indépendant du sous-type [129].

2.4.2 Seconde approche : modèle LPS

Bien que l'approche non-supervisée offrait une preuve de concept convaincante, l'instabilité inhérente à cette technique [39] rend son application difficile en clinique. Même en reclusterisant chaque échantillon séparément au sein d'une série bien caractérisée (série d'entrainement), un ou plusieurs des 3 *clusters*attendus se retrouvent fréquemment divisés ou mélangés, rendant la conclusion délicate.

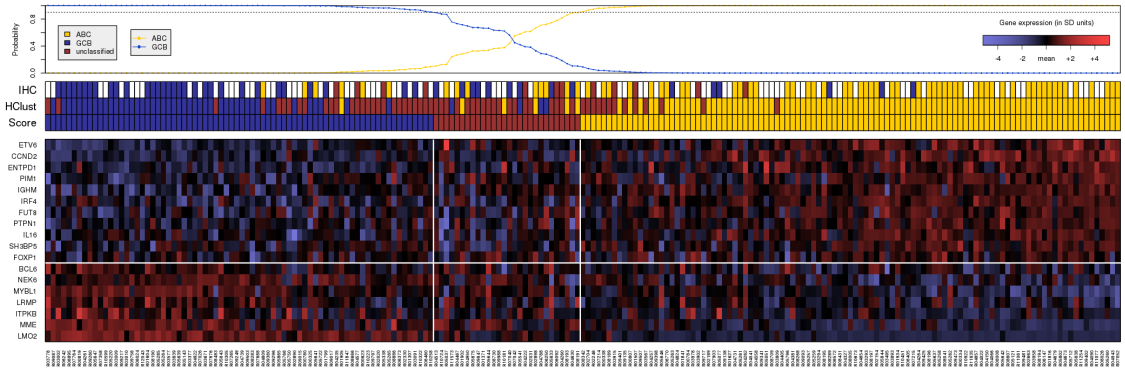


Figure 2.6 – Classification GCB / ABC par LPS dans 189 LDGCB

Figure extraite de Bohers *et al*, Genes Chromosomes Cancer 2014 [130]

Nous nous sommes donc tournés vers des modèles supervisés, tels que le modèle LPS utilisé par le classifieur de référence (§2.3.3). Face à l'absence de réel « *gold standard* » dans cette série, le modèle a été entrainé sur les résultats d'un *clustering* hiérarchique couplés aux résultats IHC de l'algorithme de Hans (Figure 2.6). À défaut de garantir une classification parfaitement fidèle aux techniques de référence, l'utilisation de ce modèle a au moins permis d'assurer de plus grandes robustesse et simplicité d'utilisation à notre outil [130]. Cette approche offre de plus un contrôle direct sur la taille du groupe indéterminé, le seuil de confiance utilisé permettant de réduire ce dernier en acceptant des ressemblances plus ténues avec les *clusters* GCB et ABC typiques utilisés pour l'entrainement. Le seuil de 90% proposé par Wright *et al* nous a ainsi permis de passer de 59 non classés en *clustering* à 28 après modélisation LPS [130]. Le niveau de confiance attribué à chaque échantillon par ce modèle est de plus une information précieuse au clinicien dans l'évaluation de la balance bénéfice-risque d'un éventuel traitement ciblé.

Cette seconde classification a été mise à profit dans deux projets :

— La caractérisation des mutations récurrentes de 5 gènes fréquemment mutés dans les LDGCB (§4.1.3) : *MYD88* (L265P chez 22% des ABC), *CD79A* et *CD79B* (16% des ABC), *CARD11* (11% des LDGCB) et *EZH2* (Y641 chez 24% des GCB). Il est apparu que la mutation d'au moins un gène impliqué dans la voie NFκB constituait un facteur pronostique chez les ABC [130].

— La mise en évidence du déséquilibre de méthylation H3K27 par IHC, y compris chez des patients ABC sans mutation d'*EZH2*. Le niveau d'expression d'*EZH2* est également apparu comme un facteur pronostique chez les patients ABC, réanimant l'intérêt porté à ce gène dans ce sous-type [131].

2.5 Classification GCB-ABC par RT-MLPA

2.5.1 Choix des gènes

Disposant en plus de la série de données mise à disposition par l'équipe de Lenz et Wright (GSE11318 [22, 132]) d'une seconde série de 225 puces Affymetrix U133+2, nous avons choisi de reconsidérer les gènes à inclure dans ce classifieur. En appliquant la méthodologie LIMMA [133] de manière indépendante dans ces deux séries, nous avons pu mettre en évidence 5802 *probesets* discriminant significativement les LDGCB ABC et GCB dans ces deux séries. En nous concentrant en priorité sur les gènes les plus significatifs présentant des distributions bimodales (Figure 2.7), nous avons choisi 4 gènes sur-exprimés chez les ABC (*IRF4*, *TNFRSF13B*, *IGHM* et *FOXP1*) et 4 gènes sur-exprimés chez les GCB (*LMO2*, *NEK6*, *MYBL1* et *BCL6*), plus 6 gènes susceptibles de présenter un intérêt diagnostique ou pronostique (*MS4A1* [134], *CCND1* [135], *CCND2* [136], *TNFRSF9* [98], *MYC* et *BCL2* [108]).

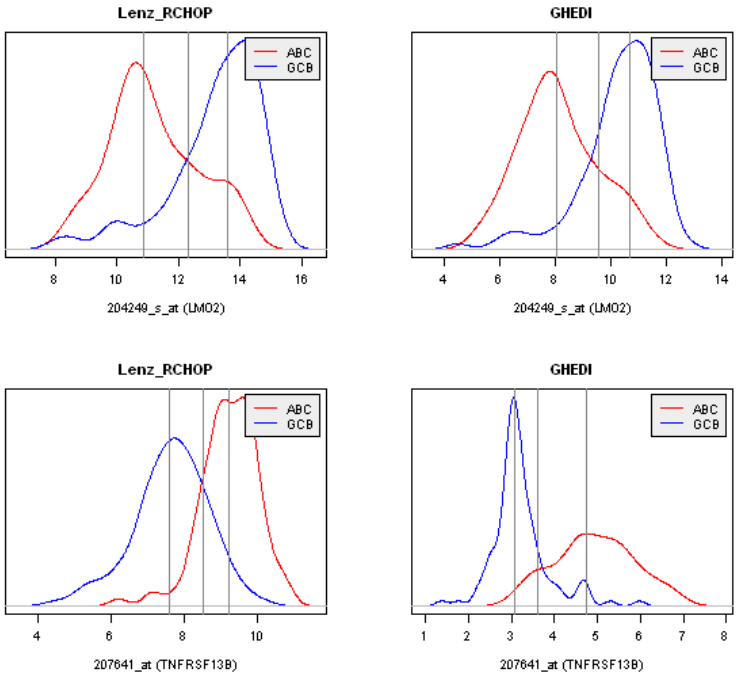


Figure 2.7 – Expression différentielle GCB / ABC de *LMO2* et *TNFRSF13B*

Les densités de distribution des valeurs d'expression de 2 gènes (*LMO2* en haut et *TNFRSF13B* en bas) dans 2 séries (GSE11318 à gauche et GHEDI à droite) sont présentées ici, pour les sous-types ABC (rouge) et GCB (bleu). L'échelle en abscisse correspond au logarithme à base 2 du niveau d'expression d'un *probeset* de la puce Affymetrix U133+2 représentatif du gène. L'échelle en ordonnée n'est pas représentée dans la mesure où elle varie d'un gène et d'un groupe à l'autre, et que seule la répartition de cette distribution le long des valeurs d'expression était recherchée. Les traits gris verticaux représentent les moyennes de chaque sous-type et de la série complète.

2.5.2 Développement de l'outil

Afin de maximiser la portabilité de notre test, nous avons choisi de développer un outil « clé en main » pour l'analyse des résultats. Une fois les différentes phases « humides » terminées, l'électrophorèse capillaire est réalisée au moyen d'un analyseur de fragment de la marque *Applied Biosystems*. Cet appareil produit un électrophorégramme numérique au format FSA, qui sera pris en charge par le logiciel que nous avons développé. La première étape de l'analyse qui s'en suit consiste à dériver des pics correspondant au marqueur de taille un modèle linéaire faisant le lien entre durée de migration et taille de fragment, afin d'identifier les pics. La seconde étape consiste à mesurer la hauteur des pics correspondant aux différents gènes analysés, et à les normaliser par la hauteur moyenne de l'ensemble des pics.

La prédiction du sous-type est enfin réalisée selon le modèle LPS, dont les paramètres ont été déterminés au préalable dans une série d'entrainement pour laquelle les sous-types GCB et ABC avaient été prédits par une technique de référence (le DASL en l'occurrence). Nous avons choisi de ne pas appliquer la normalisation suggérée par Wright *et al* [40], dans la mesure où elle a tendance à reproduire artificiellement les mêmes proportions d'échantillons GCB et ABC que dans la série de référence, et qu'elle nécessite d'être appliquée par *batches* plutôt qu'individuellement (ce dernier cas étant plus proche d'une activité clinique continue). Nous avons simplement choisi d'ajouter au modèle présenté en équation 2.1 un paramètre supplémentaire m_g correspondant au niveau d'expression moyen de chaque gène g dans la série d'entrainement (Équation 2.3).

$$LPS = \sum_{g \in genes} \frac{t_g}{m_g} \times expr_g \tag{2.3}$$

2.5.3 Résultats et discussion

Le développement du test a été réalisé sur 195 échantillons congelés issus du Centre Henri Becquerel, dont 115 bénéficiaient d'une classification GCB ou ABC par DASL (excluant les indéterminés). Parmi ces 115 échantillons, 50 ont servi à entrainer le modèle et 65 à le valider (80% de concordance parfaite, plus 13,8% d'indéterminés). Les 80 échantillons sans « *gold standard* » ont permis de démontrer une concordance avec la classification IHC de Hans dans 82% des cas, avec la particularité attendue d'associer la majorité des indéterminés RT-MLPA au statut « non-GCB » en IHC. Enfin l'effet pronostique attendu a bien été retrouvé entre GCB et ABC (p=0,004 en PFS), de même que pour l'expression univariée de plusieurs gènes (*LMO2*, *BCL6* et *TNFRSF13B*) et la sur-expression conjointe de *MYC* et *BCL2*.

Ces résultats suggèrent que le test développé a enfin atteint un niveau de maturité et de précision suffisant pour sa diffusion dans un contexte clinique. Sa simplicité de mise en œuvre et son faible coût en feront un outil précieux dans la stratification des patients atteints de LDGCB, aussi bien dans un contexte d'essai clinique (tel qu'actuellement discuté au sein du LYSA) que de recherche (tel qu'appliqué aujourd'hui dans notre unité).

2.5.4 Article publié (JMD 2015)

Le développement et la validation du test RT-MLPA a fait l'objet d'une publication dans *the Journal of Molecular Diagnostics* [137] en avril 2015 (IF 4,85). Cet article a fait l'objet d'un communiqué de presse de l'éditeur (16/04/2015) repris par la presse régionale [138]. Il a enfin été présenté en communications orales aux congrès annuels des sociétés américaine (ASH, décembre 2013 [139]) et française (SFH, mars 2014 [140]) d'hématologie, ainsi qu'aux journées scientifiques du Cancéropôle Nord-Ouest (mai 2014) où il a reçu le prix « Jeunes chercheurs » décerné par l'ARC [141].

The Journal of Molecular Diagnostics, Vol. 17, No. 3, May 2015

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Accurate Classification of Germinal Center B-Cell–Like/Activated B-Cell–Like Diffuse Large B-Cell Lymphoma Using a Simple and Rapid Reverse Transcriptase–Multiplex Ligation-Dependent Probe Amplification Assay

A CALYM Study

Sylvain Mareschal,* Philippe Ruminy,* Cristina Bagacean,† Vinciane Marchand,* Marie Cornic,*‡ Jean-Philippe Jais,§ Martin Figeac,¶ Jean-Michel Picquenot,*‡ Thierry Jo Molina,‖ Thierry Fest,** Gilles Salles,†† Corinne Haioun,‡‡ Karen Leroy,§§ Hervé Tilly,* and Fabrice Jardin*

From INSERM U918 and Department of Pathology,‡ Centre Henri Becquerel, Institute for Research and Innovation in Biomedicine, University of Rouen, Rouen, France; the Department of Clinical Hematology,† Prof. Dr. Ion Chiricuta Oncology Institute, Cluj-Napoca, Romania; INSERM UMRS 872§ and Department of Pathology,‖ APHP Necker Hospital, Université Paris Descartes, Paris, France; the Functional Genomic Platform,¶ UDSL, IFR114, IRCL, Lille, France; the INSERM U917,** Laboratoire d'Hématologie, CHU Pontchaillou, Rennes, France; the CNRS UMR 5239,†† ENS, HCL, Faculté de Médecine Lyon Sud, Université Claude Bernard, Lyon, France; and the Unité Hémopathies Lymphoïdes‡‡ and INSERM U955,§§ Equipe 09, APHP Hopital Henri Mondor, Université Paris Est, Créteil, France*

Accepted for publication January 26, 2015.

Address correspondence to Philippe Ruminy, Ph.D., INSERM U918, Centre Henri Becquerel, Rue d'Amiens, 76038 Rouen Cedex 1, France. E-mail: philippe.ruminy@chb.unicancer.fr.

Diffuse large B-cell lymphoma, the most common non-Hodgkin lymphoma, is subdivided into germinal center B-cell–like and activated B-cell–like subtypes. Unfortunately, these lymphomas are difficult to differentiate in routine diagnosis, impeding the development of treatments. Patients with these lymphomas can benefit from specific therapies. We therefore developed a simple and rapid classifier based on a reverse transcriptase multiplex ligation-dependent probe amplification assay and 14 gene signatures. Compared with the Affymetrix U133+2 gold standard, all 46 samples (95% CI, 92%–100%) of a validation cohort classified by both techniques were attributed to the expected subtype. Similarly, 93% of the 55 samples (95% CI, 82%–98%) of a second independent series characterized with a mid-throughput gene expression profiling method were classified correctly. Unclassifiable sample proportions reached 13.2% and 13.8% in these cohorts, comparable with the frequency originally reported. The developed assay was also sensitive enough to obtain reliable results from formalin-fixed, paraffin-embedded samples and flexible enough to include prognostic factors such as MYC/BCL2 co-expression. Finally, in a series of 135 patients, both overall ($P = 0.01$) and progression-free ($P = 0.004$) survival differences between the two subtypes were confirmed. Because the multiplex ligation-dependent probe amplification method is already in use and requires only common instruments and reagents, it could easily be applied to clinical trial patient stratification to help in treatment decisions. *(J Mol Diagn 2015, 17: 273–283; http://dx.doi.org/10.1016/j.jmoldx.2015.01.007)*

Diffuse large B-cell lymphoma (DLBCL) is the most frequent non-Hodgkin lymphoma, accounting for 30% to 40% of diagnoses.[1] Its prognosis has improved significantly over the past decade, predominantly because of the addition of the rituximab antibody to chemotherapy regimens.[2]

Supported by grants from the French National Cancer Institute (INCA), the 'Ligue contre le Cancer (comité Seine Maritime),' the 'Agir avec Becquerel' association, the GEFLUC association, and the CALYM consortium.

S.M. and P.R. contributed equally to this work.

Disclosures: A patent governing this method was filed by S.M., P.R., C.B., V.M., and F.J. (application number EP14305350; March 11, 2014).

Copyright © 2015 American Society for Investigative Pathology and the Association for Molecular Pathology.
Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmoldx.2015.01.007

Unfortunately, approximately one-third of the patients do not respond to therapies or relapse rapidly, and most of them rapidly succumb to the disease.[3] These differences in clinical evolutions can, at least in part, be explained by the heterogeneity of this tumor which can be classified into two major subtypes with different outcomes.[4] The first, termed germinal center B-cell–like (GCB), develops from germinal center B cells. It is the most curable, with a 5-year overall survival (OS) rate of nearly 75%. The second, termed activated B-cell–like (ABC), develops from late GC B cells or plasmablasts. It is more aggressive and can be cured in only approximately 30% of patients.[5]

Even though this GCB/ABC cell of origin classification was described >10 years ago, it still has little influence on clinical practices. However, recent advances have suggested that these tumors could soon benefit from targeted therapies.[6,7] Signaling through the B-cell receptor was, for example, shown to be essential for ABC DLBCL cell survival, and specific inhibitors directed against major components of the NF-κB pathway should rapidly become available in the clinics. Similarly, GCB DLBCLs could soon benefit from inhibitors of the EZH2, BCL2, or BCL6 oncoproteins. Unfortunately, these lymphomas are morphologically undistinguishable in routine diagnosis, which is a major problem for the development of these therapies. Furthermore, array-based gene expression profiling (GEP), which is considered the gold standard for discriminating these tumors, remains poorly transposable to routine diagnosis, and the surrogate immunohistochemical (IHC) algorithms that have been proposed are often considered poorly reliable.[8–10] Other methods, which are based on the evaluation of limited sets of genes, such as real-time PCR,[11,12] quantitative nuclease protection,[13] whole genome expression profiling,[14] or the more recently described digital gene-expression assays based on the NanoString technology,[15,16] were proposed to address this issue. However, all these methods have limitations, such as poor efficiency on formalin-fixed, paraffin-embedded (FFPE) samples or the requirement of dedicated platforms, which are not available in most laboratories.

To address this issue, we have developed a robust reverse transcriptase–multiplex ligation-dependent probe amplification (RT-MLPA)-based assay[17] which allows for a rapid and accurate classification of GCB and ABC DLBCLs in routine diagnosis practice.

Materials and Methods

Patients

A total of 259 DLBCL patients were included in this study (Figure 1 and Supplemental Table S1). All samples were obtained after receiving patient consent, and this study was approved by our institutional review board. A first series of 195 patients from a single institution (Centre Henri Becquerel, Rouen, France) was used to train and validate the GCB/ABC classifier. A GCB/ABC classification previously

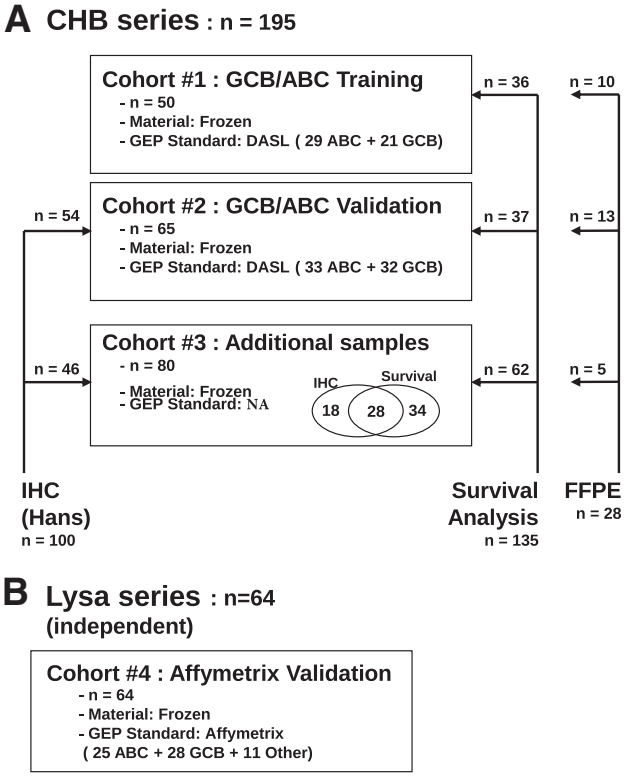


Figure 1 Description of the analyzed cohorts. General structure and overlaps between the different cohorts are indicated, along with their main characteristics (sample count, material type, GEP standard when available). Cohorts from our institution (**A**) are separated from the external cohort (**B**) to highlight their independence. ABC, activated C-cell–like; CHB, Centre Henri Becquerel; DASL, cDNA-mediated Annealing, Selection, extension and Ligation method; FFPE, formalin-fixed, paraffin-embedded; GCB, germinal center B-cell–like; GEP, gene expression profiling; IHC, immunohistochemical; Lysa, Lymphoma Study Association.

determined with a Veracode DASL assay (cDNA-mediated Annealing, Selection, extension and Ligation method; Illumina, San Diego, CA)[18] was available for 115 patients. An IHC classification was available for 100 cases and survival data for 135 of them. A second series of 64 patients included in the GHEDI (Deciphering the Genetic Heterogeneity of Diffuse large B-cell lymphoma in the rituximab era) study program conducted by the Lymphoma Study Association group and previously analyzed by U133+2 GEP arrays (Affymetrix, Santa Clara, CA) was used as an external validation set.[19] For all patients, the inclusion criteria were *de novo* DLBCL with available RNA extracted from frozen lymph node biopsies at the time of diagnosis. The diagnoses were determined according to the World Health Organization 2008 criteria by expert pathologists.[1] Primary mediastinal DLBCLs and T-cell–rich DLBCLs were excluded.

RNA Extraction

Total RNA was extracted from frozen biopsy fragments by using the RNA NOW kit (Biogentex, Seabrook, TX). The RNA samples were subsequently stored at −80°C in water.

Table 1 Sequences of the RT-MLPA Oligonucleotides

Oligonucleotides	Sequences
5′ RT-MLPA probes	
NEK6_E2L	5′-GTGCCAGCAAGATCCAATCTAGACCTGTGCATCCTCCTGACCCACAG-3′
IRF4_E6L	5′-GTGCCAGCAAGATCCAATCTAGATCTGCCGAAGCCTTGGCGTTCTCAG-3′
IGHM_E2L	5′-GTGCCAGCAAGATCCAATCTAGATGCGTCCTCCATGTGTGGCCCCG-3′
CCND1_E3L	5′-GTGCCAGCAAGATCCAATCTAGATACCTTCGTTGCCCTCTGTGCCACAG-3′
LMO2_E5L	5′-GTGCCAGCAAGATCCAATCTAGACGGAAGCTCTGCCGGAGAGACTATCTCAG-3′
FOXP1_E10L	5′-GTGCCAGCAAGATCCAATCTAGACCCTTCCCCTTCAACCTCTTGCTCAAG-3′
TNFRSF9_E2L	5′-GTGCCAGCAAGATCCAATCTAGATACGGACCTGTGACATATGCAGGCAGTGTAAAG-3′
BCL6_E3L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTCATAAAACGGTCCTCATGGCCTGCAG-3′
TNFRSF13B_E2L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTAGCGCACCTGTGCAGCCTTCTGCA-3′
CCND2_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTGACCTTCATTGCTCTGTGTGCCACCG-3′
MYC_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTTCGGGTAGTGGAAAACCAGCAGCCTC-3′
MYBL1_E10L	5′-GTGCCAGCAAGATCCAATCTAGACCAGAATTTGCAGAGACTCTAGAACTTATTGAATCT-3′
BCL2_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTACCCTGGATCCAGGATAACGGAGGCTGG-3′
MS4A1_E5F*	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTATTCTTCATGAGGGAATCTAAGACTTTGGGG-3′
3′ RT-MLPA probes	
NEK6_E3R	5′-Pho-AGGCATCCCAACACGCTGTCTTTTCCAACCCTTAGGGAACCC-3′
IRF4_E7R	5′-Pho-ACTGCCGGCTGCACATCTGCCTGTATCCAACCCTTAGGGAACCC-3′
IGHM_E3R	5′-Pho-ATCAAGACACAGCCATCCGGGTCTTCTACTATCCAACCCTTAGGGAACCC-3′
CCND1_E4R	5′-Pho-ATGTGAAGTTCATTTCCAATCCGCCCTTACTTCCAACCCTTAGGGAACCC-3′
LMO2_E6R	5′-Pho-GCTTTTTGGGCAAGACGGTCTCTGCTACTATCCAACCCTTAGGGAACCC-3′
FOXP1_E11R	5′-Pho-GCATGATTCCAACAGAACTGCAGCAGCTACTACTACTCCAACCCTTAGGGAACCC-3′
TNFRSF9_E3R	5′-Pho-GTGTTTTCAGGACCAGGAAGGAGTGTTCCTACTCCAACCCTTAGGGAACCC-3′
BCL6_E4R	5′-Pho-TGGCCTGTTCTATAGCATCTTTACAGACCAGTTGTCCAACCCTTAGGGAACCC-3′
TNFRSF13B_E3R	5′-Pho-GGTCACTCAGCTGCCGCAAGGAGCTACTACTACTACTCCAACCCTTAGGGAACCC-3′
CCND2_E2R	5′-Pho-ACTTTAAGTTTGCCATGTACCCACCGTCGATACTACTATCCAACCCTTAGGGAACCC-3′
MYC_E2R	5′-Pho-CCGCGACGATGCCCCTCAACGTTATACTACTACTACTATCCAACCCTTAGGGAACCC-3′
MYBL1_E11R	5′-Pho-GATCCTGTAGCATGGAGTGACGTTACCAGTTTTTACTACTTCCAACCCTTAGGGAACCC-3′
BCL2_E2R	5′-Pho-GATGCCTTTGTGGAACTGTACGGCCTACTACTACTACTACTTCCAACCCTTAGGGAACCC-3′
MS4A1_E6R	5′-Pho-GCTGTCCAGATTATGAATGGGCTCTTCCACTACTACTACTATCCAACCCTTAGGGAACCC-3′
Competitors	
NEK6_comp	5′-AGGCATCCCAACACGCTGTCTTT-3′
IGHM_comp	5′-ATCAAGACACAGCCATCCGGGTCTTC-3′

The sequences of the forward and 5′ phosphorylated reverse primers are presented with the symbol of the gene assessed, along with the two oligonucleotides used as competitors. The underlined sequence is designed to hybridize with PCR primers for a common amplification, whereas the rest of the sequence is a combination of a gene-specific sequence and spacer.

*MS4A1 encodes CD20.

RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

For 30 cases, RNA samples were extracted from FFPE biopsies by using the Absolutely RNA FFPE Kit (Agilent Technologies, La Jolla, CA).

Immunohistochemistry

To build tissue microarrays (Alphelys, Plaisir, France), three representative cores (0.6 mm) of FFPE tissues were used for each case. IHC staining was performed on 4-μm sections by using standard procedures. Antigen retrieval was performed on PTLink (Dako, Glostrup, Denmark) with a pH 9 buffer. CD10 (clone 56C6, 1/50e; Dako), BCL6 (clone PGB6P, 1/10e; Dako), and MUM1 (clone MUM1p, 1/100e; Dako) were stained with an AUTOSTAINER Link48 (Dako) with the envision flex system. GCB and non-GCB phenotypes were defined with the decision tree with indicated cutoffs established by Hans et al.[20]

RT-MLPA Probe Design and Mix

The RT-MLPA probes that were used are listed in Table 1. To avoid unwanted amplifications of genomic DNA, all of the probes were designed across exon−exon boundaries. They are made of a gene-specific region complementary to the cDNA target and of a tail to allow the final PCR amplification. All 5′ probes have a common tail at their 5′ end (GTGCCAGCAAGATCCAATCTAGA), which corresponded to the universal primer, U1. All 3′ probes have a different common tail at their 3′ end (TCCAACCCTTAGGGAACCC), which is complementary to the universal primer, U2. Additional spacers (TAC repeats) were inserted between the tails and the gene-specific regions to allow the separation and the identification of the PCR products according to their size. The 3′ probes were ordered with phosphorylated 5′ ends to allow for the ligation reactions.

Table 2 Candidate Genes Considered for Inclusion in the RT-MLPA Predictor

Gene	Probe set	High in	GHEDI FDR	Lenz FDR	RT-MLPA	RT-MLPA FDR
BATF	205965_at	ABC	1.95×10^{-11}	1.05×10^{-35}		
CCND2	200953_s_at	ABC	1.12×10^{-6}	8.92×10^{-14}		
FOXP1	224838_at	ABC	5.97×10^{-28}	1.35×10^{-28}	+	9.98×10^{-3}
FUT8	203988_s_at	ABC	5.19×10^{-15}	4.71×10^{-26}		
IGHM	209374_s_at	ABC	1.62×10^{-10}	2.29×10^{-21}	+	9.93×10^{-5}
IRF4	204562_at	ABC	3.92×10^{-14}	2.55×10^{-17}	+	9.93×10^{-5}
LIMD1	222762_x_at	ABC	3.28×10^{-16}	2.00×10^{-34}		
SH3BP5	201811_x_at	ABC	1.19×10^{-22}	9.98×10^{-25}		
SLA	203761_at	ABC	5.10×10^{-17}	5.64×10^{-31}		
TNFRSF13B	207641_at	ABC	3.40×10^{-11}	4.44×10^{-20}	+	9.93×10^{-5}
BCL6	203140_at	GCB	5.13×10^{-6}	2.00×10^{-34}	+	8.93×10^{-4}
LMO2	204249_s_at	GCB	1.25×10^{-10}	1.32×10^{-23}	+	1.90×10^{-10}
MME	203434_s_at	GCB	1.36×10^{-5}	6.08×10^{-21}		
MYBL1	213906_at	GCB	8.44×10^{-10}	2.23×10^{-27}	+	1.12×10^{-4}
NEK6	223158_s_at	GCB	1.06×10^{-10}	7.01×10^{-29}	+	1.03×10^{-5}
TNFRSF9	207536_s_at	GCB	9.81×10^{-4}	6.39×10^{-2}		

GCB and ABC discriminating genes [reported as false discovery rates (FDR) from two distinct Lineal Models for Microarray and RNA-Sequencing Data analyses] in two independent series of Affymetrix U133+2 arrays (GHEDI and Lenz et al[23,25]) were considered for inclusion in the predictor. The considered Affymetrix probe set, as well as the final decision to retain the gene (+) or not according to its performances in RT-MLPA, are indicated.

ABC, activated B-cell–like; FDR, false discovery rate; GCB, germinal center B-cell–like; GHEDI, Deciphering the Genetic Heterogeneity of Diffuse large B-cell lymphoma in the rituximab era; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

All of the RT-MLPA probes used in this study were synthetic oligonucleotides, ordered at the lowest purification grade (high-purity salt-free), and at the minimal synthesis scale (0.01 μmol for the left probes and 0.05 μmol for the right phosphorylated probes) (Eurofins MWG Operon, Ebersberg, Germany). For two genes, we also designed competitor probes identical to the MLPA probes, but without the PCR tails, which were used to normalize the amplification signals.

To prepare the MLPA probe mix, 10-μmol/L dilutions of each probe and competitors were prepared in Tris 10 mmol/L and EDTA 1 mmol/L (TE). The two competitors were mixed with their corresponding probes at a 1 (MLPA probe) to 4 (competitor) ratio for the *IGHM* and *NEK6* genes. Then, 2 μL of each probe dilution ($n = 26$) or probe + competitor mix dilution ($n = 2$) were mixed together to obtain a final volume of 56 μL. An equal volume of 1× TE was then added to obtain a volume of 112 μL. The final dilution was obtained by diluting 11.2 μL of the concentrated probe mix in a final volume of 1 mL of TE buffer.

RT-MLPA Procedure

The RT-MLPA protocol was adapted from Eldering et al.[17] A schematic workflow of the experimental procedure is provided in Supplemental Figure S1. The most important modifications we made were the use of random hexamer primers instead of gene-specific primers for cDNA synthesis, the reduction of the hybridization step from 16 hours to 1 hour, and the use of synthetic oligonucleotides as MLPA probes.

All RT-MLPA reactions were performed in 200-μL tubes in a thermocycler with a heated lid. Approximately 1 μg of total RNAs obtained from DLBCL biopsies were first diluted in 4 μL of RNase-free water and kept on ice. After the addition of 7.5 μL of reverse transcription mix (2.5 μL of 5× Moloney murine leukemia virus–reverse transcription buffer, 1 μL of 100 mmol/L dithithreitol, 2 μL of 10 mmol/L dNTPs, and 2 μL of 100 μmol/L random primers), the samples were heated for 1 minute at 80°C to melt the secondary structure of the mRNA, incubated for 5 minutes at 37°C to allow the hybridization of the random primers and cooled at 4°C. Next, 1 μL of Moloney murine leukemia virus–reverse transcriptase was added, and the samples were incubated for 15 minutes at 37°C to synthesize the cDNA, heated for 2 minutes at 98°C to inactivate the enzyme, and cooled at 4°C. Then, 5 μL of cDNA was transferred into a new tube, and 3 μL of a RT-MLPA probe mix (1.5 μL of SALSA-MLPA Buffer + 1.5 μL of final dilution probe mix) were added.

The samples were heated for 2 minutes at 95°C and incubated for 1 hour at 60°C to allow the annealing of the MLPA probes to their cDNA targets. The ligation of the annealed oligonucleotides was performed by cooling all samples at 56°C, adding 32 μL of a ligation mix (3 μL of SALSA-Ligase 65 Buffer A, 3 μL of SALSA-Ligase Buffer B, 25 μL of RNase-free water, 1 μL of SALSA-Ligase 65), and incubating the mixed samples for 15 minutes at 56°C. The ligase enzyme was next inactivated by heating for 5 minutes at 98°C. For the final PCR amplification, 5 μL of each ligation product was added to 15 μL of a PCR mix [10 μL Thermo Scientific Extensor Hi-Fidelity PCR Master Mix (Thermo Scientific, Marietta, OH), 1 μL of 10 μmol/L primer U1, 1 μL of 10 μmol/L FAM-labeled primer U2,

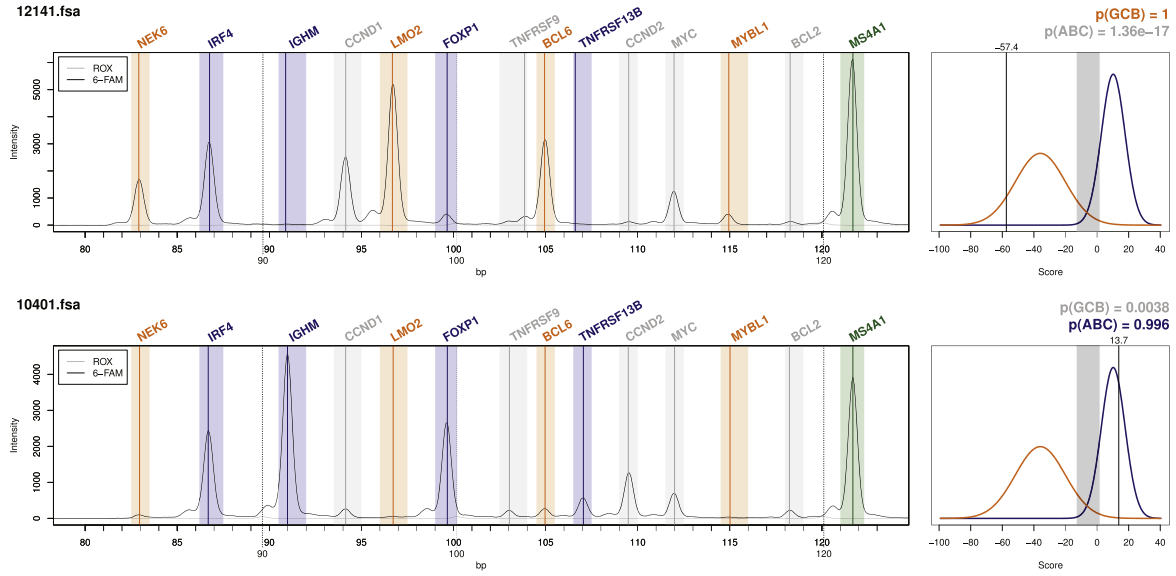


Figure 2 Representative RT-MLPA profiles. RT-MLPA profiles generated by the dedicated software are presented for two representative samples (GCB and ABC). For each sample, a fragment-analysis profile (intensity of fluorescence as a function of the PCR fragment sizes) is provided in the **left panel**, with intervals scanned for gene-related peaks highlighted by various colors [blue for ABC-related genes, orange for GCB-related genes, green for the *MS4A1* (encodes CD20) internal control, and gray for other diagnostically or prognostically relevant genes]. Size markers, as used for alignment of profiles and fragment size estimation (ROX channel), are displayed with dotted lines, indicating their theoretical size. The **right panel** illustrates the prediction made by the RT-MLPA predictor, showing the superposition of the score of the sample on the theoretic distributions of scores in the ABC and GCB subgroups. ABC, activated B-cell–like; GCB, germinal center B-cell–like; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

3 μL of RNase-free water]. The amplification was performed with the following program: 6 minutes at 94°C; 35 cycles (30 seconds at 94°C, 30 seconds at 58°C, 30 seconds at 72°C); 4 minutes at 72°C; cooled down to 16°C.

For data analysis, 1 μL of each PCR product was mixed with 1 μL of GeneScan-400 HD ROX size standard and 28 μL of formamide, denatured for 3 minutes at 95°C, and loaded onto an ABI3130 capillary electrophoresis analyzer in GeneScan mode (Applied Biosystems, Warrington, UK). The migration settings were as follows: Pre Run Voltage, 15 kV; Pre Run Time, 180 seconds; Injection Voltage, 1 kV; Injection Time, 3 seconds (reduced to 1 second when fluorescence signals were off-scale), Run Voltage, 10 kV; Run Time, 180 seconds.

Data Processing

To facilitate the interpretation of the results, interfaced software was developed to handle the entire analysis process, from the raw fragment analyzer files (.fsa) to the sample calls and GCB/ABC probabilities. This software, which can be freely downloaded (*http://bioinformatics.ovsa.fr/MLPA*, last accessed January 19, 2015) provides a graphical representation of the profile and of the GCB/ABC classification, and a table of normalized gene expressions. It consists in a Tcl-tk user interface built on an R package (*http://www.r-project.org*), both of which are freely distributed under the GNU Public License 3.

Each channel of the electrophoresis profiles is first corrected for time-related biases, using a LOWESS transformation of the signals over time. To estimate the sizes of the peaks in base pairs, the profiles are then aligned with the 50-, 60-, 90-, 100-, and 120-bp size standards. The ROX profile is smoothed with a Nadaraya-Watson Gaussian kernel, and the time indexes of local maxima are used to fit a linear model that converts time indexes into base pairs. The heights of the different peaks are next measured by looking for the FAM signal maxima in size intervals that are predefined in base pairs. Both the gene and size standard windows were tuned from the graphical profiles and were made available in the software's design file (version 1.8.5). To minimize RNA quantity bias, all expression values are finally divided by the mean of all genes measured in the sample, and the base 2 logarithms of these ratios plus 1 are used as normalized expression values.

Survival Analysis

Survival analysis was restricted to the 135 patients diagnosed between 2001 and 2011 at the Centre Henri Becquerel who were treated by a combination of rituximab and chemotherapy. It was performed with the *survival* package version 2.37.4 for the R software, considering an α risk of 5% as a significance threshold. OS was computed from the day of treatment to death from any cause or right-censored at last follow-up. Progression-free survival (PFS) was computed from the day of treatment to disease progression, relapse, or death from any cause, or right-censored at last follow-up. The survival rates were estimated with the Kaplan-Meier method that provides 95% CIs, and significant differences between groups were assessed by the log-rank test. Continuous univariate analyses were performed with Cox proportional hazards regression models, and the likelihood ratio test *P* values and derived false discovery rates

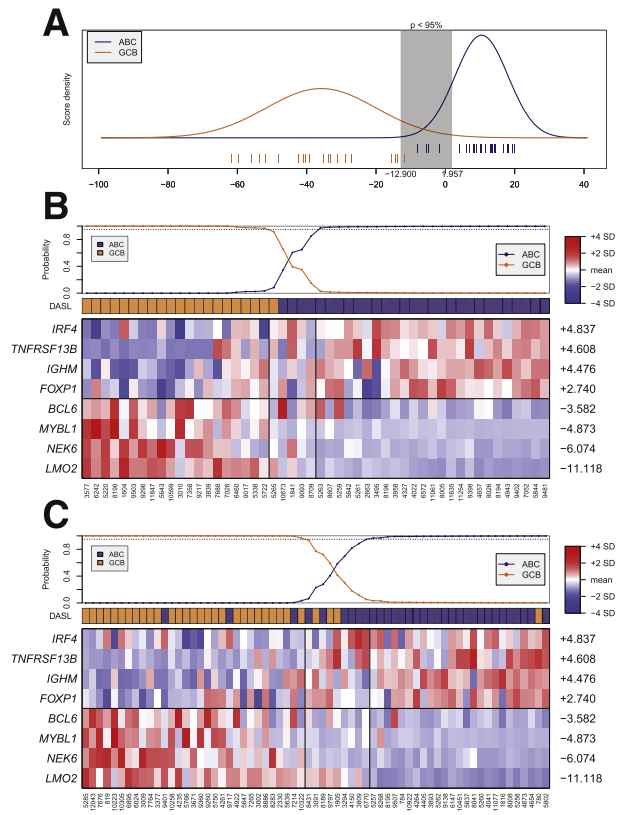


Figure 3 RT-MLPA predictor training and validation. **A:** The scores of the GCB and ABC samples of the training series are presented as vertical bars, beneath the estimated density functions used in the prediction model. **B** and **C:** The expression of the eight genes included in the GCB-ABC predictor in the training (**B**) and validation (**C**) series are presented as heat maps, along with the expected classes computed with the DASL technology (**middle panel**) and the predicted probability to belong to each of the groups (**top panel**). Samples (columns) are ordered by ascending RT-MLPA scores, whereas genes (rows) are ordered by their discriminating power (*t* statistics displayed on the right). **Black vertical bars** split samples according to the group prediction (GCB, unclassified, and ABC from left to right). ABC, activated B-cell–like; DASL, cDNA-mediated annealing, selection, extension, and ligation method; GCB, germinal center B-cell–like; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

(FDRs) are reported.[21] All combinations of thresholds were tested for PFS significance of the MYC^{+} $BCL2^{+}$ subgroup, focusing on the couple of values that led to the most significant segmentation of patients.[22]

Statistical Analysis

Prediction accuracies and unclassified proportions are reported with 95% CIs estimated by the *exact* binomial computation method, as implemented in the binom R package version 1.1-1.

Results

Gene Selection

The candidate genes for inclusion into the GCB/ABC predictor were selected according to their differential expression in two independent series of Affymetrix U133+2 arrays. The first series consisted of 203 arrays made available by Lenz et al[23] in the Gene Expression Omnibus database (*http://www.ncbi.nlm.nih.gov/geo*; Accession number GSE11318). The second consisted of an independent series of 225 arrays classified into the GCB or ABC subtypes, following the recommendations of Wrights et al[24] and a signature of 24 genes.[19] A LIMMA model was built for each series by using R version 2.15.1 (LIMMA package version 3.12.3) to discriminate two groups by using single-channel arrays. Genes with at least one probe set with FDR $< 10^{-12}$ in both series were selected for RT-MLPA testing, beginning with the lowest *P* values. The 16 markers obtained from this initial selection are listed in Table 2.

Different combinations of genes were then tested to finalize the design, and those expressed at the most comparable levels were selected. We finally retained four GCB (*LMO2*, *MYBL1*, *BCL6*, *NEK6*) and four ABC (*IRF4*, *FOXP1*, *IGHM*, *TNFRSF13B*) markers. For two genes expressed at high levels in some samples (*NEK6* and *IGHM*), competitor probes were used to reduce the height of the corresponding peaks in the RT-MLPA profiles. We finally included the *MYC*, *BCL2,* and *TNFRSF9* genes, which were shown to be prognostic in DLBCLs, and the *CCND1*, *CCND2,* and *MS4A1* (encoding CD20) genes as controls. Representative RT-MLPA profiles are presented in Figure 2, and all gene expression values are detailed in Supplemental Table S2.

Class Prediction Model

The classification of samples as GCB or ABC was achieved with an adaptation of the linear predictor score defined by Wright et al[24] and a Bayesian predictor. We adopted the same assumption that the distributions of the scores within the GCB and ABC groups followed normal distributions (Figure 3A), which allowed the estimation of the probability for each case of belonging to one or the other subtype. Scaling factors were applied to balance the overweighting of the most highly expressed genes, a marginal phenomenon in heavily normalized values used in Wright's publication which is no longer negligible with the values manipulated here. To account for these differences in expression ranges, the *t* statistics that were used as gene-specific coefficients were thus further divided by the mean expressions of the corresponding genes in the training series. This correction also allowed the avoidance of the shift and scale normalization to Lymphochip's ranges that was suggested by Wright et al[24] to adapt their model to other platforms, which may lead to incorrect calls in series with GCB/ABC distributions distinct from the original Lymphochip series. Finally, the confidence threshold to consider a sample either as GCB or ABC was changed from 90% (as recommended by Wright et al[24]) to 95%, for better specificity.

The Bayesian predictor was developed with a training series of 50 randomly selected DLBCL cases previously characterized in our laboratory by using a DASL assay (Figure 1) and a 17-gene signature (Figure 3B).[18] As

expected, the average expressions of the eight GCB/ABC discriminant genes were significantly different between the two subtypes in this series (Table 2). The predictor classified 90% of the cases within the expected subtypes (20 GCB and 25 ABC), whereas 10% (1 GCB and 4 ABC) were considered unclassified (95% CIs, 78%−97% and 3%−22%, respectively). In a second independent validation series of 65 patients (from the same institution and characterized with the same DASL assay) (Figure 1), 80% of the cases were classified within the expected subtypes (24 ABC and 28 GCB), 13.8% were considered unclassified (3 GCB and 6 ABC), and four were misclassified (95% CIs, 68%−89% and 7%−25%, respectively) (Figure 3C). When focusing on samples classified by both techniques, 93% of agreement was reached in this second series (95% CI, 82%−98%).

Reproducibility Assessment

We next performed 10 independent repeats for two samples (RNA 9402 and 9260, ABC and GCB, respectively). The calculated scores remained highly stable (mean ± SD, −32.75 ± 0.41 for the sample 9260, 17.88 ± 0.32 for the sample 9402) (Supplemental Figure S2A). To address further the robustness of the assay, we also tested 12 serial one-half dilutions of the second sample (9402), from 400 ng to <1 ng of RNA. The raw fluorescence peak heights for the 14 genes remained stable from 400 to 12 ng of RNA, resulting in a high stability of the normalized expression values (Supplemental Figure S2, C and D), scores (17.84 ± 0.26) (Supplemental Figure S2, B and E), and GCB/ABC probabilities (Supplemental Figure S2E).

Validation on an Independent Cohort of Frozen Samples

We next applied this predictor to a second independent validation series of 64 patients from the GHEDI study (Figure 1), classified within the GCB or ABC subtypes by using Affymetrix U133+2 gene expression data and a Bayesian predictor based on a 24-gene expression signature as previously described.[24] This series comprised 28 GCB, 25 ABC, and 11 unclassified cases. The RT-MLPA assay classified 84.4% (95% CI, 73%− 92%) of these cases within the expected subtypes (24 of 28 GCB, 22 of 25 ABC, and 8 of 11 unclassified; Figure 4). Four GCB and three ABC cases were considered unclassifiable, accounting for 13.2% (95% CI, 5%−25%) of samples classified by the reference technique. Finally, three previously unclassified cases were misclassified into the ABC subtype. When focusing on samples classified by both techniques, none of the 46 remaining samples was in contradiction between this assay and the Affymetrix U133+2 gold standard (95% CI, 92%−100%).

Validation on FFPE Samples

To test further the robustness of this assay, we next compared the results obtained from paired frozen and FFPE

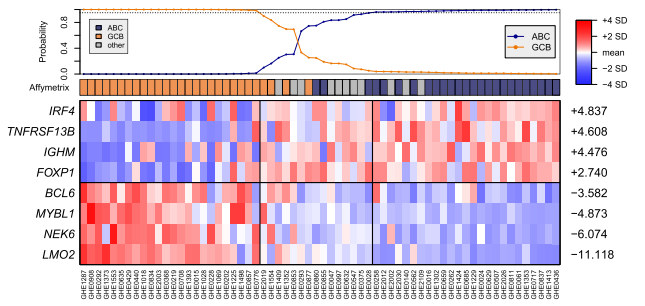


Figure 4 RT-MLPA predictor independent validation on Affymetrix samples. The expression of the eight genes included in the GCB-ABC predictor in an independent series of 64 frozen samples is presented as a heat map, along with the expected classes computed with Affymetrix U133+2 arrays and the predicted probability of belonging to each of the groups. Samples (columns) are ordered by ascending RT-MLPA scores, whereas genes (rows) are ordered by their discriminating power (*t* statistics displayed on the right). **Black vertical bars** split samples according to the group prediction (GCB, unclassified, and ABC from left to right). ABC, activated B-cell–like; DASL, cDNA-mediated annealing, selection, extension, and ligation method; GCB, germinal center B-cell–like; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

biopsies. Thirty RNA samples extracted from FFPE archival materials from the CHB cohort were tested. Except for two extraction failures, interpretable profiles were obtained for all 28 remaining cases. The RT-MLPA assay classified 25 of the FFPE samples (89.3%) within the expected subtypes (11 of 14 GCB, 12 of 12 ABC, and 2 of 2 unclassified), whereas three GCB cases were considered unclassified (Figure 5A). The statistically significant ($P = 1.2^{e-11}$) Pearson's correlation between frozen and FFPE samples is illustrated in Figure 5B.

Extended Series of Samples

We next included 80 additional independent frozen samples, all from the Centre Henri Becquerel, and for which neither DASL nor Affymetrix classification was available (Figure 1). The resulting extended series of 195 cases allowed us to test for the correlations between the RT-MLPA GCB/ABC classification, IHC (100 cases), and survival (135 cases). Excluding the 50 samples of the training cohort, the RT-MLPA predictor classified 63 of the 145 remaining cases into the GCB subtype (43.4%), 56 into the ABC subtype (38.6%), and 26 were considered unclassifiable (17.9%). These proportions correspond to the repartition which is typically observed in DLBCL populations.

Comparison with Immunochemistry

For 100 cases (54 in the validation cohort 1 and 46 in cohort 3), the GC/non-GC origin was addressed by IHC at diagnosis by using Hans algorithm, implemented in our department of pathology in 2005.[20] To avoid artifacts, only cases with positive CD10 internal controls were considered interpretable. In this series, IHC and RT-MLPA were in agreement for 36 of 44 GCB cases (81.8%) and 32 of 38 ABC cases (84.2%). The last 18 cases, considered unclassifiable by RT-MLPA, distributed

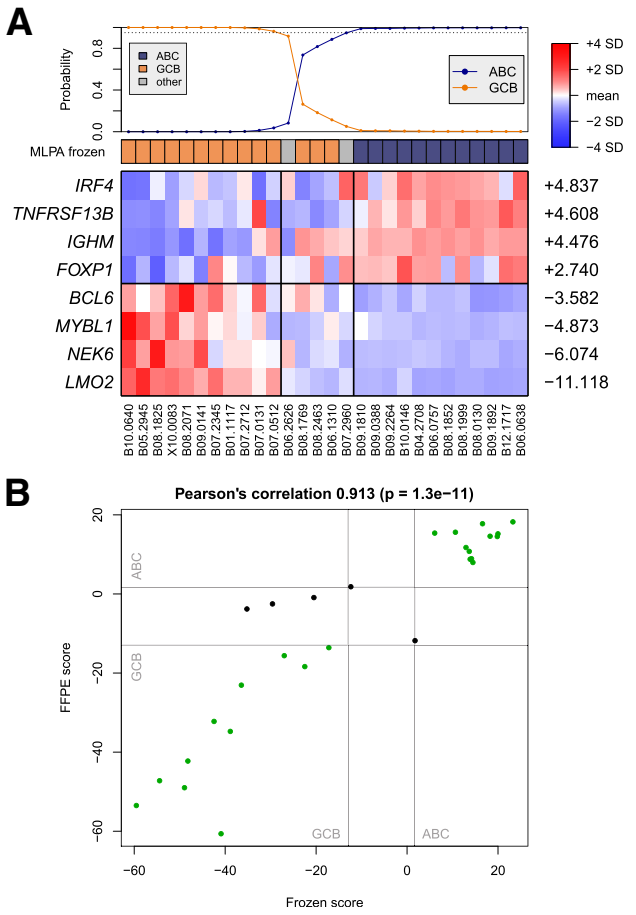


Figure 5 RT-MLPA predictor performances in FFPE samples. **A:** The expression of the eight genes included in the GCB-ABC predictor in a series of 28 FFPE samples is presented as a heat map, along with the expected classes computed with the DASL technology on matching frozen samples and the predicted probability of belonging to each of the groups. The samples (columns) are ordered by ascending RT-MLPA scores, whereas the genes (rows) are ordered by their discriminating power (*t* statistics displayed on the right). The **black vertical bars** split samples according to the group prediction (GCB, unclassified, and ABC from left to right). **B:** RT-MLPA scores measured in FFPE and frozen samples from the same 28 patients are presented as a scatter plot. Score thresholds to classify samples as GCB or ABC with a confidence of ≥95% are materialized by **grey lines. Green circles** are used for samples classified in the same subtype using both samples. **Black circles** are used for samples unclassified in one or the other sample. ABC, activated B-cell–like; DASL, cDNA-mediated annealing, selection, extension, and ligation method; FFPE, formalin-fixed paraffin-embedded; GCB, germinal center B-cell–like; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

within the GCB (4 cases) and non-GCB (14 cases) subtypes defined by IHC (Figure 6). These results are in agreement with what is typically observed between Affymetrix classifications and Hans algorithm, which identifies GCB cases but does not differentiate unclassified cases from ABC cases (both referred as non-GCB cases).[20]

Survival Analysis

To evaluate the prognostic value of this assay, we next focused on the 135 patients diagnosed between 2001 and 2011 at the Centre Henri Becquerel who were treated by a combination of rituximab and chemotherapy. As expected, ABC cases were significantly associated with a poorer PFS ($P = 0.004$) and OS ($P = 0.012$) (Figure 7A). Continuous univariate analysis also confirmed that the expression of several individual genes within this MLPA signature were significantly associated with prognosis (Supplemental Table S3). As expected, high *LMO2* (FDR = 0.006 and 0.012 in PFS and OS, respectively) and *BCL6* (FDR = 0.019 in PFS) expressions were found to be associated with a better prognosis. Similarly, low *TNFRSF13B* expression was linearly correlated with a better prognosis (FDR = 0.036 and 0.012 in PFS and OS, respectively). Simplification toward two-group models that classify patients according to the median expression of the considered gene led to similar results (Figure 7, B–D).

To further illustrate the ability of this assay to retrieve well-established prognosis factors, the survival of MYC^+ $BCL2^+$ patients was compared with the rest of the series. By following previously described methods,[22] the expression thresholds were set at values (0.256 for *BCL2*, 1.284 for *MYC*) that lead to the most significant segmentation of patients in PFS. In total, 14 patients were classified as MYC^+ $BCL2^+$ and showed a particularly poor outcome in OS [27% (95% CI, 11%–66%) versus 67% (95% CI, 58%–76%) at 3 years; $P = 0.0002$] and PFS [14% (95% CI, 4%–52%) versus 56% (95% CI, 47%–66%) at 3 years; $P = 0.0007$] (Figure 7E).

Discussion

Here, we describe a rapid and robust RT-MLPA assay that allows an accurate classification of GCB and ABC DLBCLs. This easy-to-use method allows the rapid evaluation of the relative expression of 14 genes in a single reaction. It only requires a thermocycler, a capillary sequencer, and basic

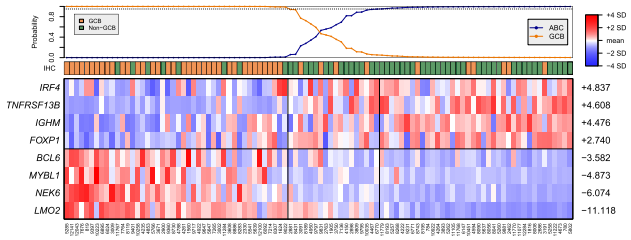


Figure 6 RT-MLPA predictor performances compared with immunohistochemistry. The expression of the eight genes included in the GCB-ABC predictor in a series of 100 frozen samples is presented as a heat map, along with Hans IHC algorithm results, and the predicted probability of belonging to each of the groups. Samples (columns) are ordered by ascending RT-MLPA scores, whereas genes (rows) are ordered by their discriminating power (*t* statistics displayed on the **right**). **Black vertical bars** split samples according to the group prediction (GCB, unclassified, and ABC from left to right). ABC, activated B-cell–like; GCB, germinal center B-cell–like; IHC, immunohistochemical; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

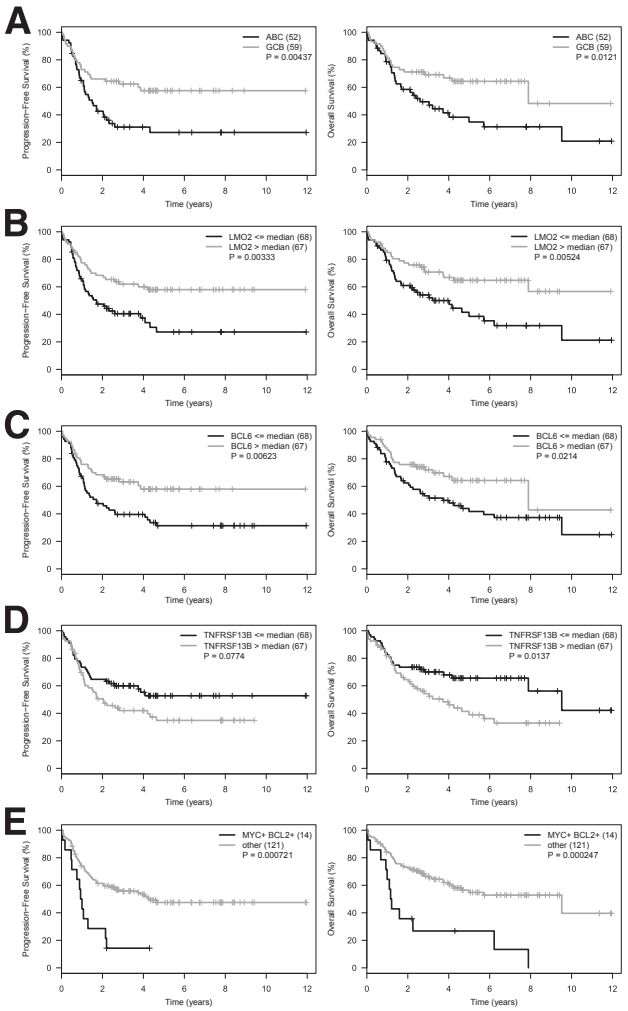


Figure 7 Patient outcomes according to RT-MLPA–based grouping. Progression-free and overall survival of a series of 135 patients treated with a combination of rituximab and chemotherapy are presented according to various grouping factors: GCB-ABC subgroup, as predicted by RT-MLPA (**A**), *LMO2* (**B**), *BCL6* (**C**), and *TNFRSF13B* (**D**) expressions as measured by RT-MLPA according to their respective medians, MYC^{+} $BCL2^{+}$ against the rest of the cohort (**E**), defined by an RT-MLPA expression higher than 1.284 for *MYC* and 0.256 for *BCL2*. Raw log-rank *P* values are displayed on the respective legend panels. ABC, activated B-cell–like; GCB, germinal center B-cell–like; RT-MLPA, reverse transcriptase multiplex ligation-dependent probe amplification.

molecular biology reagents and can thus be implemented in many laboratories.[17] Its simplicity and flexibility allow testing from a single to up to 40 samples in parallel, making it particularly suitable for a routine diagnosis workflow. Furthermore, with the few adaptations we made, the whole procedure can be achieved in <1 day, making it a rapid and inexpensive alternative (with a reagent cost of <$5/sample) to the current methods of GCB and ABC DLBCL stratification.

In the independent validation series we tested, none of the RT-MLPA calls were in contradiction with the current gold standard (Affymetrix arrays), and only a few samples moved from or to the unclassified group. Several well-known prognostic factors were also found to be significant when determined by RT-MLPA, including the GCB/ABC classification,[23] *LMO2* and *BCL6* gene expressions,[12] and *MYC-BCL2* coexpression,[22] further confirming the validity of this assay.

The proportions of unclassifiable or type III samples we observed (13.8% and 13.2% of the two frozen validation cohorts) are slightly higher than the 9% originally described with the gold standard[24] and the 7% reported by the same laboratory with the use of the NanoString technology.[16] It should be noted, however, that the confidence threshold used for decision making was raised from the 90% originally published to the more ubiquitous 95%, which can explain this small increase. Nevertheless, these proportions are consistent with the performances obtained with other gene expression-based methods: 16%[15] or 20%.[13] The same problem was also reported with Hans IHC algorithm[8–10] because CD10 and BCL6 reliable interpretation cannot always be achieved. Whether the type III subgroup corresponds to an independent biological entity or to a heteroclite collection of atypical tumors is still an open question. However, the discrepancies observed in our series about these lymphomas and the lack of consensus observed in the literature about their prognostic value[14,24] tend to support the later hypothesis.

Compared with other GEP methods, RT-MLPA is associated with a relatively poor dynamic range.[25] This characteristic affected the development of the classifier as some genes, such as *MME* (encoding CD10), one of the best markers of the GCB subtype, could not be incorporated into the final design. However, RT-MLPA also proved to be efficient in evaluating the expression of other highly discriminant markers, such as *LMO2*, which is particularly difficult to evaluate by IHC. Furthermore, even though GCB and ABC DLBCLs differ by the expression of thousands of genes,[4] the flexibility of RT-MLPA allowed us to easily and rapidly test multiple combinations of markers to reach a satisfactory design. Another important advantage of RT-MLPA over most other GEP methods is that only short cDNA fragments are necessary for the correct binding and ligation of the gene-specific oligonucleotide probes. It is thus less affected by the use of low RNA concentrations and RNA degradation and constitutes as such an interesting alternative for testing FFPE biopsies. It could thus be used for the retrospective analysis of archival collections and for the inclusion of patients in prospective clinical trials, because only a few institutions routinely collect frozen biopsy materials.

Conclusion

The robust and cost-effective RT-MLPA assay we describe, with its simple workflow and its short turnaround time, could be used in many laboratories without requiring the acquisition of any dedicated platform. It could then challenge IHC and other quantitative GEP assays to enable the stratification of DLBCL patients in prospective clinical trials

and ultimately to facilitate the generalization of GCB/ABC DLBCL-targeted therapies.

Acknowledgments

We thank Drs. Marie Hélène Delfau and Camille Laurent and Prof. Philippe Gaulard for critical reading of the manuscript and helpful comments.

S.M. and P.R. designed and performed research, analyzed and interpreted data, and wrote the manuscript; C.B. and V.M. performed experiments (RT-MLPA); M.C. performed research (IHC, FFPE samples); J.-P.J. analyzed and interpreted data (Affymetrix arrays); M.F. performed gene expression experiments (Affymetrix arrays); J.-M.P. and T.M. performed histopathologic review; T.F., G.S., and C.H. coordinated the LNH03 Lymphoma Study Association (LYSA) program and contributed vital reagents and samples; K.L. coordinated sample collection for the LYSA group; H.T. initiated the study and coordinated the LNH03 LYSA program, curated survival data, and supervised survival analyses; F.J. designed research, coordinated the GHEDI program, and wrote, edited, and approved the final manuscript.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2015.01.007*.

References

1. Swerdlow SH; International Agency for Research on Cancer, World Health Organization: WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon, France, International Agency for Research on Cancer, 2008
2. Coiffier B: Rituximab therapy in malignant lymphoma. Oncogene 2007, 26:3603–3613
3. Coiffier B, Thieblemont C, Van Den Neste E, Lepeu G, Plantier I, Castaigne S, Lefort S, Marit G, Macro M, Sebban C, Belhadj K, Bordessoule D, Fermé C, Tilly H: Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. Blood 2010, 116:2040–2045
4. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000, 403:503–511
5. Shaffer AL 3rd, Young RM, Staudt LM: Pathogenesis of human B cell lymphomas. Annu Rev Immunol 2012, 30:565–610
6. Barton S, Hawkes EA, Wotherspoon A, Cunningham D: Are we ready to stratify treatment for diffuse large B-cell lymphoma using molecular hallmarks? Oncologist 2012, 17:1562–1573
7. Roschewski M, Staudt LM, Wilson WH: Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nat Rev Clin Oncol 2014, 11:12–23
8. De Jong D, Rosenwald A, Chhanabhai M, Gaulard P, Klapper W, Lee A, Sander B, Thorns C, Campo E, Molina T, Norton A, Hagenbeek A, Horning S, Lister A, Raemaekers J, Gascoyne RD, Salles G, Weller E; Lunenburg Lymphoma Biomarker Consortium: Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications–a study from the Lunenburg Lymphoma Biomarker Consortium. J Clin Oncol 2007, 25:805–812
9. Salles G, de Jong D, Xie W, Rosenwald A, Chhanabhai M, Gaulard P, Klapper W, Calaminici M, Sander B, Thorns C, Campo E, Molina T, Lee A, Pfreundschuh M, Horning S, Lister A, Sehn LH, Raemaekers J, Hagenbeek A, Gascoyne RD, Weller E: Prognostic significance of immunohistochemical biomarkers in diffuse large B-cell lymphoma: a study from the Lunenburg Lymphoma Biomarker Consortium. Blood 2011, 117:7070–7078
10. Gutiérrez-García G, Cardesa-Salzmann T, Climent F, González-Barca E, Mercadal S, Mate JL, Sancho JM, Arenillas L, Serrano S, Escoda L, Martínez S, Valera A, Martínez A, Jares P, Pinyol M, García-Herrera A, Martínez-Trillos A, Giné E, Villamor N, Campo E, Colomo L, López-Guillermo A; Grup per l'Estudi dels Limformes de Catalunya I Balears (GELCAB): Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. Blood 2011, 117:4836–4843
11. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R: Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004, 350:1828–1837
12. Alizadeh AA, Gentles AJ, Alencar AJ, Liu CL, Kohrt HE, Houot R, Goldstein MJ, Zhao S, Natkunam Y, Advani RH, Gascoyne RD, Briones J, Tibshirani RJ, Myklebust JH, Plevritis SK, Lossos IS, Levy R: Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. Blood 2011, 118:1350–1358
13. Rimsza LM, Wright G, Schwartz M, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Cook JR, Tubbs RR, Braziel RM, Delabie J, Miller TP, Staudt LM: Accurate classification of diffuse large B-cell lymphoma into germinal center and activated B-cell subtypes using a nuclease protection assay on formalin-fixed, paraffin-embedded tissues. Clin Cancer Res 2011, 17:3727–3732
14. Barrans SL, Crouch S, Care MA, Worrillow L, Smith A, Patmore R, Westhead DR, Tooze R, Roman E, Jack AS: Whole genome expression profiling based on paraffin embedded tissue can be used to classify diffuse large B-cell lymphoma and predict clinical outcome. Br J Haematol 2012, 159:441–453
15. Masqué-Soler N, Szczepanowski M, Kohler CW, Spang R, Klapper W: Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens. Blood 2013, 122: 1985–1986
16. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM, Smeland EB, Mottok A, Braziel RM, Ott G, Delabie J, Tubbs RR, Cook JR, Weisenburger DD, Greiner TC, Glinsmann-Gibson BJ, Fu K, Staudt LM, Gascoyne RD, Rimsza LM: Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 2014, 123:1214–1217
17. Eldering E, Spek CA, Aberson HL, Grummels A, Derks IA, de Vos AF, McElgunn CJ, Schouten JP: Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. Nucleic Acids Res 2003, 31:e153
18. Bohers E, Mareschal S, Bouzelfen A, Marchand V, Ruminy P, Maingonnat C, Ménard AL, Etancelin P, Bertrand P, Dubois S, Alcantara M, Bastard C, Tilly H, Jardin F: Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell lymphoma. Genes Chromosomes Cancer 2014, 53:144–153

19. Jardin F, Mareschal S, Figeac M, Jais JP, Leroy K, Copie-Bergman C, Salles GA, Coiffier B, Delarue R, Peyrade F, Bosly A, Ketterer N, Haioun C, Tilly H, Molina TJ: Integrated analysis of high-resolution gene expression and copy number profiling identified biallelic deletion of CDKN2A/2B tumor suppressor locus as the most frequent and unique genomic abnormality in diffuse large B-cell lymphoma (DLBCL) with strong prognostic value in both GCB and ABC subtypes and not overcome by a dose-intensive immunochemotherapy regimen plus rituximab. Results of a Prospective GELA Clinical Trial Program. ASH Annual Meeting Abstracts 2012, 120:415
20. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, Müller-Hermelink HK, Campo E, Braziel RM, Jaffe ES, Pan Z, Farinha P, Smith LM, Falini B, Banham AH, Rosenwald A, Staudt LM, Connors JM, Armitage JO, Chan WC: Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004, 103:275–282
21. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B 1995, 57:289–300
22. Johnson NA, Slack GW, Savage KJ, Connors JM, Ben-Neriah S, Rogic S, Scott DW, Tan KL, Steidl C, Sehn LH, Chan WC, Iqbal J, Meyer PN, Lenz G, Wright G, Rimsza LM, Valentino C, Brunhoeber P, Grogan TM, Braziel RM, Cook JR, Tubbs RR, Weisenburger DD, Campo E, Rosenwald A, Ott G, Delabie J, Holcroft C, Jaffe ES, Staudt LM, Gascoyne RD: Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol 2012, 30:3452–3459
23. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al; Lymphoma/Leukemia Molecular Profiling Project: Stromal gene signatures in large-B-cell lymphomas. N Engl J Med 2008, 359:2313–2323
24. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM: A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A 2003, 100:9991–9996
25. Patrinos GP: Molecular diagnostics. Amsterdam, Academic Press, 2010

3 Caractères génomiques

3.1 État de l'art

3.1.1 Altérations du nombre de copies (CNA)

Mécanistique

Au même titre que la plupart des cancers, les LDGCB sont la cible de nombreux réarrangements somatiques produisant des génomes tumoraux très différents de ceux portés par les cellules saines environnantes. Cela peut se traduire par la translocation de portions chromosomiques (§3.1.2), mais également par la **perte** ou la **duplication** de régions allant du simple gène au chromosome entier, ou par des **disomies uniparentales acquises** (CNLOH). En modifiant le nombre de copies d'une région chromosomique, ces gains et pertes influent sur le niveau d'expression des gènes impliqués. Les oncogènes se trouvent préférentiellement ciblés par des amplifications pouvant atteindre plusieurs dizaines de copies, et les gènes suppresseurs de tumeurs par des délétions touchant un allèle (délétion hétérozygote), voire les deux (délétion homozygote). Ces altérations peuvent également agir comme événement secondaire à une mutation ponctuelle, augmentant ainsi le nombre d'allèles mutés ou diminuant le nombre d'allèles sauvages afin d'exprimer davantage le transcrit altéré. Les CNLOH ne s'expriment qu'à travers ce dernier mécanisme, dans la mesure où elles consistent en une perte localisée de l'hétérozygotie sans altération du nombre de copies, ce qui peut expliquer que leur étude reste encore marginale.

Les mécanismes à l'origine de ces altérations somatiques ne sont pas encore totalement élucidés, et doivent être distingués de ceux survenant lors de la fécondation ou de la gamétogenèse qui conduisent aux altérations constitutionnelles. Plusieurs mécanismes ont ainsi été décrits pour expliquer la survenue d'anomalies chromosomiques en dehors de la méiose, et ce à des échelles distinctes.

Les **anomalies de la mitose** constituent les plus simples de ces mécanismes, impliquant de larges régions de l'ordre du bras chromosomique. La non-disjonction des

chromatides lors de l'anaphase peut ainsi conduire au gain d'un chromosome pour l'une des cellules filles, et par symétrie à la perte dudit chromosome pour la seconde cellule (Figure 3.1). De la même façon, une disjonction incorrecte peut conduire à la formation d'isochromosomes monocentriques [142], caractérisés par la présence de deux bras longs ou courts au lieu d'une combinaison des deux (Figure 3.1).

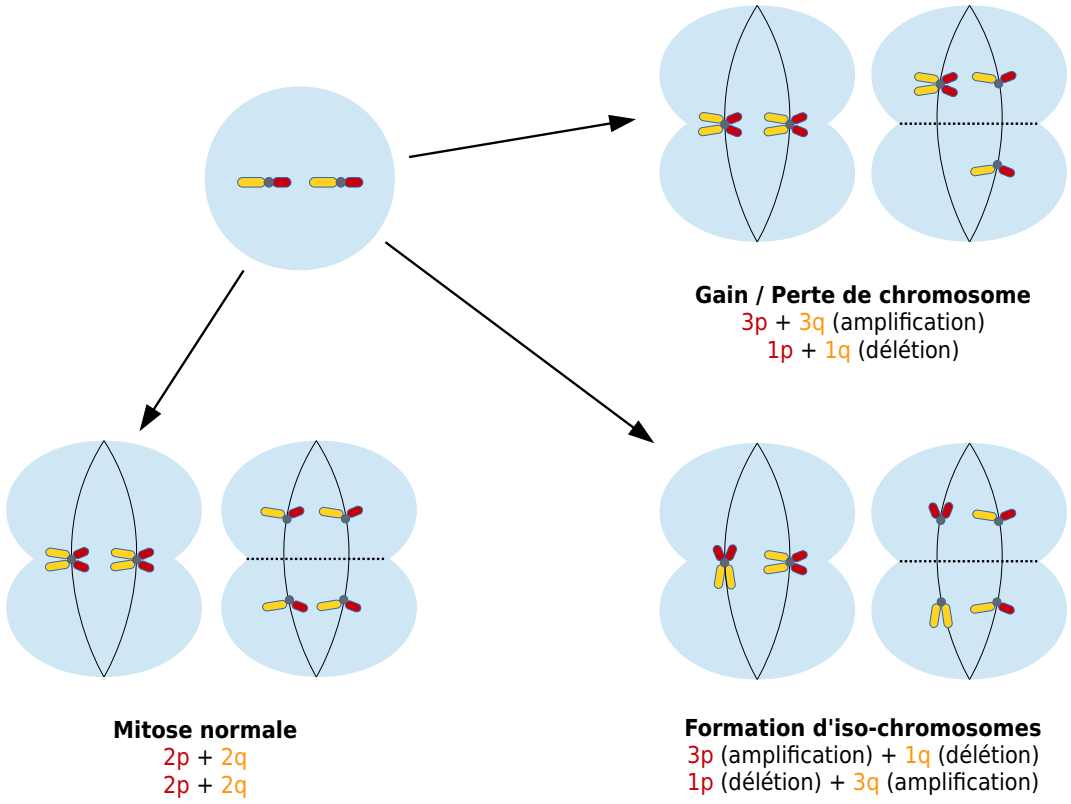


Figure 3.1 – Anomalies de mitoses et altérations chromosomiques

Illustration d'anaphases de mitoses normales et anormales pour une paire de chromosomes homologues (organisme diploïde). Les bras courts (p) sont présentés en rouge et les bras longs (q) en jaune, les centromères en gris foncé et le fuseau mitotique par des traits noirs.

Les réarrangements chromosomiques plus localisés impliquent généralement une ou plusieurs cassures double-brins de la molécule d'ADN, suivies d'une réparation incorrecte par l'un des nombreux mécanismes existants [143]. Les **cassures double-brins** sont connues pour survenir de façon aléatoire au sein de l'ensemble des cellules vivantes, causées notamment par les dérivés réactifs de l'oxygène (ROS) produits lors de l'exposition aux radiations ionisantes (naturelles, thérapeutiques ou accidentelles) [143, 144]. De nombreuses chimiothérapies reposant sur ce phénomène, conduisant à la mort cellulaire préférentiellement dans les cellules en division rapide qui constituent les tumeurs, il n'est pas rare d'observer des leucémies « chimio-induites » plusieurs années après un cancer, du fait de l'apparition d'altérations du génome dans des cellules initialement saines exposées au traitement [145].

Parmi les nombreux mécanismes mis en place par la cellule pour réduire l'impact de ces cassures, les plus conservateurs sont ceux de **réparation par recombinaison homologue**. Comme leur nom l'indique, ces mécanismes reposent sur l'utilisation d'une région homologue (chromatide sœur durant la mitose, chromosome homologue voire une homologie locale avec d'autres chromosomes) comme matrice pour rétablir la liaison entre les deux fragments. Le modèle DSBR [146] est le plus étudié de ces mécanismes, notamment pour sa faculté à produire des phénomènes d'enjambement, ou *crossing-over* (Figure 3.2). Selon ce modèle, la réparation de la molécule double-brin cassée est permise par un appariement de chaque brin au brin complémentaire de la séquence donneuse, formant après réplication et ligation un tétramère d'ADN basé sur deux jonctions de Holliday. Les contraintes physiques appliquées à cette conformation provoqueront sa séparation aléatoire en deux molécules d'ADN double-brins, pouvant résulter en un échange entre les deux partenaires (Figure 3.2).

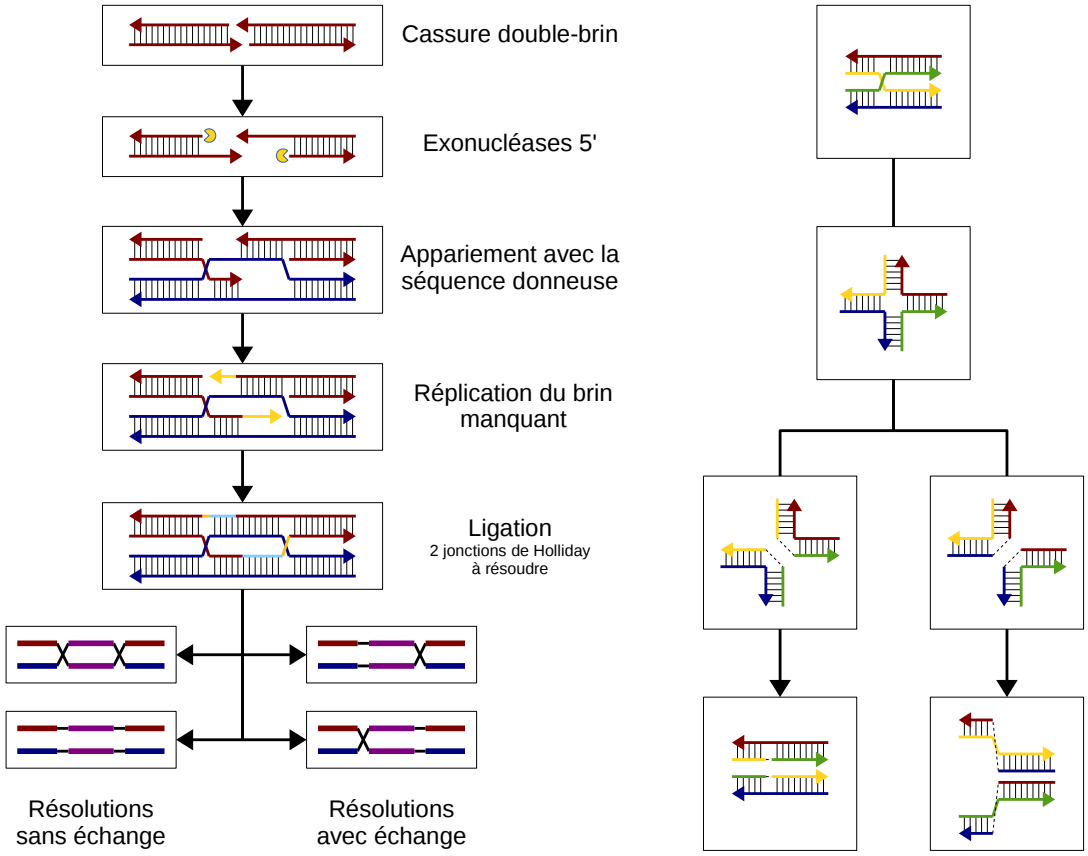


Figure 3.2 – Modèle *Double-Strand Break Repair*

La **partie gauche** présente la réparation d'une cassure double-brin sur l'ADN double-brin rouge grâce à son chromosome homologue bleu. La **partie droite** décrit les résolutions possibles d'une des deux jonctions de Holliday survenues lors du mécanisme DSBR.

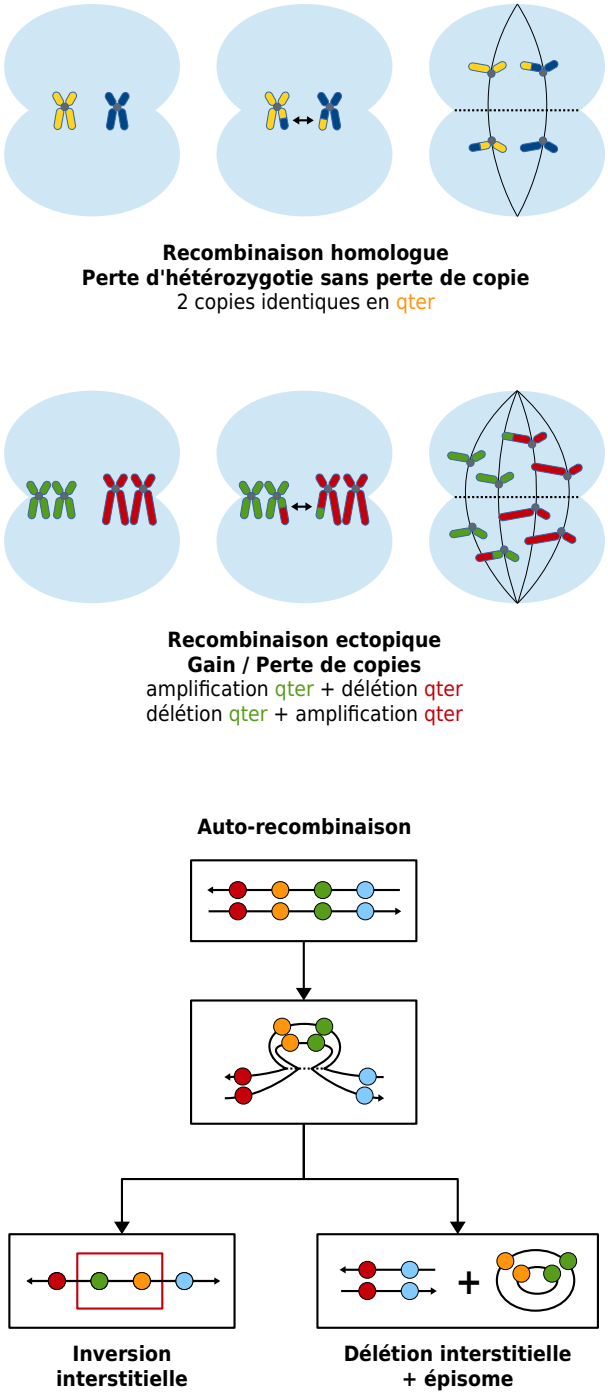


Figure 3.3 – Recombinaisons et altérations chromosomiques

Illustration des conséquences d'un enjambement chromosomique impliquant trois types de partenaires différents. Le **panneau du haut** dépeint la recombinaison survenue entre deux chromosomes homologues lors de la mitose (bleu et jaune selon l'origine paternelle ou maternelle). Le **panneau central** décrit une recombinaison survenue entre deux paires de chromosomes différents lors de la mitose (rouge et vert selon le numéro du chromosome impliqué, les chromosomes homologues étant cette fois-ci de couleurs identiques). Le **panneau du bas** explique la recombinaison survenue au sein d'un unique chromosome, les boules de couleur servant à représenter l'ordre des séquences le long de la molécule d'ADN double-brin.

Si l'échange en lui-même n'implique pas d'altération du nombre de copies, la perte ou la duplication subséquente d'un chromosome dérivé d'une telle translocation selon les mécanismes décrits précédemment (Figure 3.1) pourront en revanche conduire à des CNA plus localisés [147]. Selon le partenaire impliqué (Figure 3.3), il pourra en résulter des pertes d'hétérozygotie sans perte de copie (CNLOH), des gains et pertes de copies, des inversions ou encore des épisomes. Bien que la recombinaison ectopique entre chromosomes différents soit facilitée par la présence au sein du génome de nombreux éléments répétés (LINE, SINE, rétrotransposons...), cela reste un événement rare dans la mesure où l'utilisation d'une chromatide sœur est bien plus aisée et fréquente, et qu'une minorité des recombinaisons donnent lieu à un échange de matériel génétique [147].

Par opposition à ces mécanismes de réparation basés sur la recombinaison homologue, principalement actifs durant les phases S et G2 de la mitose où les chromatides sont doublées, d'autres mécanismes non spécifiques peuvent être sollicités. Les mécanismes de **jonction d'extrémités non-homologues** (NHEJ) comptent ainsi parmi les principales causes de translocation chez l'Homme [143, 147], et donc d'anomalies du nombre de copies. Bien qu'ils soient facilités par une courte homologie au niveau du point de cassure (dizaine de bases), ces mécanismes permettent la jonction de deux fragments d'ADN qu'il y ait homologie ou non. Ils sont de plus tolérants aux cassures non-franches formant des extrémités incompatibles, grâce à une activité exonucléase préalable à la fusion.

Vue d'ensemble des LDGCB

L'incidence relativement importante des LDGCB et leur tendance à la clonalité ont permis d'établir un profil assez clair des altérations chromosomiques présentes dans cette pathologie, et ce avec une précision croissante à mesure que les techniques de cytogénétique ont évolué. La cytométrie en flux a ainsi démontré un niveau de réarrangement modéré dans les LDGCB, avec seulement 34% des cas présentant une aneuploïdie marquée [148, 149]. La majorité de ces aneuploïdies consistent en une variation du nombre de chromosomes par rapport à l'état diploïde (21%), laissant quelques cas pseudo-tétraploïdes (8%) voire plus complexes encore (5%). L'application des techniques de caryotypage en bandes G ou Q à de larges cohortes de LDGCB a permis d'affiner ces résultats [150], mettant en évidence des caryotypes anormaux à cette résolution dans 87% des cas. Ceux-ci consistent en un gain d'un ou plusieurs chromosomes dans 72% des cas (hyperdiploïdie), en des pertes dans 9% des cas (hypodiploïdie). Parmi les anomalies les plus fréquentes dans les LDGCB, ces techniques ont pu mettre en évidence des gains des chromosomes 7 (17%), 12 (17%), X (13%), 11 (11%), 3 (10%) et 18 (10%), des pertes des chromosomes 6 (11%), Y (10%), 4

(8%) et X (8%) et de nombreuses translocations (§3.1.2). L'écart important observé entre ces deux approches peut s'expliquer par le fait que les réarrangements dans les LDGCB concernent pour la majorité des cas quelques chromosomes ou parties de chromosomes. L'analyse des 1 427 caryotypes de LDGCB publiés dans CGAP [151] montre en effet que le nombre de chromosomes dans cette pathologie se distribue autour d'une médiane à 47 chromosomes, avec des déciles à 10% et 80% de 45 et 51 chromosomes respectivement (Figure 3.4).

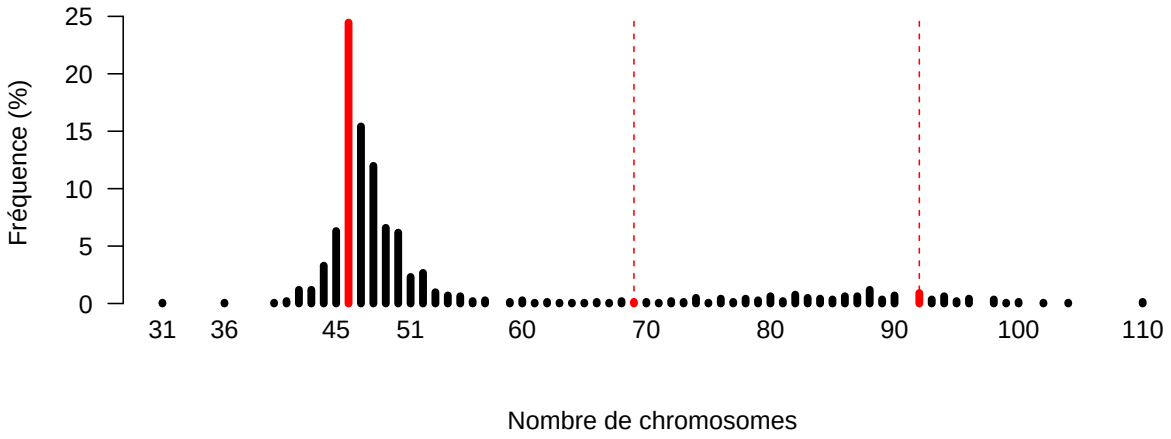


Figure 3.4 – Distribution du nombre de chromosomes dans les LDGCB

Distribution du nombre de chromosomes dans 1 427 caryotypes de LDGCB [151]. Les barres rouges correspondent aux normes diploïdes (46), triploïdes (69) et tétraploïdes (92).

Si ces techniques ont permis de mettre en évidence quelques *loci* plus réduits (délétions 1q36, 7q22-q36, 6q21, 6q23 [150]), l'identification de gènes apportant un avantage sélectif à ces anomalies n'a pris son essor qu'avec le développement de la CGH, sur métaphases puis sur puces (§3.3.1). La précision que cette technique apporte à la localisation des points de cassure a mis en évidence une grande diversité entre les patients, faisant de l'identification de tels gènes *drivers* un domaine encore étudié aujourd'hui (§3.5). La plupart des études s'intéressent aux gènes présents dans la région communément altérée chez un maximum de patients (MCR), ce qui peut conduire à des résultats différents selon les cohortes, notamment dans les régions plus floues abritant de nombreux gènes d'intérêt tels que le bras 6q. Le panorama présenté en Figures 3.5 et 3.6 est ainsi basé sur 4 grandes études utilisant des puces CGH ou SNP à haute résolution, afin d'être aussi précis que possible :

— 37 LPM publiés par Wessendorf *et al* en 2007 [152], après une analyse sur puces CGH comprenant 2800 BAC, faute de série de LPM plus récente.
— L'analyse de 203 LDGCB *de novo* sur puces CGH NimbleGen 380 000 sondes par Lenz *et al* en 2008 [22], dans les sous-types GCB, ABC et LPM définis par puces Affymetrix.

— La série de 166 LDGCB au diagnostic de Scandurra *et al* 2010 [153], analysée sur puces CGH GeneChip 250 000 sondes.
— L'étude de 180 LDGCB non-traités analysés sur puces Affymetrix SNP 6.0 (900 000 sondes SNP et 900 000 sondes CNA) par Monti *et al* en 2012 [154], classés en GCB et ABC par puces Affymetrix.

La délétion de *CDKN2A/B* en 9p21 (30%) et du bras 6q (40%), incluant de nombreux gènes suppresseurs de tumeurs tels que *PRDM1* ou *TNFAIP3*, sont ainsi les anomalies les plus emblématiques du sous-type **ABC** (Figures 3.5 et 3.6). En terme d'amplification, les régions 18q (abritant *BCL2*) et 19q13 (*SPIB*) sont également des cibles fréquentes, dans environ 40% et 30% des ABC. D'autres gains dont les gènes *drivers* sont moins bien identifiés viennent compléter ce tableau, notamment la trisomie 3 (25%) et les gains en 5p15 et 11q24 (20% environ). Les sous-types **GCB** et **LPM** démontrent là encore une forte proximité génomique, avec des délétions 1p36 dans 40% et 25% des cas respectivement, ainsi que des amplifications emblématiques de *REL* et *BCL11A* en 2p15-p16 (35% des GCB et 25% à 50% des LPM). Les gains du bras 7q, et plus particulièrement des *loci* 7q22 et 7q36 constituent également des anomalies fréquentes (30% des cas environ), de même que les délétions en 13q34 (30% et 20%). Les gains de la partie centromérique du bras 12q concernent également 20% et 30% des GCB et LPM, mettant en avant des gènes tels que *MDM2* ou *BEST3*. Enfin de manière plus spécifique, les **LPM** se caractérisent par des amplifications du chromosome 9, et notamment de la région 9p22-p24 abritant *JAK2* et *PDCD1LG2*, pouvant toucher jusqu'à 70% de ce sous-type.

Altérations du nombre de copies courantes

La délétion très localisée des gènes ***CDKN2A*** et ***CDKN2B*** en 9p21 constitue un des CNA majeurs dans les LDGCB, retrouvé chez 35% des ABC et 16% des GCB [154]. *CDKN2A* code par épissage alternatif les protéines p14 (ARF) et p16 (INK4A), tandis que *CDKN2B* code la protéine p15 (INK4B). Les protéines p15 et p16 sont connues pour inhiber les *Cyclin-Dependent Kinases* 4 et 6 en cas de perte d'intégrité du génome, blocant ainsi la transition du cycle cellulaire de la phase de quiescence G1 à la phase de réplication S [155]. p14 serait quant à elle impliquée dans l'activation de p53 via MDM2 en réponse à des signaux hyper-prolifératifs tels que l'expression de *MYC*, conduisant là encore à des effets suppresseurs de tumeurs variés (contrôle du cycle cellulaire, apoptose...) [156]. La délétion conjointe de ces inhibiteurs fournit ainsi à la cellule tumorale un moyen d'échapper à deux mécanismes de régulation majeurs et peu spécifiques des lymphomes, ce qui explique que cet événement soit retrouvé dans de nombreux cancers non-hématologiques [154] et associé à un moins bon pronostic, y compris au sein du groupe ABC [22, 157]. L'inactivation de ce gène

Altération					Région	Gènes / Positions d'intérêt	Sous-type			Autres cancers	Source
−−	−	LOH	+	++			ABC	GCB	LPM		
-	✔	-		-	1p13	*CD58*	9	4	-	non	Monti
-	✔	-		-	1p36.11	*SLC9A1 ?*	11	5	-	non	Monti
✔	✔	-			1p36.32	*TP73*	4 / 16	10 / 42	26	oui	Lenz
-	✔	-		-			7	18	-		Monti
	✔						26		-		Scandurra
		-	✔		1q	-			11		Wessendorf
-		-	✔	-	1q21-q23	*FCRLA, FCGR2B, FCGR2C*	11	20	-	oui	Monti
		-	✔		1q31.3	*DENND1B ?*	7	21	0		Lenz
-	✔	-		-	1q32.1	*RASSF5, IKBKE, IL10, PIGR*	9	0	-	non	Monti
		-	✔	✔	2p15-p16	*REL, BCL11A, COMMD1*	12 / 5	35 / 28	26 / 20	non	Lenz
-		-	✔	-			4	23	-		Monti
			✔	✔			20 / 4		-		Scandurra
		-	✔	✔			-	-	51 / 11		Wessendorf
		-	✔	✔	2q35	*CXCR2*	-	-	11 / 3		Wessendorf
		-	✔		3	...	26	1	0		Lenz
		-	✔		3p	...	30	1	3		Lenz
			✔				6		-		Scandurra
	✔	-			3p	...	-	-	13		Wessendorf
		-	✔		3p14	*FOXP1*	12	3	0		Lenz
✔	✔				3p14	*FHIT*	1 / 8		-		Scandurra
		-	✔		3q	...	39	3	7		Lenz
			✔				11		-		Scandurra
-		-	✔	-		3q29	24	13	-		Monti
		-	✔	✔	3q11-q12	*NFKBIZ*	12 / 10	0 / 0	0 / 0		Lenz
		-	✔		3q27	*BCL6, miR-28*	-	-	13		Wessendorf
	✔				4p11-p12	*FRYL ?*	14		-		Scandurra
-	✔	-		-	4q35	*ZFP42*	2	10	-	oui	Monti
	✔						10		-		Scandurra
			✔		5p	...	10		-		Scandurra
-		-	✔	-	5p15	*TERT*	22	12	-	oui	Monti
-		-	✔	-	5q35	*CANX*	13	13	-	non	Monti
			✔		6p	...	7		-		Scandurra
	✔	-			6p21	MHC class II	-	-	11		Wessendorf
		✔			6p21-p25	...	15		-		Scandurra
	✔	-			6q	...	39	17	10		Lenz
	✔	✔					30 / 6		-		Scandurra
	✔	-			6q12	Ø	9	0	19		Lenz
	✔	✔			6q16	*PRDM13 ?*	28 / 16		-		Scandurra
-	✔	-		-	6q21	*PRDM1*	29	21	-	oui	Monti
	✔						31		-		Scandurra
-	✔	-		-	6q23	*TNFAIP3*	29	22	-	oui	Monti
✔	✔	✔					1 / 32 / 14		-		Scandurra
			✔		6q27	*DLL1 ? TBP ?*	7		-		Scandurra
			✔		7	...	23		-		Scandurra
			✔		7p	...	19		-		Scandurra
			✔		7q	...	19		-		Scandurra
-		-	✔	-	7q22	*CUX1*	15	29	-		Monti
		-	✔				-	-	32		Wessendorf
-	✔	-		-	7q34	*PRSS1/2*	11	9	-	non	Monti
	✔						15		-		Scandurra
-		-	✔	-	7q36	*PTPRN2 ?*	18	27	-		Monti
	OS	✔			8p21-p23.1	8p23.1 : *NEIL2 ? GATA4 ?*	13 / 7		-		Scandurra
	✔	-				8p21.3 : *TNFRSF10*	-	-	11	oui	Wessendorf
-	✔	-		-			13	7	-		Monti
			✔		8p22-p23.1	Ø	13		-		Scandurra
-		-	✔	-	8p23.3	*ERICH1*	7	12	-	non	Monti
-		-	✔	-	8q24	8q24.3 : *PARP10 ? ...*	9	14	-	oui	Monti
		-	✔			8q24.21 : *MYC*	-	-	13		Wessendorf
		-	✔		9	...	7	0	26		Lenz
		-	✔	✔	9p	...	11 / 3	7 / 1	45 / 26		Lenz
		✔	✔				7 / 10		-		Scandurra
✔	✔	-			9p21	*CDKN2A/B, MTAP*	20 / 30	0 / 4	0 / 7		Lenz
-	✔	-		-		*CDKN2A/B*	35	16	-	oui	Monti
✔	✔	✔				*CDKN2A/B*	3 / 19 / 10		-		Scandurra
		-	✔		9p22-p24	*JAK2, PDCD1LG2*	2	3	10		Lenz
		✔	✔	**DFS**			10 / 13 / 2		-		Scandurra
		-	✔	✔			-	-	68 / 11		Wessendorf
-	✔	-		-	9p22-p24	*JAK2*	6	17	-	oui	Monti
		-	✔		9q	...	7	3	32		Lenz
		✔	✔				5 / 5		-		Scandurra
		✔			9q31	*TMEFF1 ?*	13		-		Scandurra
-		-	✔	-	9q34	*TRAF2 ?*	13	9	-	non	Monti
		-	✔				-	-	32		Wessendorf
	✔	-			10	...	1	1	16		Lenz
	✔	-			10q23-q24	*PTEN, FAS*	0	11	0		Lenz
OS	✔	✔				*FAS*	1 / 10 / 7		-	oui	Scandurra
-	✔	-		-		*FAS*	7	8	-		Monti

Figure 3.5 – Bilan des CNA répertoriés dans les LDGCB (partie 1)

Suite et légende en Figure 3.6

Altération −−	Altération −	Altération LOH	Altération +	Altération ++	Région	Gènes / Positions d'intérêt	Sous-type ABC	Sous-type GCB	Sous-type LPM	Autres cancers	Source
			✔		11p	...		11	-		Scandurra
			✔		11q	...		14	-		Scandurra
	✔	-			11q13	*FADD, miR-139*	-	-	11		Wessendorf
		-	✔		11q23	...	-	-	18		Wessendorf
			✔		11q24-q25	*ETS1 ? FLI1 ? TP53AIP1 ?*		30	-	non	Scandurra
-		-	✔	-			22	10	-		Monti
		-	✔		12	...		9	-		Scandurra
		-	✔		12p	...		10	-		Scandurra
		-	✔			...	-	-	16		Wessendorf
		-	✔		12q	...		10	-		Scandurra
		-	✔		12q14-q21	12q14 : *MDM2*	0	13	3		Lenz
		-	✔			12q14 : *MDM2*	-	-	30		Wessendorf
-		-	✔	-		12q15 : *BEST3 ?*	6	22	-		Monti
		✔	✔			12q21 : Ø		7 / 19	-		Scandurra
			✔		13q	...		10	-		Scandurra
-	✔	-		-	13q14	*MED4, RB1*	7	12	-	oui	Monti
	✔	✔				*DLEU1/2*		13 / 7	-		Scandurra
	✔	-				*miR-15A, miR-16*	-	-	13		Wessendorf
		-		✔	13q31-q32	*MIR17HG*	0	13	3		Lenz
-		-	✔	-			4	10	-	oui	Monti
		✔	✔	✔				7 / 17 / 1	-		Scandurra
	✔	-			13q34	*ING1*	10	31	19		Lenz
	✔	✔						13 / 6	-		Scandurra
	✔	-			15q11	*miR-211*	-	-	16		Wessendorf
-	✔	-		-	15q15-q21	*B2M*	4	12	-	non	Monti
	✔	✔						19 / 9	-		Scandurra
		-	✔		16p13	*SOCS1, CIITA*	12	0	7		Lenz
	✔	✔			17p	...		14 / 6	-		Scandurra
	✔	✔			17p13	*TP53*		17 / 11	-		Scandurra
-	✔	-		-			13	10	-	oui	Monti
	✔	-					-	-	13		Wessendorf
			✔		17q21	Ø		25	-		Scandurra
-		-	✔	-	17q25	*RAB40B*	11	8	-	oui	Monti
			✔		18	...		11	-		Scandurra
	✔		✔		18p	...		9 / 11	-		Scandurra
-	✔	-		-	18p11	Ø	4	5	-	oui	Monti
	✔					*YES1 ? THOC1 ? TYMS1 ?*		18	-		Scandurra
			✔		18q	...		20	-		Scandurra
		-	✔				38	15	16		Lenz
		-		✔	18q21	*BCL2*	8	5	0		Lenz
-		-	✔	-			31	12	-	oui	Monti
		-	✔				-	-	24		Wessendorf
-		-	✔	-	19p13	*MUM1, TCF3, ...*	15	5	-	non	Monti
-	✔	-		-	19p13	*CD70*	4	9	-	oui	Monti
✔	✔					*CD70, C3, TNFSF9, TNFSF14*		1 / 10	-		Scandurra
		-	✔		19q	...	19	4	0		Lenz
		-	✔		19q13	*SPIB, ...*	26	3	3		Lenz
-		-	✔	-		*IL11, FIZ1, ...*	28	4	-	non	Monti
			✔	✔		*IL11, FIZ1, ... / SPIB*		8 / 1	-		Scandurra
		-	✔		20p	...	3	6	23		Lenz
			✔		21	...		5	-		Scandurra
-		-	✔	-	21q22	...	4	16	-	non	Monti
-	✔	-		-	22q11	*PRAME, BCR*	7	8	-	non	Monti

Figure 3.6 – Bilan des CNA répertoriés dans les LDGCB (partie 2)

Les altérations du nombre de copies touchant 10% ou plus des LDGCB sont présentées par ordre génomique, après croisement de différentes sources. Les anomalies représentatives touchant 20% d'un sous-type ou plus sont signalées par un **carré noir**. Les colonnes **altérations** décrivent le type de chaque altération (de gauche à droite : délétions homozygote et hétérozygote, perte d'hétérozygotie sans perte de copie, gain d'une ou plusieurs copies), les types d'anomalies non rapportés étant marqués d'un tiret et les anomalies prédictives de la survie par le type de survie prédite. Les **gènes** rapportés pour chaque région ont soit été mis en évidence par les auteurs, soit retrouvés après un nouvel examen des coordonnées publiées, les hypothèses les plus incertaines étant signalées par un « ? ». L'occurrence dans chaque **sous-type** est rapportée en proportion de la série étudiée, selon disponibilité du sous-type dans l'étude source. La présence de l'anomalie dans des **cancers non-hématologiques** est rapportée telle que décrite par Monti *et al* [154]. La **source** est rapportée parmi les études suivantes : *Monti* (180 LDGCB) [154], *Lenz* (203 LDGCB) [22], *Scandurra* (166 LDGCB) [153] et *Wessendorf* (37 LPM) [152].

pourrait enfin concerner une part plus importante des LDGCB, l'hyper-méthylation de son promoteur ayant été retrouvée de manière fréquente et corrélée à un niveau d'expression plus bas [158], aboutissant similairement à un moins bon pronostic [128].

Le *locus* 2p15-p16, dupliqué voire amplifié dans 35% des GCB [22] et jusqu'à 50% des LPM [152], regroupe également plusieurs gènes d'intérêt. En temps que sous-unité du facteur de transcription NFκB, voie métabolique d'importance majeure dans les LDGCB menant à l'activation de nombreux facteurs anti-apoptotiques tels que *BCL2* [42], le gène ***REL*** (c-Rel) constitue un candidat de choix dans ce *locus*. Le gène ***BCL11A***, distant de quelques centaines de kilo-bases et généralement co-amplifié, revient fréquemment dans la littérature comme second *driver* potentiel de cette amplification. Les liens établis entre ce gène et des phénomènes clés de la lymphopoïèse tels que la recombinaison VDJ [159], l'hyper-mutation somatique [160, 161] ou la protéine Bcl-6 [162] en font en effet un candidat valable, bien que des études plus fines du *locus* 2p15-p16 tendent à démontrer une préférence pour l'amplification de *REL* [163]. ***COMMD1***, bien que situé environ 1 méga-base en aval de *REL* légèrement hors du MCR habituel, est apparu plus récemment comme un troisième candidat lié au pronostique [164]. Le peu d'études s'intéressant à sa fonction dans un contexte onco-hématologique suggérant plutôt un rôle d'inhibiteur de NFκB [165], on peut cependant penser qu'il s'agirait d'un gène passager de cette amplification. Pour finir, le biais de distribution de ces amplifications en faveur du groupe GCB pourrait s'expliquer par la présence de cette anomalie dans les LF, et par conséquent dans les LDGCB issus de leur transformation [166].

La duplication voire l'amplification du *locus* 9p22-p24 constitue un événement emblématique des LPM (jusqu'à 70% des cas [152]), pouvant se matérialiser par des gains plus ou moins focalisés du bras 9p ou du chromosome 9 dans son intégralité. La tyrosine kinase ***JAK2*** est la cible la plus communément admise dans ce *locus*, de par son rôle clé dans la voie JAK-STAT et les réponses aux cytokines et facteurs de croissance qui en découlent [167]. Les mutations inactivatrices de *SOCS1*, inhibiteur bien connu de cette voie, observées de manière quasi-systématique dans les LPM et les LH génétiquement proches (§4.1.3) tendraient à conforter cette hypothèse. Des études *in vitro* ont également pu démontrer l'effet marqué de l'inhibition de *JAK2* sur la croissance de ces deux types de tumeurs [168], ouvrant la voie à des essais cliniques de phase I et II [169, 170]. On peut tout de même s'étonner que la mutation activatrice de *JAK2* V617F, observée fréquemment dans les syndromes myéloprolifératifs, n'ait pas été retrouvée dans ces lymphomes [171], ce qui suggérerait d'autres avantages à cette amplification. La présence dans ce *locus* de ***KDM4C*** a ainsi été avancée, soulignant un mécanisme de dérégulation épigénétique qui peut également être imputé à *JAK2* [42]. *KDM4C* serait en effet capable de déméthyler la lysine 9 des histones 3 (H3K9), et *JAK2* de phosphoryler la tyrosine 41 de ces mêmes

histones, ces deux mécanismes pouvant mener au relâchement de la chromatine. ***PDCD1LG2*** constitue enfin un troisième candidat de poids, l'amplification 9p22-p24 résultant en une sur-expression marquée de ce ligand de PD-1 capable de réprimer l'activation des lymphocytes T environnants, et donc de potentiellement échapper à ce type de réponse immunitaire [42, 172].

La délétion du bras 6q, relativement ubiquitaire au sein des cancers [154], est également retrouvée fréquemment dans les LDGCB, et ce préférentiellement dans le sous-type ABC de mauvais pronostique. Contrairement aux *loci* décrits précédemment, la délétion 6q présente un aspect discontinu [173], vraisemblablement dû à la présence de nombreux gènes d'intérêts, qui rend son analyse délicate. Lenz *et al* considèrent ainsi qu'environ 40% des LDGCB ABC et 20% des deux autres sous-types seraient touchés, en appliquant comme critère qu'au moins 60% du bras soit délété [22]. Parmi les MCR les plus communément admis, la position 6q23 hébergeant ***TNFAIP3*** (A20) semble particulièrement intéressante. En effet, la délétion de cet inhibiteur de la voie NFκB préférentiellement dans le sous-type ABC est en adéquation avec le lien établi entre ce sous-type et l'activation constitutive de la voie NFκB [42]. Le *locus* 6q21 abritant ***PRDM1*** (Blimp-1) a également un rôle majeur dans la délétion 6q chez les ABC, ce facteur de transcription étant impliqué dans la différenciation et l'arrêt du cycle cellulaire. L'amplification 19q13 impliquant ***SPIB*** dans 25% des ABC peut d'ailleurs être perçue comme un moyen détourné d'arriver au même résultat, *SPIB* et *PRDM1* réprimant leurs expressions respectives [42]. En amont, les translocations activatrices de *BCL6* observées chez les ABC (§3.1.2) peuvent également être reliées à cette voie, *BCL6* réprimant l'expression de *PRDM1* [42].

L'amplification 18q impliquant ***BCL2*** peut être considérée comme une alternative à la translocation t(14;18) développée dans les sous-types ABC et LPM (30% [154] et 25% [152] respectivement), cette dernière translocation étant plutôt l'apanage des GCB (§3.1.2). Ces deux mécanismes semblent conduire à la sur-expression de *BCL2*, au moins au niveau protéique [174], permettant à la cellule d'échapper aux mécanismes d'apoptose.

Enfin les délétions 13q34 et 1p36, retrouvées majoritairement dans les GCB (30% et 40% respectivement), peuvent être toutes deux reliées à la voie p53. En 13q34, ***ING1*** apparait comme un gène dont la suppression conduit à une sensibilité plus élevée aux dommages génomiques et à la survenue de lymphomes chez la souris [175], un effet indépendant quoique potentialisé par la perte conjointe de p53 [176]. Le gène ***TP73*** en 1p36 présentant une homologie forte avec le bien connu *TP53* et le potentiel d'interagir avec les mêmes cibles, des effets sur la résistance aux altérations du génome, l'apoptose ou la prolifération peuvent être attendus après sa délétion [177]. La délétion de ***TP53*** lui-même en 17p13 peut également être ob-

servée dans 10% à 15% des LDGCB, indépendamment du sous-type et même dans d'autres cancers que les lymphomes [154].

3.1.2 Translocations

Mécanistique

Une translocation chromosomique consiste en l'échange de matériel génétique entre deux chromosomes, et conduit à la juxtaposition de deux parties du génome normalement distantes. Au même titre que les altérations du nombre de copies (§3.1.1), ce phénomène est le résultat de cassures double-brins de la molécule d'ADN prises en charge de manière incorrecte par l'un des nombreux mécanismes de réparation ou de recombinaison à l'œuvre dans la cellule [178]. On distingue les translocations **déséquilibrées**, durant lesquelles un ou plusieurs fragments sont perdus et empêchent donc la réciprocité de l'échange, des translocations **équilibrées** durant lesquelles deux jonctions symétriquement opposées sont réalisées.

Outre les anomalies du nombre de copies pouvant survenir dans le cas des translocations déséquilibrées (§3.1.1), cette juxtaposition peut avoir deux types d'effets : la création d'une **protéine de fusion** ou la **dérégulation de l'expression** des gènes impliqués. Dans le premier cas, les cassures survenant au sein de deux gènes conduisent à la création d'un gène hybride, formé en 5' des exons du premier gène et en 3' des exons du second. La traduction d'un tel gène produit une protéine couplant des caractéristiques des deux gènes, l'exemple emblématique étant la fusion *BCR-ABL1* caractéristique de la LMC, qui permet une activation constitutive de la fonction tyrosine-kinase d'ABL1 grâce au domaine d'homodimérisation de BCR [179]. Dans le second cas, la cassure conduit à rapprocher un oncogène d'une séquence promotrice ou régulatrice de la transcription différente, de sorte à déréguler son niveau d'expression. Les LDGCB offrent de nombreux exemples de ce second phénomène impliquant les gènes d'immunoglobulines, qui seront repris en détail ci-après.

Si les cassures double-brins impliquées dans les translocations peuvent survenir de manière aléatoire selon les mécanismes décrits précédemment (§3.1.1), les lymphopathies y sont particulièrement propices du fait des nombreux mécanismes de réarrangement de l'ADN génomique à l'œuvre dans les lymphocytes B issus du centre germinatif dont elles sont généralement issues [42]. En effet, la **recombinaison des segments V, D et J** des gènes d'immunoglobulines durant la lymphopoïèse repose sur la création et la réparation de cassures double-brins (§1.3.2), et fournit régulièrement et de manière constitutive l'une des deux cassures nécessaire à une translocation. La reconnaissance des sites de clivage des protéines RAG

étant médiée par un motif nucléotidique spécifique (RSS), ce mécanisme peut également être engagé avec d'autres gènes que les immunoglobulines qui présentent des motifs similaires. Si l'implication de ce manque de spécificité a été confirmée pour plusieurs translocations dans les LAL T, impliquant des gènes tels que *LMO1*, *LMO2* ou *TAL1*, son rôle dans les translocations des lymphomes B est encore source de débats [180]. L'**hyper-mutation somatique** due à AICDA (§1.4.5) fournit également de nombreuses occasions à l'apparition de cassures double-brins. En effet, cette enzyme induit la désamination des cytosines et 5-methylcytosines en uraciles et thymines dans des conditions bien spécifiques qui peuvent également être réunies sur d'autres gènes que les immunoglobulines, tels que *BCL6* [181]. La réparation de ces erreurs peut conduire à la formation de cassures simple-brins, pouvant aisément évoluer en cassures double-brins dans les régions d'ADN en cours de transcription favorables à l'action d'AICDA [178]. Le phénomène de **commutation isotypique** met à profit les cassures induites par AICDA dans les régions dites « switch » situées entre les parties constantes d'immunoglobulines, de manière à permettre l'excision d'une ou plusieurs constantes en aval (§1.4.6). Là encore il s'agit d'un phénomène induisant de manière constitutive l'apparition de cassures dans les gènes d'immunoglobulines, pouvant dégénérer en translocation chromosomique.

Translocations des gènes d'immunoglobulines

Les nombreuses cassures physiologiques survenant dans les gènes d'immunoglobulines en font tout naturellement les partenaires de translocation les plus fréquents dans les LDGCB. Bien que le *locus* codant les chaînes lourdes (*IGH*, situé en 14q32.33) soit le plus fréquemment impliqué, ces événements peuvent occasionnellement concerner les *loci* codant les chaînes légères (IGK et IGL, respectivement situés en 2p12 et 22q11.2) [182]. Compte tenu du haut niveau d'expression de ces gènes dans la lignée B, le second point de cassure de ces translocations a généralement lieu près d'un oncogène, qui verra son expression fortement augmentée par la proximité des séquences régulatrices de l'expression des immunoglobulines [182]. Bien que cette sur-expression apporte un avantage sélectif certain aux cellules concernées, il ne s'agit pas d'un événement somatique suffisant à l'apparition d'un LDGCB comme cela a pu être suggéré pour la t(9;22) dans la LMC [179], la preuve en étant la proportion significative de cellules non-tumorales arborant de telles translocations chez les sujets sains [183]. Pour finir, on peut noter que ces réarrangements pathologiques ne sont pas incompatibles avec un réarrangement VDJ fonctionnel sur l'autre allèle, l'expression d'un BCR normal étant nécessaire à la survie de la cellule tumorale dans la majorité des cas [184].

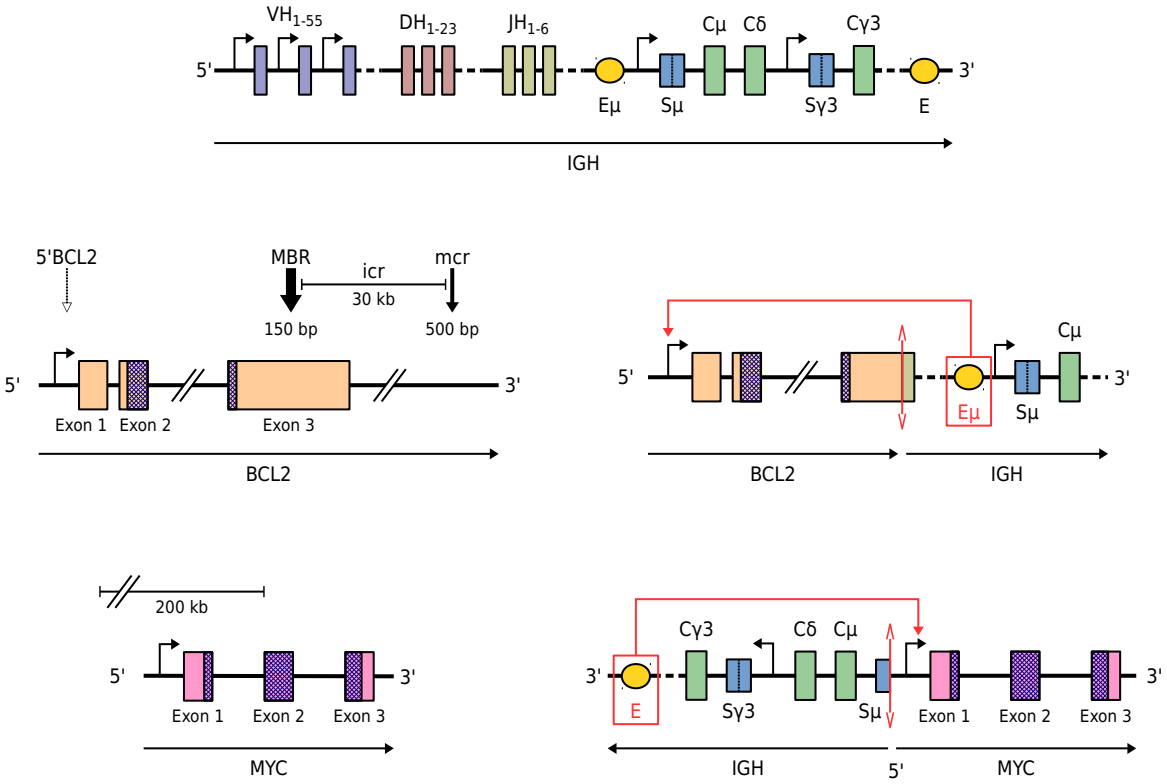


Figure 3.7 – Principales translocations impliquant *IGH*

Le *locus IGH* non réarrangé est présenté en haut, avec une sélection de segments V (*Variable*), D (*Diversity*), J (*Joining*), C (*Constant*), séquences de commutation isotypique S (*Switch*) et amplificateurs transcriptomiques E (*Enhancer*). Le gène *BCL2* est représenté à hauteur intermédiaire, dans sa configuration native mettant en évidence les différents points de cassure possibles (à gauche), et dans le cadre d'une translocation t(14;18) MBR / JH6 (à droite). Enfin le gène *MYC* est décrit en bas, dans sa configuration germinale (à gauche) et dans le cadre d'une translocation t(8;14) Sμ(à droite).

La translocation **t(14;18)(q32;q21)** impliquant l'oncogène *BCL2* est la plus fréquente de ces translocations dans les LDGCB (20% des cas). Cette proportion pourrait s'expliquer par la transformation d'environ 30% des LF en LDGCB [185], cette translocation étant présente dans 80% des LF [186] et majoritairement dans les LDGCB de type GCB, biologiquement proches [187]. Les points de cassure de cette translocation se situent majoritairement en 5' du segment JH6 pour *IGH* [188] et dans la partie 3'-UTR de l'exon 3 de *BCL2*, répartis dans 3 régions de fréquences décroissantes (Figure 3.7) : *Major Breakpoint Region* (MBR), *Minor Cluster Region* (MCR) et *Intermediate Cluster Region* (ICR) [182]. Des cassures en 5' de *BCL2* ont également été rapportées dans de rares LLC, conduisant à des réarrangements 5'/5' préférentiellement avec les chaînes légères d'immunoglobulines [182]. La protéine *BCL2* permettant la survie des lymphocytes B ayant reconnu un antigène dans le centre germinatif, sa sur-expression permanente confère à la cellule tumorale une résistance aux signaux apoptotiques [186]. Cette translocation en elle-même ne semble pourtant pas déli-

miter un sous-groupe de LDGCB plus agressifs, d'autres mécanismes (dont les gains du chromosome 18 observés chez les ABC, §3.1.1) contribuant au même résultat. La sur-expression de la protéine Bcl-2, quelle qu'en soit la cause, a en revanche été démontrée comme pronostique dans les LDGCB traités par R-CHOP [189].

Une translocation **t(8;14)(q24;q32)** menant à la sur-expression du gène *MYC* peut également être observée dans 10% des LDGCB, bien qu'elle soit plutôt caractéristique des lymphomes de Burkitt [182]. Contrairement à la t(14;18) qui semble majoritairement due à une erreur lors du réarrangement VDJ, la t(8;14) serait plutôt due à l'action d'AICDA et aux erreurs de commutation isotypique [190], si l'on fait abstraction des lymphomes de Burkitt endémiques liés à la prévalence particulièrement forte de l'EBV dans certaines régions d'Afrique [191]. Cette dualité de mécanismes et de régions ciblées sur *IGH* autorise par ailleurs d'anecdotiques double-translocations t(8;14;18) associées à un pronostic particulièrement sombre [192]. D'un point de vue biologique, l'expression continue de *MYC* obtenue par le biais de la t(8;14) permet de pallier à sa faible stabilité dans sa configuration non-phosphorylée, et d'accentuer ses effets sur le cycle cellulaire et l'expression d'un large panel de gènes difficiles à identifier. Cette difficulté peut s'expliquer par le caractère inhabituel de *MYC*, qui agirait comme régulateur secondaire de la transcription sur des gènes déjà activés par d'autres mécanismes [193]. Établir l'impact pronostique des translocations de *MYC* en elles-mêmes est délicat dans les LDGCB, du fait de l'importante co-occurrence des translocations de *BCL2* [194].

En effet, les fréquences relativement élevées de ces deux translocations font qu'il est possible d'observer des tumeurs arborant les deux translocations à la fois, dénommées ***double-hits***. Compte tenu de la chronologie des mécanismes mis en jeu, il est probable que la t(14;18) survienne antérieurement à la t(8;14), cette hypothèse étant supportée par l'observation occasionnelle de clones porteurs de la t(14;18) seule chez des patients atteints d'un lymphome *double-hits* [195]. On peut noter également l'implication plus fréquente des chaines légères dans la t(8;14) de ces lymphomes, ce qui est cohérent avec la nécessité d'exprimer un BCR fonctionnel propre à de nombreux lymphomes [195]. Les lymphomes porteurs de ces deux translocations présentent des caractéristiques intermédiaires entre LDGCB et lymphome de Burkitt, et commencent à former une nouvelle entité dans la classification de l'OMS [14]. Cette entité, qui représente moins de 5% des LDGCB, se caractérise par un pronostic particulièrement sombre qui appelle aujourd'hui des solutions thérapeutiques [194, 195]. On peut finalement noter que cette désignation *double-hit* est la source d'une certaine ambiguïté dans la littérature, dans la mesure où elle peut s'étendre ou non à d'autres translocations telles que celles de *BCL6* (§3.1.2), et inclure ou non une translocation de *MYC*.

De nombreux autres gènes ont pu être décrits comme réarrangés de manière récurrente avec les gènes d'immunoglobulines, même si ces anomalies restent rares. La translocation **t(6;14)(p25;q32)** conduisant à la sur-expression d'*IRF4* a pu être mise en évidence dans 20% des myélomes multiples et 5% des LDGCB, et ce préférentiellement dans le sous-type GCB. Bien que ce lien au sous-type GCB en fasse un facteur de bon pronostic, l'impact de l'expression de ce gène emblématique des ABC au sein des GCB n'a pas été décrit [89]. La translocation **t(9;14)(p13;q32)** impliquant *PAX5*, un facteur de transcription contribuant à la lymphopoïèse B, a ainsi pu être occasionnellement mise en évidence dans les lymphomes lymphocytiques et leurs éventuelles transformations en LDGCB [196]. La translocation de *NBEAP1* (*BCL8*) identifiée **t(14;15)(q32;q11-13)** a finalement pu être répertoriée dans 3% à 4% des LDGCB, mais reste encore peu caractérisée à l'heure actuelle [182].

Translocations de BCL6

BCL6 est un partenaire fréquent d'*IGH* dans les LDGCB (5% à 10% des LDGCB présentent une **t(3;14)(q27;q32)**). Son rôle clé dans la prolifération et la survie des lymphocytes B dans le centre germinatif en font une cible de choix pour ces translocations [182]. De nombreux effets de la sur-expression de ce gène ont pu être identifiés : le blocage de la différentiation plasmocytaire par la répression de *PRDM1* et *IRF4*, l'accélération du cycle cellulaire par la répression de *CDKN1A* et *CDKN1B*, ou encore l'inhibition des mécanismes d'inflammation et de réparation de l'ADN [42]. Il est donc surprenant de constater que cette translocation semble plus fréquente dans les ABC [197], d'autant plus que la sur-expression de *BCL6* tendrait à inhiber la voie NFκB pourtant centrale dans ce sous-type [42]. Les effets d'inhibition observés sur l'expression de *BCL2* et *MYC* tendraient cependant à suggérer que d'autres mécanismes compensent ces effets anti-tumoraux [42], écartant l'hypothèse d'une lymphomagénèse déclenchée par un événement isolé.

BCL6 a toutefois la particularité d'être impliqué dans des translocations avec de nombreux autres partenaires dotés de promoteurs ou d'éléments régulateurs alternatifs, selon le même modèle que dans la t(3;14)(q27;q32). Parmi la vingtaine de partenaires actuellement identifiés [198], on retrouve aussi bien des gènes d'intérêt hématologique tels que *PIM1*, *MYC*, *LRMP*, *CIITA* ou *IL21R* que des gènes fortement exprimés de manière non-spécifique tels que *EIF4A2*, *HIST1H4I* ou *GAPDH*. La grande diversité des partenaires et la difficulté d'observer une expression plus marquée de *BCL6* font qu'il est parfois difficile d'établir la présence d'une telle translocation avec certitude. On estime néanmoins que 30% à 40% des LDGCB présentent un tel réarrangement [182].

D'un point de vue clinique, les réarrangements de *BCL6* ont été associés à un pronostic plus favorable dans les LDGCB [199, 200], bien que ces résultats restent à valider à l'ère du Rituximab. On peut noter également qu'une distinction a été suggérée entre translocations de *BCL6* impliquant ou non les immunoglobulines, ces partenaires alternatifs étant associés à un moins bon pronostic qu'IGH [201]. L'association en LDGCB *double* (avec un réarrangement de *MYC*) voire *triple hits* (*MYC* et *BCL2*) définit en revanche des cas de très mauvais pronostic [195, 202].

Protéines de fusions

Bien que l'essentiel des translocations retrouvées dans les LDGCB consistent en une dérégulation de l'expression, des translocations menant à la production de protéines de fusion ont été occasionnellement décrites.

Les translocations impliquant *ALK*, bien que plutôt spécifiques en hématologie des lymphomes T anaplasiques à grandes cellules (LAGC), ont également pu être rapportées dans les LDGCB. Elles définissent un sous-type de LDGCB très rare et de mauvais pronostic, marqué par l'absence des marqueurs CD20 et CD30 caractéristiques des lymphomes B et des LAGC respectivement [203]. Les partenaires d'*ALK* dans ce type de LDGCB sont principalement *CLTC* (**t(2;17)(p23;q23)**, 70% des cas) et *NPM1* (**t(2;5)(p23;q35)**, 10% des cas), bien que d'autres partenaires tels que *RANBP2* (t(2;2)(p23;q12)) aient pu être identifiés de manière anecdotique [204, 205]. *ALK* semble impliqué dans de nombreux réarrangements à travers différentes catégories de tumeurs, avec cependant un schéma commun : la partie cytoplasmique d'*ALK* supportant ses fonctions tyrosine kinase (extrémité 3' codée par les exons 20 à 29) se retrouve associée à une protéine fournissant à la fois un promoteur plus efficace et un domaine d'homodimérisation permettant une auto-activation. Cette substitution du promoteur explique que les partenaires d'*ALK* varient selon les cancers, un partenaire significativement exprimé dans le tissu cellulaire à l'origine de la tumeur étant requis [206].

De nombreux réarrangements de *CIITA*, principal régulateur de la transcription des gènes du MHC de classe 2, ont pu être mis en évidence dans la maladie de Hodgkin, et dans les LPM qui partagent de nombreux caractères génomiques avec cette entité. Bien que ces réarrangements impliquent de nombreux partenaires encore à identifier, il semblerait que la majorité des cassures ait lieu dans le premier intron de *CIITA*, n'offrant à la fusion qu'un exon de *CIITA* et un promoteur susceptible d'entraîner la sur-expression de gènes tels que *CD273* et *CD274*, également décrits comme sur-exprimés par d'autres mécanismes dans ces mêmes pathologies [207]. La présence en FISH d'un réarrangement de *CIITA* a ainsi été mise en évidence dans 38% des LPM et associée à un pronostic défavorable sous CHOP.

Enfin, une translocation **t(3;3)(q26;q28)**, vraisemblablement cryptique en cytogénétique conventionnelle, a récemment été mise en évidence par des techniques de séquençage d'ARN à haut débit dans 5% des LDGCB [208]. Issue d'une inversion interne du chromosome 3, cette translocation remplace les 3 premiers exons de *TP63* par un nombre variable d'exons de *TBL1XR1*. Bien que la conservation du point de cassure dans *TP63*, paralogue de *TP53*, laisse supposer un impact biologique voir pronostique, celui-ci reste encore à déterminer.

3.2 Objectifs et approche suivie

Malgré la disponibilité de séries publiées de LDGCB analysés par puces CGH [22], il nous paraissait important de caractériser les CNA au sein des échantillons utilisés de manière récurrente dans les études de notre unité, afin de pouvoir croiser cette information avec les données déjà recueillies. C'est dans cette optique qu'une série de 108 puces CGH avaient été réalisées quelques années avant mon arrivée, mais n'avaient été que peu exploitées faute d'outils bioinformatiques efficaces.

Les outils d'analyse fournis par le fabricant n'étant pas adaptés à l'étude de larges séries de données, et les outils publiés par la communauté scientifique n'exploitant pas pleinement les spécificités des LDGCB étudiés (forte clonalité, réarrangements modérés, contamination normale variable), nous avons fait le choix de développer une nouvelle suite logicielle dédiée à ce problème. Celle-ci devait inclure un modèle statistique étendant celui décrit (quoi que non-implémenté) par Lenz *et al* [22], tout en offrant une interface graphique simple d'utilisation et en gérant les polymorphismes présents dans une partie des puces de notre série.

Ce projet a ainsi été réalisé en trois phases :

1. L'élaboration d'une bibliothèque logicielle permettant de prendre en charge des données génomiquement localisées comme les gènes, sondes et régions altérées : Rgb (§3.4). Développée en R, langage *open-source* offrant à la fois les fonctionnalités statistiques et graphiques nécessaires à ce projet, celle-ci devait également permettre l'utilisation en ligne de commande et les calculs intensifs pour mes propres besoins.
2. L'extension de cette bibliothèque par une seconde implémentant les différents algorithmes nécessaires à l'analyse de puces CGH : cghRA (§3.5). L'ensemble des classes développées dans cghRA héritent ainsi des classes définies dans Rgb, permettant de séparer la partie graphique de la partie purement calculatoire en deux projets *open-source* indépendants. La mise au point des différents algorithmes a été réalisée au sein des 108 puces du Centre, et a ainsi permis de les caractériser au mieux.

3. L'application des outils et de l'expertise développée durant ce projet à une série de 195 puces à plus haute résolution réalisées dans le cadre du projet GHEDI (§3.6), projet pour lequel des puces transcriptomiques étaient également disponibles.

3.3 Méthodes

3.3.1 Hybridation Génomique Comparative (CGH)

Principe de la technique

L'Hybridation génomique comparative sur puces telle qu'elle est réalisée aujourd'hui est une évolution moléculaire d'une technique homonyme pratiquée initialement en microscopie à fluorescence à l'échelle cellulaire [209]. Ces deux variantes reposent sur l'hybridation compétitive entre deux ADN fragmentés marqués de fluorochromes différents sur une même matrice représentant le génome humain normal. L'excès d'un des deux fluorochromes dans une région donnée permet ainsi d'identifier un nombre supérieur de copies de cette région dans l'ADN correspondant (Figure 3.8). En mettant en compétition un ADN tumoral et un ADN sain, on identifie ainsi les régions gagnées ou perdues par la tumeur au cours de son évolution.

La CGH traditionnelle était pratiquée sur des métaphases de cellules saines, les chromosomes étant ensuite identifiés selon les critères morphologiques et de coloration habituellement utilisés en cytogénétique conventionnelle. Afin d'éliminer les signaux relatifs aux nombreuses régions naturellement répétées du génome, un ADN Cot-1 enrichi en ce type de séquences était également ajouté en excès et sans marquage, avant une lecture informatisée au microscope [209].

Les progrès réalisés en biologie moléculaire ces 20 dernières années ont permis l'avènement des puces à ADN, dont l'application à cette technique a permis d'améliorer grandement sa résolution [210]. Une première étape a été l'utilisation comme matrice d'hybridation de lames de verre sur lesquelles plusieurs milliers de BAC (150 à 350 kb) représentant des régions connues du génome humain étaient déposés sous forme de gouttes individuelles, réparties selon une grille. La lecture précise des deux intensités de fluorescence était ainsi facilitée, et l'utilisation de BAC bien caractérisés permettait une plus grande précision dans la localisation des signaux. Avec l'achèvement du séquençage du génome humain et la mise au point de techniques « d'impression » d'oligonucléotides sur supports solides, il est désormais possible de réaliser cette hybridation compétitive sur plusieurs centaines de milliers de courtes

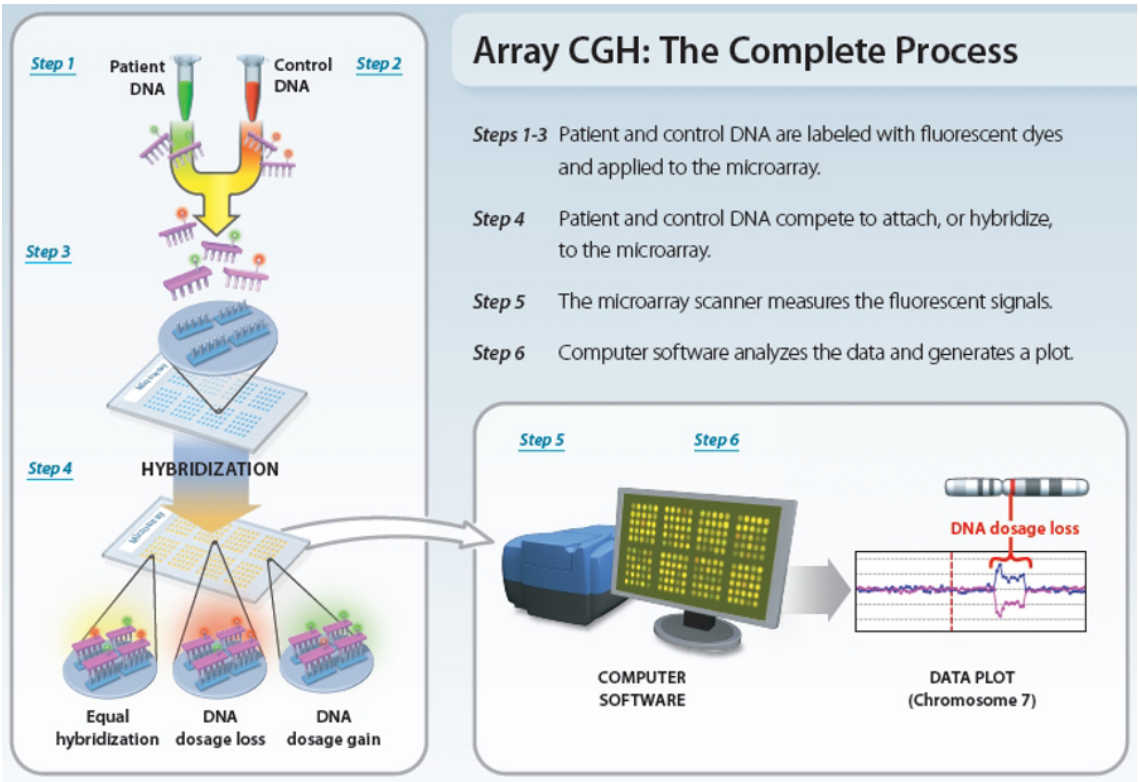


Figure 3.8 – Principe de la CGH

Figure extraite de Theisen *et al*, Nature Education 2008 [210]

régions d'une soixantaine de nucléotides (sondes), offrant une résolution de l'ordre de la centaine de kb.

Analyse des données

L'analyse de résultats issus de cette technique se découpe en plusieurs parties distinctes, pour lesquelles des algorithmes performants sont rendus nécessaires par le nombre croissant de sondes analysées et la complexité calculatoire qui en résulte.

La **normalisation** mathématique des signaux est la première étape, visant à éliminer les nombreux biais qui peuvent exister dans ce genre d'expérience et à faciliter la comparaison de lots de puces différents. La plupart de ces biais sont inhérents aux puces à ADN et ont déjà fait l'objet de développements pour les puces transcriptomiques au début des années 2000 [211], réappliqués ici aux puces CGH. On peut citer à titre d'exemple les artefacts spatiaux, consistant en une intensité accrue d'une fluorescence dans une région de la puce, alors qu'aucun lien n'existe entre position sur la puce et position dans le génome. La relation entre quantité d'ADN et fluorescence n'étant pas parfaitement identique pour les deux fluorochromes, un recalibrage adéquat est également courant dans ce genre d'études. Il résulte généralement de cette analyse pour chaque sonde un log-ratio corrigé (formule 3.1),

correspondant au logarithme à base 2 du ratio entre les intensités de fluorescence respectives des échantillons tumoral et normal (0 en l'absence d'anomalie, négatif en cas de délétion et positif en cas d'amplification).

$$LR_{sonde} = log_2 \frac{fluorescence_{tumeur}}{fluorescence_{reference}} \tag{3.1}$$

La **segmentation** consiste ensuite à grouper les sondes voisines sur le génome qui présentent des log-ratios équivalents, aux imprécisions de mesure près. Résoudre ce problème constitue la partie la plus complexe d'un point de vue algorithmique, le nombre k de segmentations possibles d'un chromosome couvert par n sondes en un nombre variable s de segments tendant rapidement vers l'infini (de l'ordre de 10^{30} et 10^{300} pour 100 et 1 000 sondes respectivement, selon la formule 3.2). L'algorithme CBS développé en 2004 [212] et amélioré en 2007 [213] fournit encore à l'heure actuelle le meilleur compromis entre rapidité et qualité des résultats [214], en appliquant des segmentations binaires successives basées sur le test de Student.

$$k = \sum_{s=1}^{n} C_{n-1}^{s-1} \tag{3.2}$$

Une fois ces segments définis, il est nécessaire de leur associer un **nombre de copies**, en fonction de leurs log-ratios moyens. Bien que certaines équipes se contentent encore aujourd'hui d'appliquer des seuils arbitraires pour déterminer gains et pertes [215, 216], des modèles plus robustes et plus complexes ont vu le jour (§3.5). L'analyse de tissus tumoraux pose en effet des problèmes supplémentaires qui rendent l'utilisation de ces seuils communs impropre, tels que la présence en quantité variable de cellules saines dans les échantillons (marges d'exérèse ou infiltration de la tumeur par des cellules saines) ou encore la présence d'anomalies sous-clonales. Ces deux phénomènes contribuent à faire varier d'un échantillon à un autre les log-ratios à associer aux états amplifié et délété, et nécessitent d'être pris en compte par des modèles paramétrés pour chaque échantillon.

Il convient en outre de rappeler que le log-ratio mesuré traduit une comparaison entre deux ADN, un pool d'ADN normaux étant utilisé comme référence en l'absence d'un prélèvement sain issu du patient correspondant. Dans ce cas, certaines des amplifications et délétions détectées ne reflètent pas la tumeur, mais plutôt la variabilité naturelle entre individus d'une même espèce (polymorphismes). C'est pourquoi une phase de **filtration des polymorphismes** est parfois ajoutée, bien qu'identifier *a posteriori* ces régions soit un problème complexe qui n'a pas encore trouvé de solution satisfaisante (§3.5).

La recherche de **récurrence** au sein d'une série d'échantillons de même nature constitue la dernière étape. Compte tenu de la grande variabilité des points de cassure entre échantillons, il convient d'établir un *modus operandi* pour identifier les événements les plus caractéristiques de la série étudiée. La pratique courante consiste à s'intéresser à la région communément altérée chez un maximum de patients (MCR), dans la mesure où c'est la plus susceptible de contenir le gène ayant permis à ces différents CNA de perdurer au sein de leurs tumeurs respectives. Face à la variabilité de ces régions d'une série à une autre, des méthodes statistiques plus complexes ont là encore été développées pour prendre en compte cette incertitude de position dans les résultats [217].

3.4 Bibliothèque logicielle génomique : Rgb

3.4.1 Choix technologiques

Le traitement des données de CGH repose principalement sur la notion de **position génomique**, chaque sonde de la puce et chaque région résultant de l'analyse étant définie par des coordonnées précises au sein du génome humain (numéro de chromosome, positions de départ et d'arrivée numérotées en paires de bases). L'interprétation des résultats nécessite de plus une annotation des régions amplifiées ou délétées, en terme de gènes impactés ou polymorphismes connus par exemple, là encore définis par des coordonnées de départ et d'arrivée. Ce paradigme nécessite des structures de données et fonctions adaptées, qu'il était nécessaire de mettre en place indépendamment du type de données à manipuler.

Compte tenu du nombre important d'algorithmes d'analyse de puces CGH disponible via Bioconductor [218], R [219] est apparu comme un langage de programmation tout à fait adapté à ce problème. Ses qualités sont nombreuses, et tout à fait en accord avec les besoins de ce projet :

- Sous licence GNU GPL [220], il est entièrement gratuit et **open-source**, soutenu par une communauté internationale de plusieurs milliers d'utilisateurs.
- Il est **compatible** avec Windows, MacOS, Debian, RedHat et Solaris.
- Il intègre nativement les modèles et **tests statistiques** les plus avancés.
- Il intègre un système de **représentation graphique** élaboré, capable de produire des images de qualité sous la plupart des formats d'image courants.
- Il s'utilise nativement en **ligne de commandes**, selon une syntaxe de haut niveau déjà maitrisée par une partie importante de la communauté scientifique.
- Il propose également la création d'**interfaces graphiques**, via Tcl/tk.

— Il propose plusieurs systèmes de **programmation orientée objet**, facilitant la conceptualisation et la maintenance du logiciel.
— Il facilite la création de **packages**, permettant de diffuser les outils développés avec leur documentation et de gérer simplement les dépendances.
— Il est adapté au **calcul intensif**, via l'exploitation de sa syntaxe vectorisée ou via la liaison dynamique à des fonctions écrites dans les langages de plus bas niveaux tels que C ou Fortran.

Bioconductor propose le package « GenomicRange » [221] pour intégrer ce paradigme de position génomique dans R, néanmoins la complexité de son implémentation n'offrait pas un temps de réponse adapté au calcul intensif impliqué dans l'analyse d'une série de puces CGH à haute résolution (voir le test de performance présenté en Figure 1 de l'article, §3.4.4). Une nouvelle implémentation a donc été réalisée *ab initio*, en mettant l'accent sur la rapidité d'exécution.

3.4.2 Implémentation

L'implémentation du « R genome browser » (Rgb) a volontairement été tournée vers la représentation graphique des données génomiques, l'inspection visuelle constituant une part importante de l'analyse des données de CGH. Selon les principes de la programmation orientée objet [222], qui permet de compartimenter les fonctionnalités dans des entités bien distinctes héritant les unes des autres, Rgb a été construit autour d'une classe **track.table** destinée à contenir un groupe d'éléments positionnés dans le génome (Figure 3.9). Cette classe hérite ses fonctionnalités de deux cascades différentes, l'une implémentant le stockage et la manipulation des données et l'autre implémentant l'aspect graphique.

Le stockage des données en mémoire et leur manipulation en temps que tableau classique à deux dimensions sont gérés dans la classe **refTable** (Figure 3.9). Bien que R propose nativement la classe « data.frame » pour gérer ce type de données tabulées, une nouvelle implémentation était nécessaire afin de garantir une meilleure gestion de la mémoire et des ressources de calcul. En effet, le langage R repose sur le principe de « copie à l'écriture », qui veut que les variables ne soient jamais modifiées directement dans la mémoire mais copiées entièrement avant toute modification, de sorte à préserver l'ancienne version des données. Ce fonctionnement est peu propice aux jeux de données génomiques, qui peuvent impliquer plusieurs millions d'exons ou de sondes de CGH au sein d'un génome. La classe « refTable » propose donc un mécanisme de modification *in situ* des données basé sur le système de « références de classes » proposé par R, divisant l'utilisation de la mémoire par 10 et le temps de calcul par 100 (Figure 1 de l'article, §3.4.4).

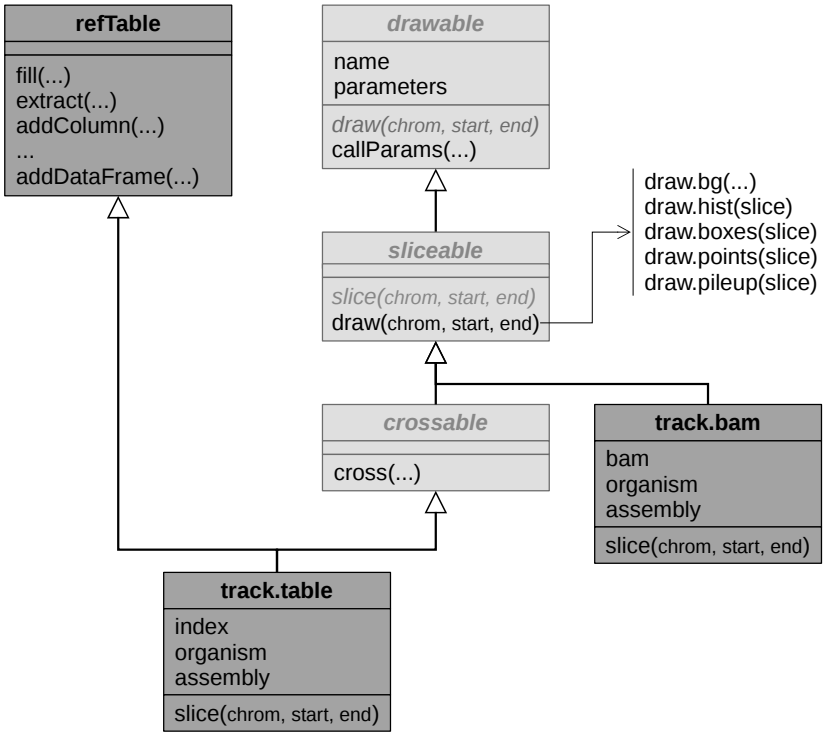


Figure 3.9 – Diagramme UML des principales classes de Rgb

Chaque boîte représente une classe, implémentée (gris foncé) ou virtuelle (gris clair), avec ses principaux attributs et méthodes. L'héritage est symbolisé par une flèche blanche, pointée de la classe fille vers la classe mère.

Le système de représentation graphique repose quant à lui sur la classe virtuelle **drawable** (Figure 3.9), qui permet la prise en charge par les interfaces graphiques et en ligne de commande de Rgb de n'importe quelle classe implémentant une méthode « draw ». Cette méthode se doit de dessiner grâce au système graphique classique de R le contenu de l'objet dans une fenêtre génomique donnée, définie par un numéro de chromosome et des positions de départ et d'arrivée. Cette classe gère également les paramètres de ces différentes représentations, au sein d'un système permettant de définir des valeurs au niveau des fonctions, classes et objets.

La classe drawable est étendue par la classe virtuelle **sliceable**, qui propose une implémentation de cette méthode « draw » basée sur une méthode virtuelle « slice » retournant un sous-ensemble de l'objet correspondant à la fenêtre génomique demandée. Cette méthode constitue le point critique en terme de temps de calcul, dans la mesure où la plupart des méthodes reposent sur des appels fréquents à celle-ci. C'est pour cette raison qu'elle a été implémentée en langage C plutôt qu'en R, au sein de la classe « track.table », permettant de diviser le temps d'exécution d'un facteur 1 000 par rapport à « data.frame » et d'un facteur 10 000 par rapport à « GRanges »

(Figure 1 de l'article, §3.4.4). Plusieurs fonctions sont également fournies pour les représentations les plus classiques (histogrammes, boîtes, nuages de points...), bien qu'il soit aisé à l'utilisateur de définir ses propres fonctions de représentation.

Finalement la classe virtuelle **crossable** apporte une méthode « cross » pour peu qu'une méthode « slice » soit implémentée, permettant l'annotation d'un groupe de données par un autre. Il s'agit d'une tâche récurrente en génomique, la liste des gènes ou le nombre de sondes présents dans chaque région définie dans un jeu de résultats étant une information précieuse à son interprétation.

3.4.3 Résultats et discussion

Rgb permet ainsi la manipulation de données génomiques au sein d'une interface graphique simple similaire aux *genome browsers* actuels (Figure 3.10). Bien que cette partie graphique offre un degré de personnalisation rarement retrouvé dans les autres logiciels disponibles, Rgb se caractérise avant tout par son interface en ligne de commande et les différents degrés de mélanges qu'il autorise entre ces deux interfaces. Une utilisation majoritairement via l'interface graphique pourra ainsi être occasionnellement étendue par l'utilisation conjointe de la console, afin d'étendre ses possibilités. Inversement, une utilisation majoritairement en console pourra être facilitée par l'appel ponctuel de sous-parties de l'interface graphique, pour choisir ou modifier les pistes de données étudiées par exemple. Enfin une utilisation entièrement en ligne de commande permettra l'automatisation de tâches répétitives, telles que la génération de représentations graphiques fixes de chacune des centaines de régions altérées dans une série de puces CGH.

Rgb est actuellement distribué gratuitement dans sa version 1.4.1 sous licence GPL, via GitHub [223] et une page dédiée [224], et ses fonctionnalités continuent d'être étendues selon mes besoins. Il a servi de base au développement de cghRA (§3.5), la plupart des classes de ce dernier héritant directement des classes de Rgb.

3.4.4 Article publié (Bioinformatics 2014)

Rgb a fait l'objet d'une communication affichée en août 2013 à Berlin, lors des congrès conjoints des sociétés internationale et européenne de biologie computationnelle (ISMB/ECCB) [225]. Il a également été publié dans le journal *Bioinformatics* (IF 4,98) en avril 2014 sous forme d'une *Application Note* [226]. Une formation a également été organisée pour les membres de l'unité intéressés en février 2015, dans le cadre d'une formation plus large au langage R.

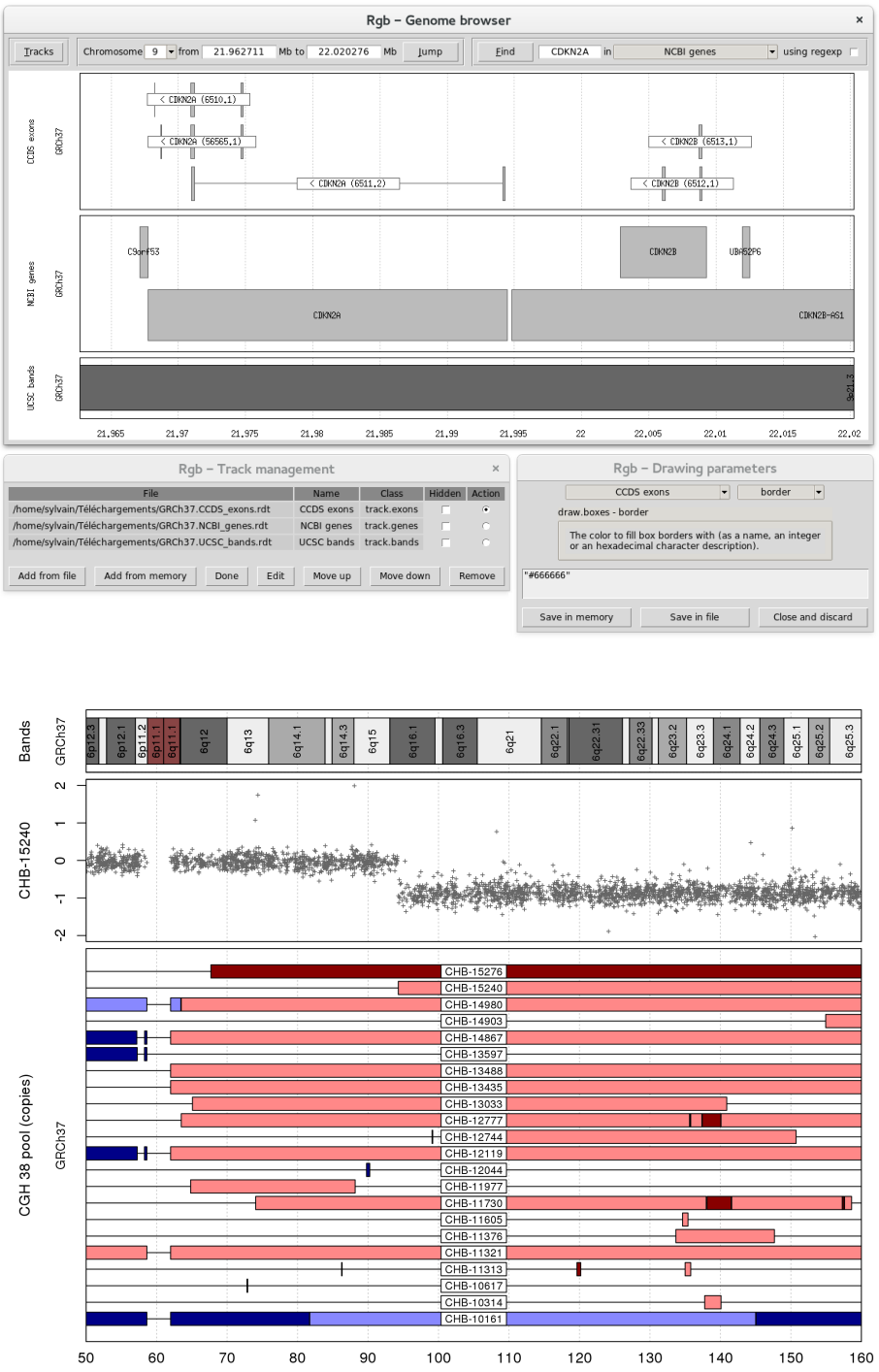


Figure 3.10 – Exemples de représentations générées par Rgb

La **moitié supérieure** présente l'interface graphique de Rgb, avec trois pistes d'annotation (bandes chromosomiques, exons et transcrits) représentées dans la région 9p21 contenant *CDKN2A*. Une partie de la délétion 6q est représentée dans la **moitié inférieure**, mettant en vis-à-vis les bandes chromosomiques, les log-ratios de fluorescence pour une puce donnée (CGH-15240) et l'ensemble des altérations retrouvées dans cette région parmi les 108 puces du Centre Henri Becquerel (délétions en rouge, amplifications en bleu).

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 15 2014, pages 2204–2205
doi:10.1093/bioinformatics/btu185

Genome analysis

Advance Access publication April 9, 2014

Rgb: a scriptable genome browser for R

Sylvain Mareschal[1,2,3,*], Sydney Dubois[1,2,3], Thierry Lecroq[2,3,4] and Fabrice Jardin[1,2,3,*]

[1]Centre Henri Becquerel, INSERM UMR 918, 76038 Rouen Cedex 1, France, [2]Normandy University, University of Rouen, 76821 Mont-Saint-Aignan, France, [3]Institute for Research and Innovation in Biomedicine (IRIB), Haute-Normandie, 76183 Rouen Cedex, France and [4]LITIS, INSA EA 4108, 76801 Saint-Etienne-du-Rouvray, France

Associate Editor: Inanc Birol

ABSTRACT

Summary: Thanks to its free licensing and the development of initiatives like Bioconductor, R has become an essential part of the bioinformatics toolbox in the past years and is more and more confronted with genomically located data. While separate solutions are available to manipulate and visualize such data, no R package currently offers the efficiency required for computationally intensive tasks such as interactive genome browsing. The package proposed here fulfills this specific need, providing a multilevel interface suitable for most needs, from a completely interfaced genome browser to low-level classes and methods. Its time and memory efficiency have been challenged in a human dataset, where it outperformed existing solutions by several orders of magnitude.

Availability and implementation: R sources and packages are freely available at the CRAN repository and dedicated Web site: http://bioinformatics.ovsa.fr/Rgb. Distributed under the GPL 3 license, compatible with most operating systems (Windows, Linux, Mac OS) and architectures.

Contact: maressyl@gmail.com or fabrice.jardin@chb.unicancer.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on December 20, 2013; revised on March 3, 2014; accepted on April 2, 2014

1 INTRODUCTION

The growing demand from the biology community for statistically robust approaches has made the R statistics-oriented scripting language an essential part of the bioinformatics toolbox. Its graphical capabilities make it a valuable tool to produce publication-grade complex figures, whereas its computational efficiency allows it to handle huge datasets, as currently required in fields like transcriptomics or next-generation sequencing. These qualities come with an open-source licensing and various operating system ports that make it available virtually everywhere. Thanks to the Bioconductor initiative (Gentleman *et al.*, 2004), large amount of software is freely available as R packages for tasks as diverse as microarray processing (Smyth, 2005; Venkatraman and Olshen, 2007), feature annotation (Zhang *et al.*, 2003) or sequence analysis (Anders and Huber, 2010).

Much of this software generates genomic data, i.e. lists of chromosome regions defined by starting and ending coordinates. Such data are usually subset using chromosomal coordinates rather than row indexes, a paradigm R was not developed to deal with. Bioconductor historically addressed this issue with the *RangedData* class from the IRanges package, handling genomic regions as ranges of integers (base positions). Its flexibility and efficiency were extended a few years later by the GenomicRanges package, making direct use of IRanges components for the subsetting. For visualization, Bioconductor provides two solutions: rtracklayer and GViz. The former sends data to visualize to the UCSC web genome browser (Kent *et al.*, 2002), a model that implies frequent comings and goings between programs and consequent network burden. The latter is more integrated within Bioconductor and produces static graphics from the classes described above, in delays incompatible with user interactivity and intensive computing.

The package described here reconciles these two aspects in a coherent way, thus offering an interactive interface responsive enough for comfortable browsing and atomic operations suitable for computer-intensive algorithms.

2 IMPLEMENTATION

Rgb is implemented as an R package, providing various classes and methods for scripts. It makes use of the Reference class system, which offers an object-oriented framework similar to what can be found in other languages such as Java or C++. It is flexible enough to suit the needs of the four categories of R users:

- **R beginners** will find in Rgb a complete graphical user interface (GUI) allowing them to convert and visualize genomic data without any command line. Stand-alone builds of Rgb can even make the R dependency totally transparent.
- **Script writers**, needing an exclusive Command Line Interface to automate analysis, will find in Rgb classes and methods able to handle their genomic data and produce high-quality graphics from them.
- **Console users** who use R as an exploratory tool without a special need for reproducibility, will find Rgb's ability to mix the two interfaces particularly time saving.
- **Package developers** may be interested in extending classes to handle new data storage modes or representations, a process greatly facilitated by the object-oriented design of Rgb.

The core of the subsetting system consists of a collection of C functions making direct use of R libraries, and the graphical interface relies on Tcl-tk. Both of them are natively managed

*To whom correspondence should be addressed.

 © The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at INSERM on July 28, 2015

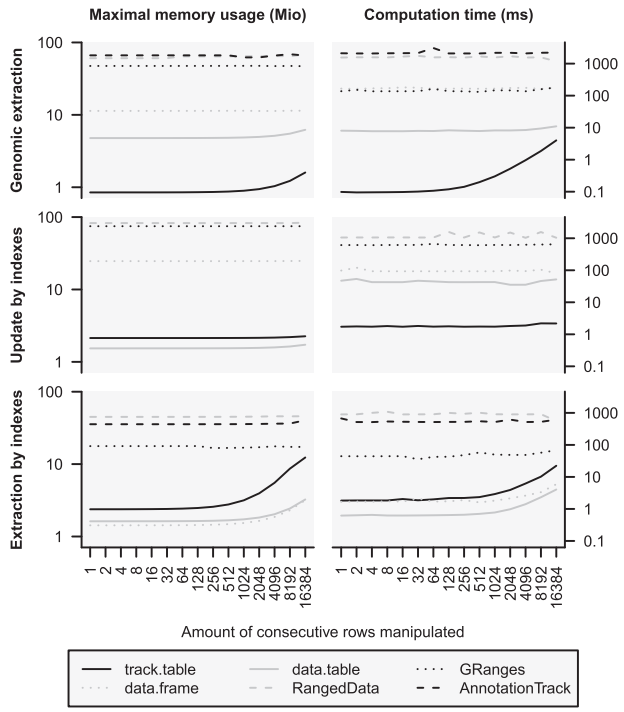


Fig. 1. Performance comparison of Rgb's *track.table* class with existing software. Three atomic tasks were benchmarked on a common dataset, in terms of maximal memory usage and computing time, as recorded by R

by R, and have been successfully compiled and tested on Windows and Linux operating systems. BAM file support can be added by the Rsamtools package, which is part of Bioconductor.

3 PERFORMANCES

To assert Rgb's *track.table* class suitability for computer-intensive tasks such as responsive genome browsing or overlap computation, its time and maximal memory consumption on usual atomic tasks were compared with existing solutions: standard R *data.frame*, the more efficient *data.table*, IRanges' *RangedData*, GenomicRanges' *GRanges* and GViz's *AnnotationTrack*. Three tasks were monitored as the most common atomic operations in such data: genomic extraction by chromosomal coordinates, extraction and modification of consecutive rows by indexes. The 291 128 exons recorded in the CCDS database (Pruitt *et al.*, 2009) for the human genome were used as a common dataset. Benchmarking was performed on a mid-range desktop computer (3.1 GHz Intel i3-2100 processor with 8 GB RAM) running R 3.0.2 in a 64-bit Fedora 18 distribution. Each measure was made with R functions *proc.time* and *gc* in a fresh session, to normalize garbage collection effects (scripts and dataset available as Supplementary Data).

With the genomic extraction task (Fig. 1), GViz and *RangedData* proved their unsuitability to intensive tasks, with computing times nearing the second for a single extraction (typical genome browser representations contain several tracks to be subset), and extravagant memory usage (hundreds of Mio for a dataset of 30 Mio). *GRanges* and generic R solutions perform better, but are still outperformed by Rgb by a magnitude of 50–1500 on small genomic extractions.

Similar results can be observed with the modification task, with Rgb outperforming R generic classes by a factor of 100 in time consumption and 10 in memory usage. It still outperforms bioinformatic solutions with the last task, but is overtaken here by generic solutions such as *data.frame* and *data.table*. However, as this kind of extraction is far less justified in a genomic context than the previous one, this average performance can be accepted.

4 CONCLUSION

Rgb provides several entry points to a consistent genome browsing system, which may prove equally useful to users looking for an interactive genome browser able to handle their R datasets and developers needing efficient genomic subsetting capabilities for their scripts. Its object-oriented paradigm and open-source licensing make it easily extendable, and its performances over existing solutions have been proven in a real genomic dataset.

ACKNOWLEDGEMENTS

The authors thank Christian Bastard for his helpful comments and feedback during Rgb development and the many CGH-array datasets he provided for testing.

Funding: This work was supported by a PhD grant from the Région Haute-Normandie (France).

Conflict of Interest: none declared.

REFERENCES

Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.

Gentleman,R.C. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.

Kent,W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

Pruitt,K.D. *et al.* (2009) The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.*, **19**, 1316–1323.

Smyth,G.K. (2005) limma: linear models for microarray data. In: Gentleman,R. *et al.* (ed.) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Statistics for Biology and Health*. Springer, New York, NY, pp. 397–420.

Venkatraman,E.S. and Olshen,A.B. (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*, **23**, 657–663.

Zhang,J. *et al.* (2003) An extensible application for assembling annotation for genomic data. *Bioinformatis*, **19**, 155.

Downloaded from http://bioinformatics.oxfordjournals.org/ at INSERM on July 28, 2015

3.5 Bibliothèque logicielle d'analyse CGH : cghRA

3.5.1 Vue d'ensemble

cghRA a été implémenté sous la forme d'un package R, qui permet son exploitation en ligne de commande ou via une interface graphique (Figure 3.11). Une version *standalone* pour Windows est également en développement, se contenant de lancer R en tâche de fond et d'invoquer cette interface et mimant ainsi le fonctionnement d'un logiciel traditionnel. Cette interface découpe l'analyse en 4 phases :

— La préparation du fichier **design**, qui vise à reformater les informations relatives à la puce utilisée telles qu'elles sont fournies par le fabricant. Cette étape peut inclure de manière optionnelle le réalignement des sondes via le logiciel BLAT [227], afin de travailler avec une version spécifique du génome de référence, potentiellement plus récente. Le pré-calcul des biais de proportions en GC soustraits par l'algorithme WACA [228] est également optionnel.
— L'analyse des **puces**, consistant à calculer le log-ratio pour chaque sonde et les segmenter en régions chromosomiques présentant un log-ratio équivalent, auxquelles un nombre de copies sera attribué via l'algorithme que nous avons développé (§3.5.2).
— La **validation manuelle des modèles** de prédiction du nombre de copies, via une interface permettant de modifier les paramètres proposés par notre algorithme et d'en observer instantanément les effets sur la distribution des segments (Figure 3.12).
— L'analyse d'une **série de puces**, via des algorithmes de mise en évidence de régions communément altérées tels que SRA et LRA [22] ou l'algorithme STEPS encore en développement (§3.5.4). La filtration préalable des segments suspectés d'être des polymorphismes plutôt que des événements somatiques sera également proposée à cette étape (§3.5.3).

Dans la mesure où le logiciel *Feature Extraction* utilisé pour la lecture des puces inclut une phase de normalisation du signal, nous n'avons pas souhaité inclure de phase supplémentaire de régression des log-ratios en fonction de l'intensité de la fluorescence comme il est d'usage en CGH [211]. Le contrôle des biais spatiaux, consistant en une proximité anormale sur la puce des sondes présentant un log-ratio élevé, est réalisé visuellement par la production automatisée d'un graphique de distribution. Compte tenu de l'absence de tels biais dans notre série, nous n'avons pas implémenté d'outils de correction dans cghRA, bien que le package MANOR [229] puisse être envisagé si cela devenait nécessaire. Enfin la répartition en « vagues » des valeurs de log-ratio qui peut parfois être observée en CGH est prise en charge via

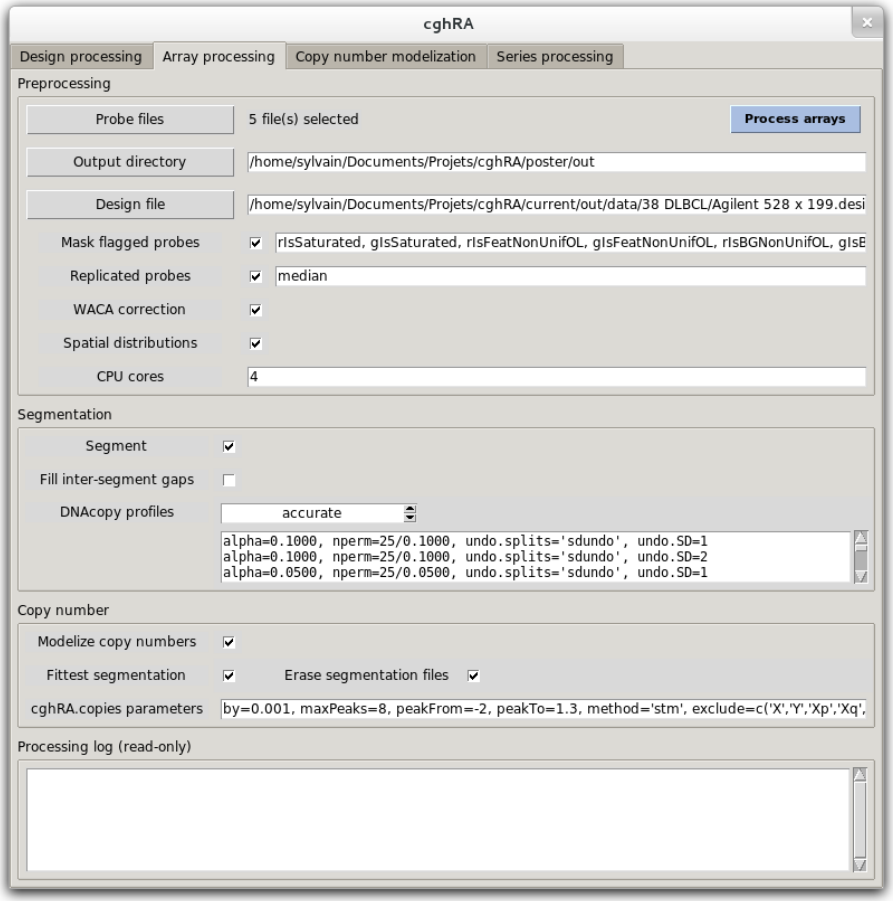


Figure 3.11 – Interface de cghRA

une nouvelle implémentation de l'algorithme WACA [228] dans cghRA. Cet algorithme se propose de gommer ce phénomène en appliquant 5 lissages LOWESS successifs aux log-ratios, gommant des liens non désirés entre ce signal d'intérêt et différentes caractéristiques de la sonde en elle-même et de la région dans laquelle elle se trouve (proportions de bases G et C notamment).

La mise au point de cghRA a été réalisée sur 108 puces commercialisées par *Agilent Technologies*, réparties en 3 séries impliquant chacune un design différent. Une première série de 19 puces 44B (44 290 sondes) regroupait autant de LDGCB réfractaires, et avait déjà fait l'objet d'une première publication en temps qu'outil de validation pour une technique de PCR quantitative [230]. Une seconde série de 51 puces 4x44K (45 220 sondes) a permis de compléter cette série avec des LDGCB « tout-venants », ces deux séries étant hybridées contre un pool d'ADN Promega. Une troisième série plus récente de 38 puces 105A (105 072 sondes) avait en revanche était hybridée contre les ADN constitutionnels appariés des patients étudiés, est étaient donc exempts de tout polymorphisme.

3.5.2 Modélisation du nombre de copies

Modèle théorique

Comme de nombreux outils actuellement disponibles [231–233], cghRA repose sur un modèle de mélange gaussien. En se basant sur la formulation du log-ratio (formule 3.1), on attendrait en effet des valeurs discrètes, correspondant aux différents nombres de copies possibles : $log_2(1/2) = -1$ pour une délétion hétérozygote, $log_2(2/2) = 0$ pour un segment normal, $log_2(3/2) = 0,58$ pour un gain d'une copie, etc. cghRA considère ainsi qu'un nombre inconnu de ces états sont représentés au sein du génome étudié, et que les variations mesurées sont le résultat d'un bruit technique réparti selon une loi normale autour de la valeur attendue. Si certains modèles appliquent cette modélisation aux log-ratios des sondes [231], nous avons fait le choix de l'appliquer aux log-ratios moyens des segments, le bruit technique étant largement diminué par la segmentation.

Une simple transformation mathématique permet d'extraire de ces log-ratios des nombres de copies mesurés (NCM), considérant que chaque sonde correspond à un point unique du génome et présente donc 2 copies dans un ADN normal (équation 3.3). Selon le modèle décrit précédemment, ces NCM suivent une loi normale dont la moyenne correspond au nombre de copies réel (NCR) du segment considéré, et l'écart-type à un bruit technique fixe (équation 3.4).

$$\left\{ \begin{array}{l} LR = log_2(tumeur/reference) \\ reference = 2 \end{array} \right\} \Rightarrow NCM = tumeur = 2 * 2^{LR} \tag{3.3}$$

$$\begin{array}{l} NCM \sim \mathcal{N}(NCR, \sigma^2) \\ NCR \in \mathbb{N} \end{array} \tag{3.4}$$

L'analyse d'échantillons tumoraux issus de biopsies pose cependant une difficulté supplémentaire : l'effet de dilution. Le tissu analysé est un mélange entre des proportions variables de cellules normales (p_{nrm}) et tumorales (p_{tum}), le log-ratio mesuré (LR_{obs}) correspond donc à une moyenne du signal observé sur le pool de cellules de l'échantillon. Considérant que les cellules normales présentent un log-ratio de 0 pour l'ensemble du génome, les log-ratios mesurés ont ainsi tendance à tendre vers 0 à mesure que la contamination en cellules normales augmente (équation 3.5).

$$\left\{ \begin{array}{l} LR_{obs} = LR_{tum} * p_{tum} + LR_{norm} * p_{nrm} \\ LR_{nrm} = 0 \\ p_{tum} + p_{nrm} = 1 \end{array} \right\} \Rightarrow LR_{obs} = LR_{tum} * (1 - p_{nrm}) \tag{3.5}$$

Il convient donc d'annuler cet effet de dilution, en incluant dans le modèle un facteur d'échelle directement lié à la proportion en cellules tumorales ($ecart$). Cette proportion pouvant varier du simple au double d'un échantillon à un autre, chaque puce fait l'objet d'une modélisation indépendante dans cghRA, par opposition à d'autres outils [233]. Un paramètre de recentrage ($centre$) est également inclus, afin de gommer de légers décalages de la moyenne de l'état normal. Enfin un paramètre de ploïdie ($ploidie$) définissant le nombre de copies le plus récurrent au sein du génome est laissé à l'appréciation de l'utilisateur, dans la mesure où la CGH ne permet pas de compter les copies de manière absolue. Dans les LDGCB, cette ploïdie est généralement de 2, et peut être modifiée *a posteriori* si plus de deux états délétés sont présents, ou laissée à 0 pour obtenir des nombre de copies relatifs. Le modèle de cghRA peut ainsi être résumé par l'équation 3.6, présentant deux paramètres clés ($ecart$ et $centre$) et deux paramètres accessoires ($ploidie$ et σ^2).

$$\begin{aligned} &NCR_{segment,puce} = \frac{NCM_{segment,puce} - centre_{puce}}{ecart_{puce}} + ploidie_{puce} + \epsilon \\ &\epsilon \sim \mathcal{N}(0, \sigma^2_{puce}) \end{aligned} \tag{3.6}$$

Application du modèle

D'un point de vue plus pratique, les paramètres du modèle sont estimés à partir de la densité de distribution des segments en fonction de leur log-ratio. La fonction de densité est reconstruite par superposition de noyaux gaussiens (fonction « density » implémentée dans R), pondérant chaque segment par le nombre de sondes qui le composent (Figure 3.12). Ce calcul étant réalisé dans un espace linéaire (NCM) et non pas logarithmique (log-ratio), cette fonction de distribution présente une succession de pics espacés de manière régulière, l'écart entre les pics correspondant au paramètre $ecart$ discuté précédemment. Les *maxima* locaux sont ainsi détectés dans cette distribution, et l'écart médian entre 2 pics consécutifs utilisé comme estimateur de ce paramètre. Le pic correspondant à la densité la plus forte est considéré comme le centre du modèle, et son NCM sert d'estimateur au paramètre $centre$.

Dans la mesure où plusieurs paramétrages de l'algorithme CBS utilisé pour la segmentation [213] peuvent produire des résultats très différents, plusieurs modèles sont comparés pour chaque puce. De même la reconstruction de la fonction de distribution dépend d'un paramètre *bandwidth* critique, pour lequel il n'est pas toujours évident d'identifier *a priori* une valeur optimale. Le modèle minimisant ϵ est ainsi sélectionné, calculant pour chaque modèle une distance STM selon l'équation 3.7.

$$STM = \sum \left(\frac{nSondes_{segment}}{nSondes_{genome}} * |NCR_{segment} - arrondi(NCR_{segment})| \right) \tag{3.7}$$

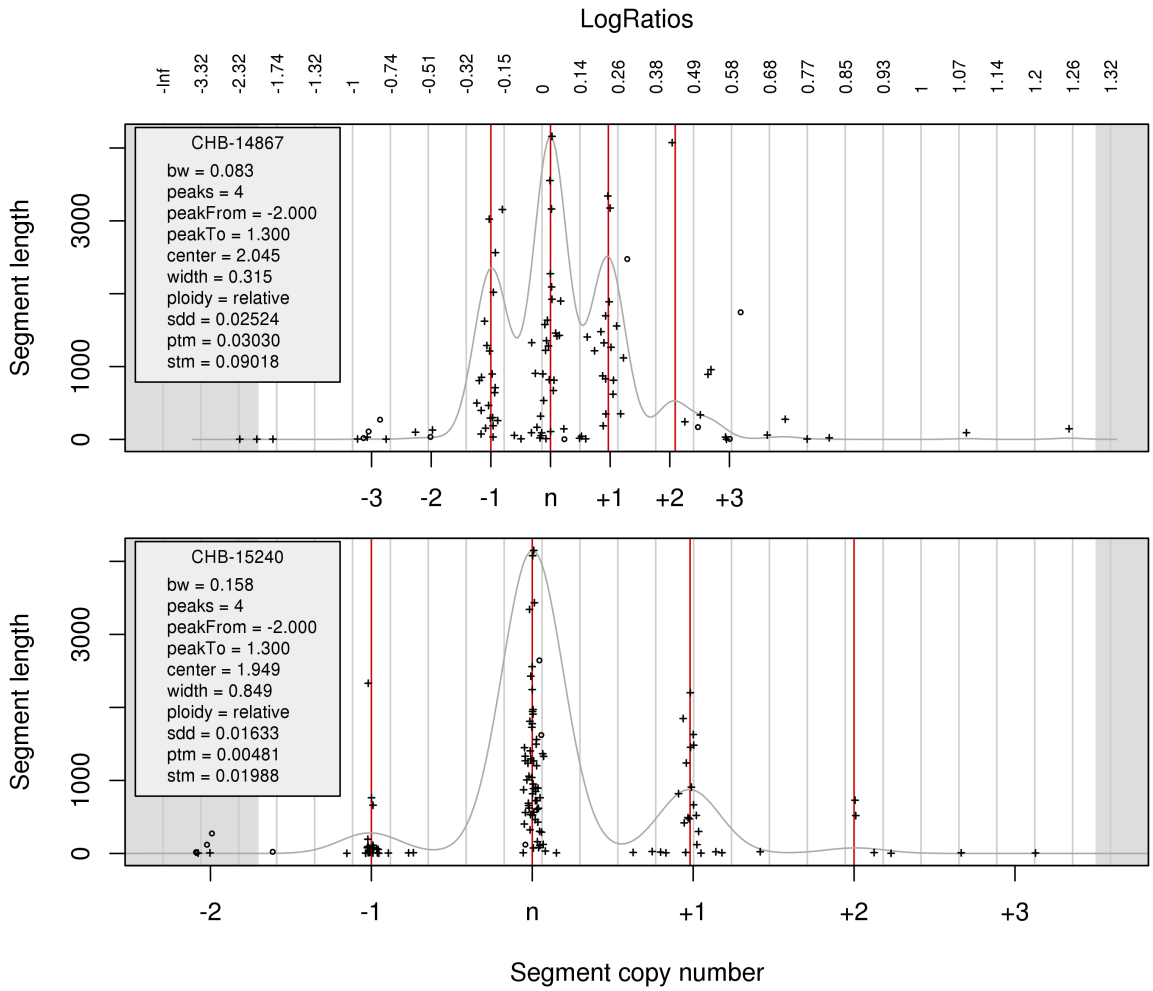


Figure 3.12 – Modèles de prédiction du nombre de copies

Deux modèles de prédiction du nombre de copies sont présentés, afin d'illustrer les différences de pureté tumorale entre échantillons. Pour chaque patient, les segments chromosomiques sont représentés par des croix, dont l'abscisse correspond au log-ratio moyen (échelle du haut) et l'ordonnée au nombre de sondes de la puce impliqués dans ce segment. La densité de distribution de ces segments est mise en évidence par une courbe grise. Les traits rouges correspondent aux pics de cette distribution, auxquels sont associés un nombre de copies (échelle du bas) relatif à la ploïdie (n=2 généralement). Les cercles correspondent à des segments situés sur les chromosomes sexuels, qui ne respectent pas le modèle.

Validation du modèle

Les prédictions opérées par cghRA ont été comparées à celles produites par deux outils également implémentés en R, l'un présentant un modèle similaire (CGHcall [233]), l'autre étant considéré comme une référence dans le domaine (GLAD [234]). Afin de départager ces 3 modèles, les résultats de caryotypage conventionnel ont pu être obtenus pour 77 des 108 patients analysés par CGH. La distance entre le nombre de copies prédit par un algorithme et le nombre de copies rendu par le caryotype a ainsi été calculée pour chaque patient, selon l'équation 3.8. Si le caryotypage

conventionnel n'offre pas une résolution aussi forte que la CGH, la pondération de cette distance par le nombre de sondes impliquées met l'accent avant tout sur les différences touchant les régions les plus larges, en supposant que le seuil de détection reste le même pour les petites régions.

$$d = \sum \left(\frac{nSondes_{segment}}{nSondes_{genome}} * |CN_{CGH,segment} - CN_{caryotype,segment}| \right) \tag{3.8}$$

Afin de réaliser cette comparaison, les résultats de cghRA ont d'abord été convertis. CGHcall et GLAD reposent en effet sur un nombre fixe d'états, 3 pour GLAD (gain, perte et normal) et 5 pour CGHcall (délétion homozygote et hétérozygote, normal, gain et amplification). Les nombres de copies rendus par cghRA ont ainsi été ramenés à 3 ou 5 états selon l'algorithme auquel il était comparé, et il est intéressant de remarquer que contrairement à ses concurrents cghRA offre un nombre virtuellement illimité d'états, sans que la complexité du modèle n'augmente.

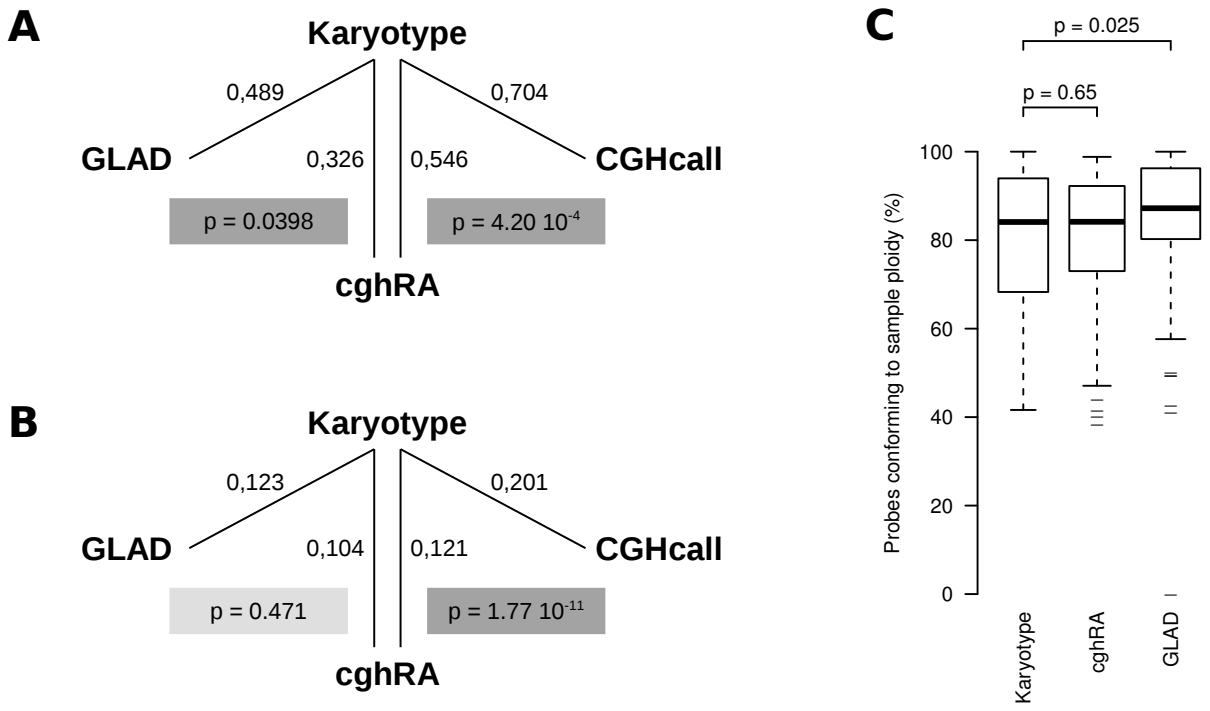


Figure 3.13 – Validation des prédictions du nombre de copies

Le panneau **A** s'intéresse aux 17 patients hyperdiploïdes, le panneau **B** aux 49 diploïdes. Ces deux panneaux représentent pour chacun des 3 algorithmes testés (GLAD, cghRA et CGHcall) la différence moyenne du nombre de copies prédit à celui obtenu par caryotypage conventionnel. Les algorithmes sont comparés 2 à 2 afin d'établir si cette distance est significativement plus faible pour l'un ou l'autre (p valeurs sur fond gris). Le panneau **C** décrit dans 66 puces la proportion des sondes pour lesquelles est prédit un état de ploïdie normale (2 copies pour la majorité des cas) par 3 méthodes différentes : le caryotype conventionnel (inféré à partir des coordonnées cytogénétiques des anomalies), cghRA et GLAD.

Comparé à CGHcall, cghRA démontre une distance significativement plus faible au caryotype, sur des échantillons hyperdiploïdes (Figure 3.13A) ou diploïdes (Figure 3.13B). La différence entre cghRA et GLAD reste marginale, non-significative chez les diploïdes et légèrement en faveur de cghRA chez les hyperdiploïdes (p=0,04). Nous avons cependant pu noter que GLAD semblait plus conservateur que cghRA, dans la mesure où un état normal était prédit sur des proportions significativement plus élevées des puces que ce que le caryotype suggère (Figure 3.13C).

Pour conclure sur la validité de ce nouveau modèle, il convient de remarquer que 13 des puces n'ont pas permis d'établir un modèle robuste avec cghRA (12%). Ces cas correspondent soit à des génomes trop peu réarrangés pour permettre la mise en évidence de plusieurs pics dans la distribution des log-ratios, soit à des échantillons fortement bruités ou combinant plusieurs sous-clones. Si CGHcall et GLAD proposent effectivement des prédictions pour ces échantillons, leur pertinence reste à démontrer. La détection et la prise en charge des modèles marginalement réarrangés est actuellement une piste de développement pour cghRA, qui pourrait réduire d'environ deux tiers cette proportion d'échantillons exclus.

3.5.3 Filtration des polymorphismes

Dans la mesure où 70 des 108 puces étudiées ont été hybridées contre des pools d'ADN témoins plutôt que l'ADN constitutionnel apparié, il est possible qu'une minorité d'événements soient en fait des polymorphismes. Afin de filtrer ces événements, leur comparaison avec la collection de variations constitutionnelles du nombre de copies (CNV) décrite dans la *Database of Genomic Variants* [235] a été implémentée dans cghRA. Ce type d'approche présente cependant un certain nombre de risques [236, 237], qui ont nécessité quelques précautions.

Ces données sont en effet issues de techniques variées, qui présentent des résolutions très différentes et donc un degré de précision variable quant à la localisation des points de cassure. Le jeu de données a ainsi été limité à 3 015 103 de CNV obtenus par des techniques jugées précises (séquençage de nouvelle génération, puces CGH ou SNP oligonucléotidiques), dans des séries de grande envergure (20 patients au moins). Les coordonnées de ces CNV ont ensuite été converties, de leur système de coordonnées génomiques originel en un ensemble de sondes de la puce CGH en cours d'analyse qui seraient potentiellement touchées par cet événement. La comparaison entre un CNV connu et un segment issu de cghRA est ainsi réalisée grâce au coefficient de Jaccard, considérant les sondes communes et propres à chacun (équation 3.9). Ce coefficient varie entre 0 (aucun chevauchement) et 1 (identité parfaite).

$$J_{CNV} = \frac{segment \cap CNV}{segment \cup CNV} \tag{3.9}$$

Dans la mesure où un segment observé chez un patient peut résulter de plusieurs CNV voisins, un certain degré de combinaison a été autorisé. Tout ajout d'un CNV permettant d'augmenter de 0,1 au moins le score est ainsi autorisé, une pénalité progressive étant appliquée au coefficient de Jaccard (3.10). Compte tenu de cette pénalité exponentielle, seules les combinaisons présentant une forte homologie avec le segment analysé sont autorisées, impliquant rarement plus de 2 ou 3 CNV.

$$J_{combinaison} = \left(\frac{segment \cap (CNV_1 \cup CNV_2 \ldots CNV_n)}{segment \cup (CNV_1 \cup CNV_2 \ldots CNV_n)} \right)^n \tag{3.10}$$

La récurrence d'un CNV au sein de la DGV est enfin prise en compte, cette dernière étant utilisée comme un pool de segments dans lequel l'algorithme puise pour construire un maximum de combinaisons de CNV ressemblant au segment analysé. Les scores des n combinaisons produites sont ensuite additionnés pour attribuer au segment un score final, le $cnvScore$. Lors de ce calcul, ces combinaisons sont traitées par ordre décroissant de score J, et pondérées par leur ordre d'arrivée (équation 3.11). Ceci permet au $cnvScore$ de privilégier la qualité des chevauchements sur leur quantité, dans la mesure où la puissance réduit rapidement à une quantité négligeable les scores J inférieurs à 1. Cette approche permet également de maitriser l'échelle, certains CNV étant décrits plusieurs milliers de fois, tout en offrant la possibilité de gagner du temps en ne calculant qu'un nombre réduit de combinaisons.

$$cnvScore = \sum_{i=1}^{i=n} J_i^i \tag{3.11}$$

L'efficacité de ce score a été mise à l'épreuve dans une série composite, comprenant des segments d'origine somatique et constitutionnelle. Les 1 020 CNA mis en évidence dans la série de 38 puces CGH hybridées contre l'ADN constitutionnel apparié ont été utilisés comme segments somatiques, en se limitant aux 31 puces analysables par cghRA (§3.5.2). La série ESTD212 de la banque NCBI dbVar, totalisant 71 178 segments détectés par puces SNP CytoScan HD chez 873 volontaires sains [238], a été utilisée comme source de segments constitutionnels extérieure à la DGV. 8 325 de ces segments couvraient au moins 2 sondes de la puce 105A utilisée pour la recherche de segments somatiques, et furent retenus pour l'analyse. Avec un seuil de 1, ce score permet ainsi de faire la dichotomie entre segments somatiques et constitutionnels avec une sensibilité de 86,7% et une spécificité de 89,1%, l'aire sous la courbe ROC étant de 92,8% (Figure 3.14 gauche). Environ 10% des CNV ne

présentaient aucun chevauchement dans la DGV (score de 0), ce qui souligne qu'en dépit des milliers d'échantillons inclus dans cette base elle peine encore à répertorier toute la variabilité constitutive du génome humain.

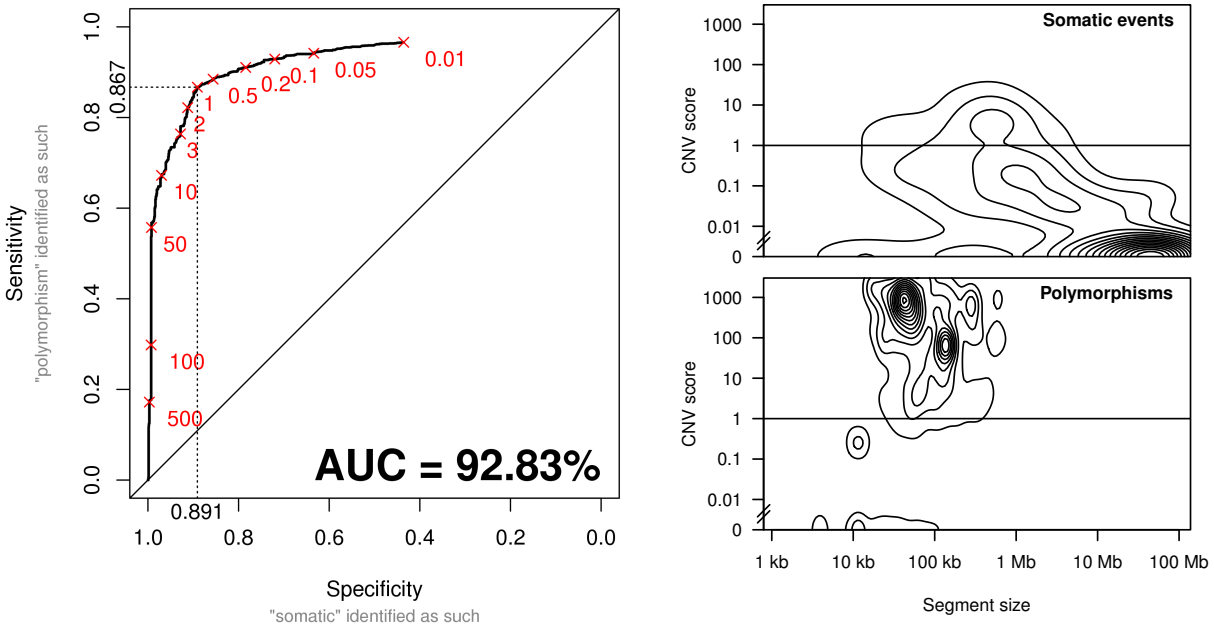


Figure 3.14 – Validation des prédictions de polymorphismes

Le panneau de gauche présente les performances de la prédiction du statut de polymorphisme en fonction du seuil de score utilisé (courbe ROC). La sensibilité et la spécificité associées au seuil retenu (1) sont mises en exergue. Le panneau de droite présente sous forme de courbes de niveaux la densité de distribution des segments d'origine somatique (partie supérieure) ou polymorphique (partie inférieure), en fonction de leur taille (abscisse) et du score qui leur est attribué par notre algorithme.

Nous avons par ailleurs constaté comme d'autres avant nous que les CNV sont limités en taille, l'essentiel de ces segments étant inférieurs à 1 Mb (Figure 3.14 droite, médiane de 62 kb). Il convient de rappeler que nous avons limité cette liste de CNV à ceux couvrant au moins 2 sondes, dans un design présentant une distance inter-sondes moyenne de 30 kb environ. Sachant que presque 90% des CNV présentaient une taille inférieure, la figure 3.14 n'est que peu informative sur la distribution des CNV en dessous du seuil de détection de notre puce. Si les segments somatiques sont généralement plus longs (médiane de 3,1 Mb), une part non négligeable de ceux-ci présentent une taille inférieure à 1 Mb, notamment les délétions du *locus* 9p21 présentant un intérêt majeur dans les LDGCB (§3.1.1).

3.5.4 Priorisation des MCR

La mise en évidence de régions amplifiées ou délétées de manière récurrente et un dernier aspect de l'analyse des puces CGH abordé lors du développement de

cghRA. Ce type d'analyse de focalise généralement sur les MCR, régions communément gagnées ou perdues par le plus grand nombre de patients (Figure 3.15), dans la mesure où le(s) gène(s) conférant à ces événements somatiques un avantage sélectif ont le plus de chances de s'y trouver. Si le nombre de patients présentant un segment chevauchant le MCR (pénétrance) est un bon indicateur de son importance, il présente néanmoins le défaut d'être parasité par des événements moins spécifiques, tels que les gains ou pertes de bras ou de chromosomes entiers. L'algorithme STEPS se concentre ainsi sur ce motif « d'escaliers », considérant qu'il témoigne d'une forte pression de sélection, tout en s'affranchissant du contexte génomique. A titre d'exemple, les régions en jaune et en vert dans la figure 3.15 présentent un score STEPS identique, alors que la région en jaune présente au niveau de son MCR une pénétrance deux fois plus élevée. Considérer les régions par ordre de pénétrance conduirait ainsi à négliger la région verte, qui présente pourtant le même niveau de réarrangements locaux suggestifs de la présence d'un gène d'intérêt.

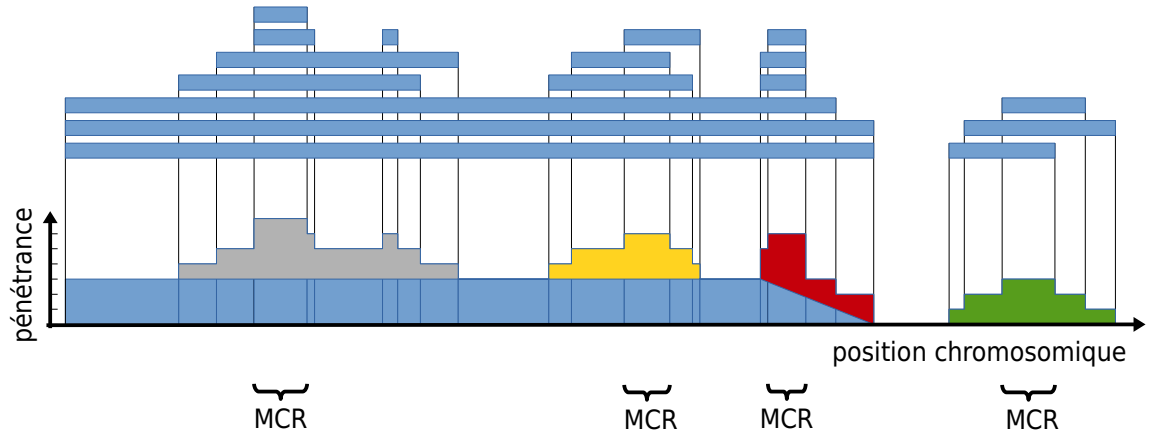


Figure 3.15 – Pénétrance et MCR

Différents segments obtenus via l'analyse d'une série de puces CGH sont présentés en vis à vis du profil de pénétrance qu'ils génèrent. Les cas particuliers surlignés en couleur sont présentés en détail dans le texte.

D'un point de vue pratique, l'algorithme STEPS consiste à calculer de manière itérative un score pour chaque coté de l'escalier formé par un MCR. En parcourant le chromosome depuis le MCR dans une des deux directions, toute diminution de la pénétrance est additionnée au score et toute augmentation soustraite deux fois, jusqu'à ce que le score atteigne 0. Ce score autorise ainsi de légères augmentations de la pénétrance dues à des segments non-spécifiques (MCR gris sur la figure 3.15), tout en s'interrompant au pied d'un MCR voisin. Les scores calculés pour chacun des deux cotés sont ensuite combinés selon l'équation 3.12, qui privilégie les escaliers symétriques, afin qu'un MCR situé à l'extrémité d'un bras fréquemment altéré ne soit pas trop avantagé (MCR rouge sur la figure 3.15).

$$mcrScore = 2 * (scoreGauche + scoreDroit) - |scoreGauche - scoreDroit| \quad (3.12)$$

La validation de cet algorithme dans la série de 108 puces CGH décrite ici est en cours, de même que sa comparaison avec des algorithmes de référence dans ce domaine tels que GISTIC [217]. L'approche de priorisation que nous avons choisi offre des listes de candidats plus fournies que ce dernier, notamment dans de petites séries comme la nôtre, mais implique d'utiliser un seuil empirique plutôt qu'un seuil α de significativité classique. Considérant les scores des événements déjà connus dans les LDGCB, nous nous sommes limités lors de l'analyse de la série GHEDI (§3.6) par cet algorithme aux MCR présentant un score d'au moins 10.

3.6 Application à la série GHEDI

3.6.1 Plan d'expérience

L'application des algorithmes de CGH décrits précédemment à la série GHEDI a permis d'offrir une nouvelle description des CNA récurrents dans les LDGCB, avec l'avantage d'un suivi clinique protocolaire. 195 patients de cette série disposaient ainsi de puces CGH Agilent 4x180k et de puces transcriptomiques Affymetrix U133+2 interprétables. Les sous-types de LDGCB ont pu être déterminés par l'application de la signature transcriptomique proposée par Wright *et al* [40], décrivant 92 GCB, 81 ABC et 22 intermédiaires, excluant les LPM. Les patients inclus dans cette série provenaient des essais cliniques LNH-03 menés par le GELA (aujourd'hui LYSA) [239–244], et avaient été traités principalement par R-CHOP (n=104) ou R-ACVBP (n=86).

L'algorithme GEDI décrit par Lenz *et al* [22] pour analyser une série similaire a été mis en œuvre afin de détecter les gains et pertes de copies impactant l'expression des gènes concernés. Brièvement, cette approche consiste à réaliser un test t de Student unilatéral pour chaque gène, entre expression des cas altérés et sauvages (2 copies). L'ensemble des p-valeurs de ces tests sont ensuite combinés en un score (somme des logarithmes), dont la significativité est mesurée par permutations.

3.6.2 Résultats et discussion

La recherche d'altérations récurrentes du nombre de copies a produit un tableau similaire à celui décrit dans d'autres séries d'envergures (§3.1.1). 56 régions minimales communément amplifiées ont pu être mises en évidence avec un score supérieur à 10, dont 28 présentaient un score GEDI jugé significatif (p=0,1). Les gains

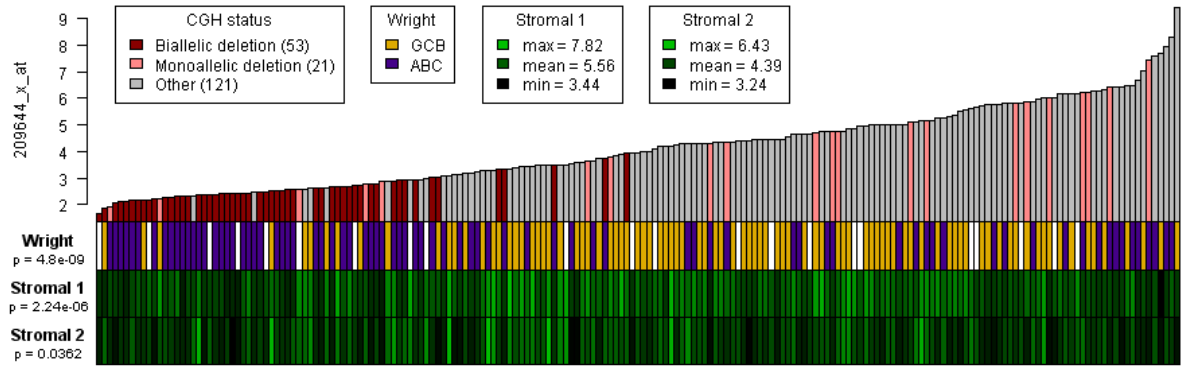


Figure 3.16 – Relation entre délétion de *CDKN2A* et expression génique

L'expression de *CDKN2A* mesurée sur puces Affymetrix est présentée en fonction de ses délétions déterminées sur puces CGH, chez 195 LDGCB de la série GHEDI. Les signatures GCB / ABC [40], Stromal 1 et 2 [41] sont également mises en relation avec ces résultats.

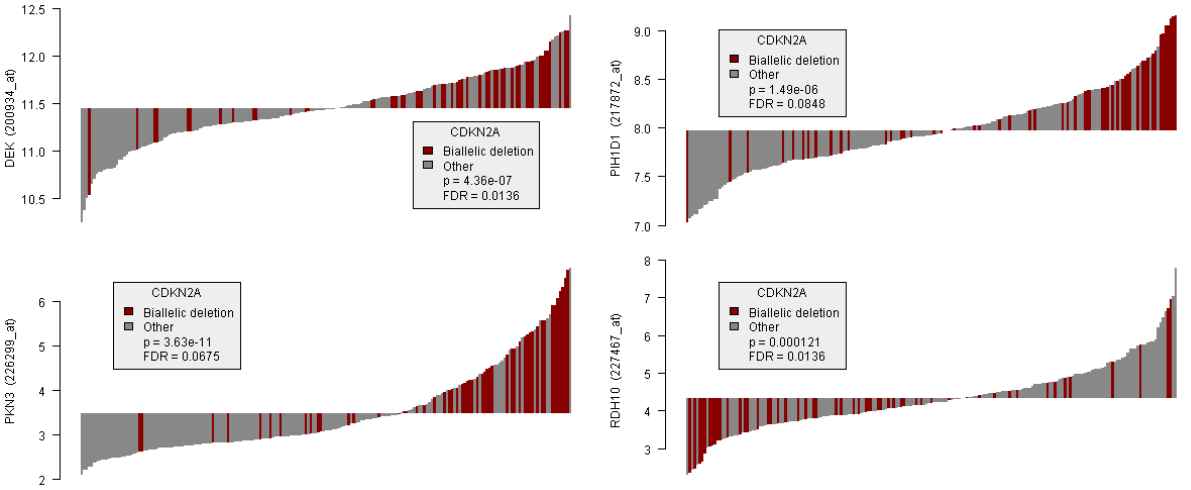


Figure 3.17 – Impact des délétions de *CDKN2A* sur l'expression de ses cibles

L'expression de 4 cibles transcriptionelles de *CDKN2A* est présentée en fonction des délétions homozygotes de ce dernier, chez 195 LDGCB de la série GHEDI.

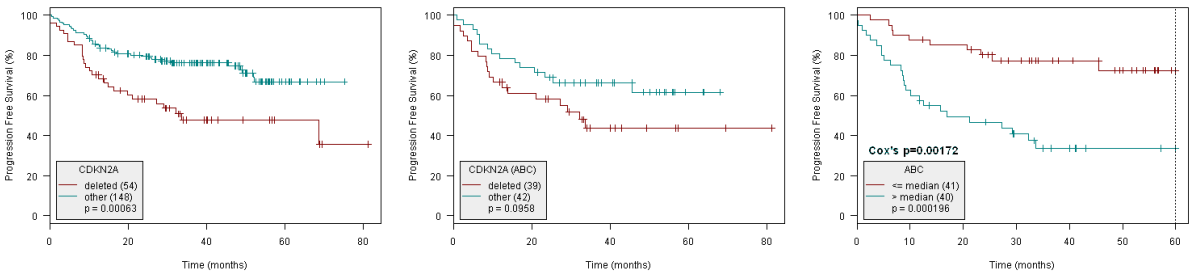


Figure 3.18 – Impact pronostique des délétions de *CDKN2A*

La survie sans progression est représentée en fonction de la délétion de *CDKN2A* (gauche et centre) ou du score d'inactivité de ce dernier (droite), dans la série GHEDI complète (gauche) ou le sous-groupe des LDGCB ABC (centre et droite).

d'*IKZF1* en 7p12 et de *REL* en 2p12 figuraient parmi les gains les plus récurrents. Plusieurs MCR montraient une association non-significative avec le pronostic, dont les gains en 3q13. 130 régions minimales communément délétées furent définies, dont 54 avec un score GEDI suffisant. La délétion de *CDKN2A* en 9p21 dominait largement cette liste avec un score de 102, et une association significative avec la survie sans progression. Des MCR centrés sur *FHIT* (3p14) et *HACE1* (6q16) furent également mis en évidence, entre autres. Des résultats similaires furent obtenus en ne considérant que les délétions homozygotes, mettant une fois de plus l'accent sur *CDKN2A*.

L'impact de la délétion de *CDKN2A* fut caractérisé plus finement, suggérant que seule la délétion homozygote de ce gène avait un impact notable sur son niveau d'expression (Figure 3.16). Cette délétion homozygote fut trouvée fortement associée au sous-type ABC ($p=5x10^{-9}$) et à une signature Stromal 1 plus faible, deux facteurs de mauvais pronostic attendus [41] (Figure 3.16).

Une signature transcriptomique de cette délétion homozygote de *CDKN2A* fut élaborée, grâce à un modèle LIMMA multivarié incluant également le sous-type (GCB ou ABC). Parmi les 17 gènes dont l'expression était significativement modifiée par la délétion de *CDKN2A* (FDR < 10%) figuraient *CDKN2A* lui même et son voisin *MTAP*, ainsi que 4 gènes d'intérêt clinique : *DEK*, *RDH10*, *PKN3* et *PIH1D1* (Figure 3.17). *DEK* est un oncogène situé en 6p22, dont la surexpression a été associée à un mauvais pronostic dans plusieurs types de cancers [245, 246]. *RDH10* (8q21) code pour une enzyme ayant démontré des propriétés anti-prolifératives sur des lignées hépatiques [247], et dont l'expression serait activée par *FOXO1*, un gène dont les mutations inactivatrices sont pronostiques dans les LDGCB [248]. *PKN3* (9q34) a été démontré comme un médiateur de la croissance tumorale liée à l'activation de la voie PI3 kinase [249]. Enfin l'expression de *PIH1D1* (19q13) a été associée à une résistance à la doxorubicine, un composant des chimiothérapies actuelles [250].

Un modèle de Cox basé sur l'expression de ces 4 gènes se révéla prédictif de la survie sans progression dans la cohorte complète ($p=8x10^{-4}$) et plus particulièrement dans le sous-type ABC ($p=2x10^{-4}$, Figure 3.18). Cette dernière observation suggère que l'effet mis en évidence va au delà du simple enrichissement de la délétion de *CDKN2A* chez les ABC, et que des inactivations de ce gène via des mécanismes alternatifs tels que la méthylation pourraient également être captés par cette approche [128, 158]. Une tendance similaire (p=0,07) fut mise en évidence en survie globale chez les LDGCB ABC publiés par Lenz *et al* [22] (faute de données de survies sans progression; Figure 3.18), ainsi que dans la cohorte complète (p=0,002).

Ces résultats ont été présentés par le Pr Fabrice Jardin lors d'une communication orale au 54e congrès annuel de la société américaine d'hématologie (ASH), ayant pris place à Atlanta en décembre 2012 [251].

4 Caractères mutationnels

4.1 État de l'art

4.1.1 Mécanistique

Avec la découverte de techniques à même de séquencer l'ADN et de réduire l'immense complexité du vivant à une succession des 4 nucléotides (A, C, G et T) est rapidement apparue la notion de **mutation**, c'est à dire de variabilité dans cette séquence. Variabilité entre individus d'une part (**polymorphismes**, ou SNP), transmise à la descendance et causée par des mutations survenues dans des cellules germinales. Mutations **somatiques** d'autre part (SNV), limitées à un groupe de cellules au sein de l'organisme et généralement étudiées dans le cadre d'un cancer. La grande flexibilité de ce système, qui à l'échelle de l'espèce permet l'évolution par l'apparition de modifications successives pouvant apporter un avantage sélectif, est ainsi à l'échelle de l'individu à l'origine des cancers, préoccupation majeure des sociétés occidentales actuelles. Nous nous limiterons dans ce chapitre aux mutations somatiques, survenant par des mécanismes dont Helleday *et al* ont publié en 2014 une revue très détaillée [252] que nous nous contenterons de survoler. Notons enfin que les translocations (§3.1.2) et CNA (§3.1.1) décrits précédemment sont parfois considérés comme des « mutations ». Nous réserverons toutefois ce terme dans le présent mémoire aux **modifications ponctuelles** de la séquence d'ADN, en y incluant les courtes insertions et délétions (moins d'une dizaine de bases ; ***indels***) qui sont fréquemment étudiées conjointement.

Helleday *et al* [252] définissent ainsi le portrait qui peut être esquissé d'une tumeur à un instant t par ses mutations comme la somme de nombreux mécanismes, chacun ayant contribué une partie des mutations présentes. Ces mécanismes ont généralement des signatures bien précises, liées aux réactions chimiques et enzymatiques sur lesquelles ils reposent, et des niveaux d'activités qui peuvent varier au cours de l'histoire d'une tumeur. Ils peuvent être **endogènes** (déméthylation des

sites CpG avec l'âge, erreurs de réplication, stress oxydant...), **exogènes** (radiations naturelles ou provoquées, produits cancérigènes...) ou une tentative de l'organisme de **réparer** un dommage plus important encore causé à l'ADN (Figure 4.1). Des processus tels que l'hypermutation somatique impliquée dans le développement des lymphocytes B (§1.4.5) sont ainsi des suspects de choix pour expliquer les mutations observées dans les lymphomes, de même que les mécanismes de réparation des cassures à l'origine des translocations (§3.1.2). Les autres membres de la famille APOBEC, à laquelle l'enzyme AICDA impliquée dans la SHM appartient, sont également fréquemment suspectées dans les cancers.

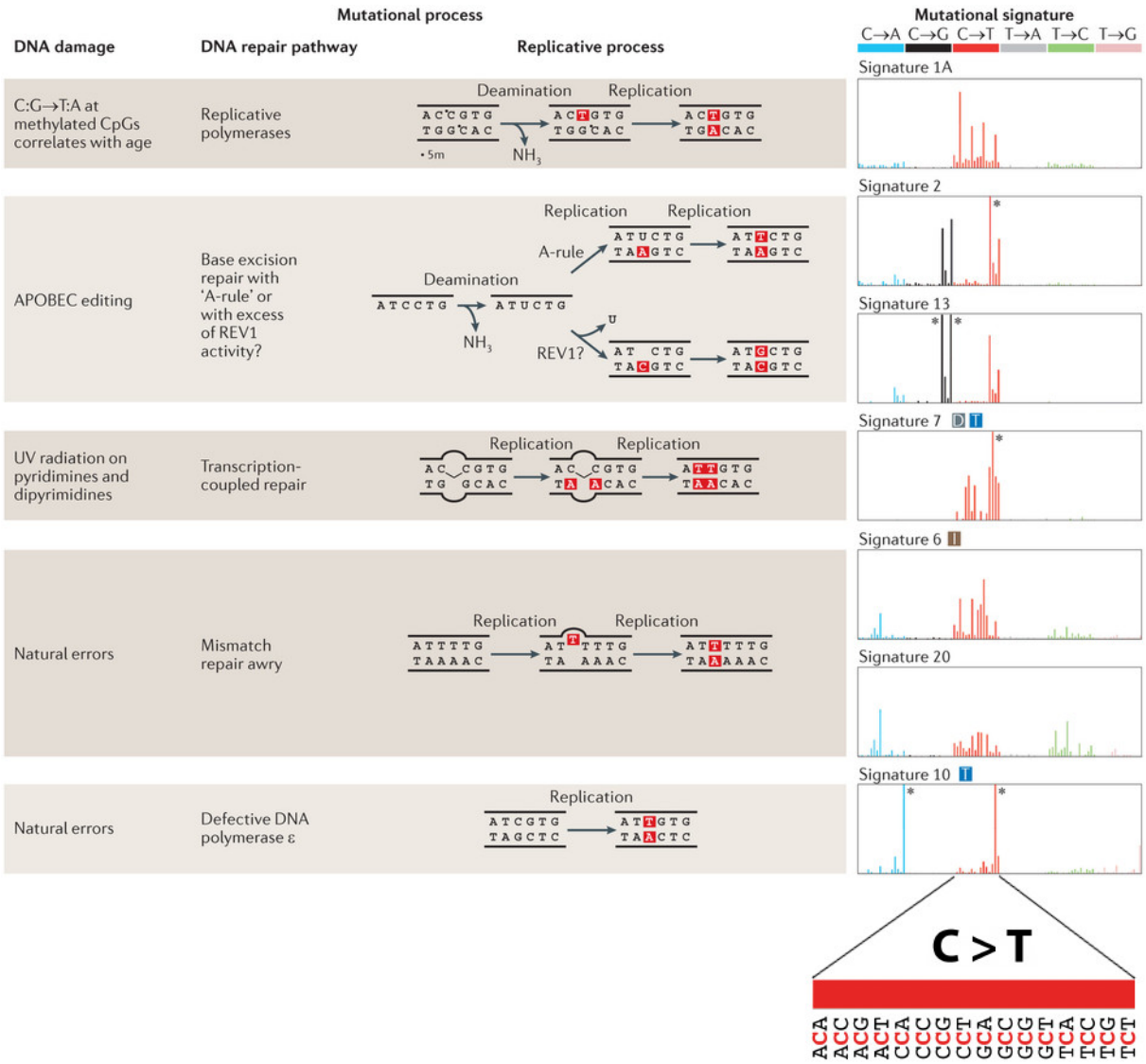


Figure 4.1 – Principales signatures mutationnelles dans les cancers

Correspondance entre 7 des 21 signatures définies par Alexandrov *et al* [253] avec des processus mutationnels connus. Les signatures marquées d'un T bleu sont préférentiellement retrouvées sur le brin transcrit, celle marquée d'un D gris impliquent fréquemment la substitution d'un dinucléotide plutôt que d'un nucléotide seul, celle marqué d'un I est également associée à des *indels*. Figure adaptée de Helleday *et al*, Nature Reviews Genetics 2014 [252].

Plusieurs équipes s'intéressèrent tout particulièrement à l'**hypermutation somatique**, cherchant à caractériser les gènes touchés par ce phénomène sélectif. Considérant que ce mécanisme était limité à un environnement bien particulier, Jiang *et al* [254] utilisèrent la ChIP-seq pour ne séquencer que ces régions proches d'un promoteur où les modifications des histones (méthylations H3K4, acétylations H3K9 et H3K27) sont favorables à l'accessibilité de la chromatine. Ils mirent ainsi en évidence dans deux lignées de LDGCB les gènes *BCL2*, *BCL6*, *PIM1* et *MYC* attendus, de même que 35 nouvelles cibles potentielles, incluant *CIITA*, *CXCR4*, *EBF1*, *MEF2B* ou *IGLL5*. Khodabakhshi *et al* [181] choisirent quant à eux une approche pan-génomique, mettant à profit les 40 génomes de LDGCB publiés précédemment par leur équipe [255] (§4.1.2). En appliquant des filtres basés sur la position (jusqu'à 2 000 bases en aval du site d'initiation de la transcription) et le contexte nucléotidique (WRCY, où le C est muté), ils identifièrent 44 gènes cibles. Là encore beaucoup de gènes déjà suspectés à l'époque furent mis en avant (*PAX5*, *IRF4*, *SOCS1*, *BCL7A*...), avec tout de même le bémol que *BCL2* et *BCL6* ne soient pas significatifs dans leurs résultats.

Alexandrov *et al* [253] cherchèrent plutôt à caractériser les mécanismes mutationnels à l'œuvre de manière non-supervisée, mettant en œuvre une méthodologie développée précédemment [256] dans plusieurs milliers d'exomes et de génomes de tumeurs publiés. Ils extrayèrent ainsi **21 signatures** de cette masse de données (Figure 4.1), chacune caractérisée par une préférence pour un ou plusieurs des 96 types de mutations possibles, considérant les bases avoisinant directement la base mutée. A titre d'exemple, la déméthylation d'un CpG correspond à une signature riche en substitutions A**CG**>A**TG**, C**CG**>C**TG**, G**CG**>G**TG** et T**CG**>T**TG**, dans la mesure où la base précédente n'a pas d'importance. L'analyse manuelle de ces signatures permit d'en relier certaines à des processus bien connus (voir pour revue [252, 253]), et d'identifier les signatures 1A et 1B d'une part (déméthylation des CpG liée l'âge), 2 et 13 d'autre part (APOBEC) comme le dénominateur commun d'une large majorité des 30 types de cancers étudiés. Si les lymphomes B étaient faiblement représentés dans cette étude (environ 50 cas incluant lymphomes de Burkitt et folliculaires), les signatures 1B, 2, 5, 9 et 17 furent mises en avant. En plus des deux mécanismes précédents, la signature 9 a été attribuée aux erreurs de la polymérase η [252] et les signatures 5 et 17 restent à identifier, la signature 5 présentant peu de spécificité et la signature 17 une préférence marquée pour la transition GCG>GTG.

4.1.2 Séquençage à haut débit des LDGCB

Études fondatrices

Si la recherche de mutations somatiques récurrentes à même d'expliquer la lymphomagénèse a longtemps été réalisée par séquençage Sanger ou *Southern blot* sur quelques gènes choisis *a priori* [257–259], les techniques de séquençage à haut débit ont permis de grandes avancées dans ce domaine. Plusieurs équipes ont ainsi entrepris plus ou moins en même temps de séquencer le génome (WGS), l'exome (WES) ou le transcriptome (RNA-seq) complet de ces tumeurs, avec cette fois-ci une approche non-supervisée.

Morin *et al* [260] publièrent ainsi pour la première fois dans Nature en 2011 une large série de lymphomes analysés avec ces 3 techniques, alors en plein essors, produisant une masse de données rarement égalée à ce jour. Le séquençage génomique de 12 tumeurs (11 LDGCB et 1 LF) additionné de 2 exomes de LDGCB supplémentaires leur permit de définir une liste de 109 gènes mutés de façon récurrente dans ces pathologies proches. Ils étendirent ensuite cette cohorte à 127 génotypes par RNA-seq, incluant au passage 10 lignées cellulaire fréquemment étudiées et 12 LF supplémentaires. Bien que cette seconde méthode ne puisse fournir une couverture aussi homogène que le WGS, elle permit néanmoins d'ajouter 10 gènes à leur liste de cibles récurrentes, parmi lesquels *PIM1*, *TP53*, *IRF4*, *B2M* ou *CD79B*. Près de 2 000 séquençages Sanger et la réalisation d'un design de reséquençage ciblé à haut débit furent nécessaires à la validation de ces mutations, mettant en évidence de nouveaux gènes tels que *GNA13*, *MLL2* (*KMT2D*) ou *MEF2B*. L'analyse statistique des mutations survenues dans ces deux derniers gènes permit de mettre en évidence des enrichissements en mutations inactivatrices et non-inactivatrices respectivement, et de placer sous les projecteurs les mécanismes de régulation épigénétique qu'ils contrôlent (méthylation et acétylation des histones). Par ailleurs la technique RNA-seq choisie par ces auteurs avait l'avantage de pouvoir mettre en évidence des transcrits de fusion (§3.1.2), issus notamment des translocations de *BCL2* (24 cas) ou de *BCL6* (17 cas) avec les gènes d'immunoglobulines. La quantification de l'expression des gènes est également possible avec cette approche, et permit d'établir la classification GCB / ABC de ces patients (§2.1.1). La préférence des mutations de *BCL2* pour l'allèle exprimé fut enfin confirmée par cette technique, en accord avec les mécanismes de SHM suspectés de les causer (§1.4.5).

Pasqualucci *et al* [261] firent à quelques mois d'intervalle les mêmes observations, avec des moyens plus modestes. Là encore un petit groupe de 6 tumeurs fut tout d'abord caractérisé par puces SNP et séquençage haut débit (WES à faible profondeur, 10x environ), afin d'établir une liste de gènes candidats. 56 gènes furent

ainsi reséquencés en Sanger dans 48 à 115 tumeurs selon les cas, mettant en évidence les mêmes gènes impliqués dans la modification des histones (*MLL2*, *CREBBP*, *EP300*, *EZH2*). 73 puces SNP supplémentaires furent réalisées, n'apportant que peu de nouveautés dans le domaine déjà longuement étudié des CNA (§3.1.1). L'accent fut tout de même mis sur l'interaction avec le micro-environnement et les lymphocytes T, par le biais de mutations inactivatrices de *B2M* (HLA classe I), *CIITA* (HLA classe II) ou *CD58* et d'amplifications de *PDL1* et *PDL2*.

Toujours à quelques mois d'intervalle, **Lohr *et al*** [262] suivirent un plan d'expérience plus en accord avec les pratiques actuelles, mettant en jeu une cohorte de découverte plus conséquente et une analyse bioinformatique plus poussée. Les exomes normaux et tumoraux de 55 patients atteints de LDGCB furent ainsi séquencés, avec cette fois-ci une profondeur de 150x susceptible de fournir une liste exhaustive des variants, y compris dans des tumeurs de faible cellularité. Les gènes rapportés ne le furent plus qu'uniquement sur des critères de récurrence, mais selon deux approches distinctes : en rapportant le nombre de mutations observées à leurs tailles d'une part, et en essayant de prédire la pathogénicité et la récurrence locale (« points chauds ») de ces mutations d'autre part. Les séquençages Sanger furent limités dans cette étude à une cinquantaine de validations, afin d'estimer la spécificité des mutations rapportées (98%). Cette approche leur permit de confirmer des gènes identifiés avant l'avènement du NGS (*CD79B*, *TP53*, *CARD11*, *MYD88*, *EZH2*), mais également les résultats plus récents rapportés par les deux études précédentes (*MLL2*, *TNFRSF14*, *BTG1*, *MEF2B*, *GNA13*). Quelques nouveautés furent mises en évidence, parmi lesquelles figurent *ACTB*, *P2RY8*, *PCLO* ou les protéines d'histone. Ces auteurs s'intéressèrent également à *BCL2*, mais s'attachèrent à démontrer à l'inverse de Morin *et al* [260] que les mutations fréquentes retrouvées dans ce gène sont « *passengers* » et non pas « *drivers* ». Ils firent en effet l'observation que les mutations non-synonymes étaient essentiellement confinées hors des domaines fonctionnels, à l'inverse des mutations synonymes. Le fort taux de mutation observé pouvait effectivement être attribué à la SHM (§1.4.5), comme l'enrichissement en mutations sur des motifs nucléotidiques WRCY pouvait l'attester, et ce par le biais de la translocation t(14;18) qui rapproche *BCL2* du *locus IGH* normalement ciblé par ce phénomène (§3.1.2). Une sélection négative des cellules perdant la fonctionnalité de *BCL2*, nécessaire à leur survie, pourrait alors expliquer ce biais de distribution.

Zhang *et al* [263] ajoutèrent rapidement une nouvelle cohorte de taille significative aux précédentes, soulignant au passage les nombreuses études ayant été menées en parallèle avec le même objectif. Une paire d'ADN normal et tumoral fut séquencée en génome complet, 73 LDGCB et 21 lignées cellulaires en exome avec une profondeur limitée (47x). Cette cohorte fut divisée en deux parties, l'une de « découverte » pour laquelle les ADN normaux appariés furent également séquencés

(n=34), l'autre servant uniquement à vérifier la prévalence des mutations observées dans l'ADN tumoral seul. Cette approche mit en exergue pour la première fois dans les LDGCB le problème des polymorphismes, des filtres analytiques étant nécessaires dans cette seconde cohorte pour palier le manque d'un ADN normal apparié. La validation fut menée par 3 techniques différentes (Raindance, puces SNP et séquençage Sanger), se contentant de valider la spécificité de l'étude (95% à 98%) sans y apporter de nouvel échantillon. La mise en évidence des gènes d'intérêt fut cette fois-ci réalisée par l'établissement d'un modèle de classification basé sur de nombreux paramètres (taux de mutation, taille du gène, mutations à fort impact prédit...), entrainé grâce à une cinquantaine de gènes déjà décrits comme pertinents dans la littérature. 322 gènes furent ainsi sélectionnés, couvrant l'essentiel des gènes utilisés pour l'entrainement du modèle. On peut noter pour la première fois l'absence de *MLL2*, ce gène ne faisant pas partie du kit de capture utilisé. Les auteurs soulignèrent le peu de recouvrement entre les gènes rapportés par les 4 études de l'époque (Figure 4.2), mis sur le compte de l'hétérogénéité biologique des LDGCB bien que d'importantes différences méthodologiques soient tout de même présentes. L'accent fut finalement mis sur la voie PI3 kinase, avec des mutations dans les gènes *PIK3CD* ou *MTOR*. Des expériences fonctionnelles pratiquées sur les lignées cellulaires mutées dans *MTOR* mirent en évidence une sensibilité accrue aux inhibiteurs de cette voie, suggérant l'importance que la recherche de mutations récurrentes peut avoir pour le développement de la « médecine personnalisée ».

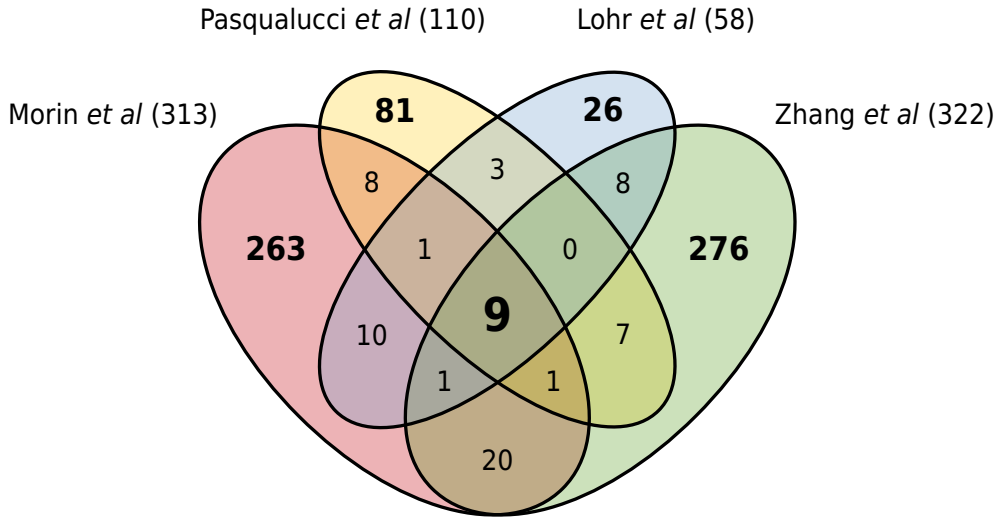


Figure 4.2 – Gènes d'intérêt dans les LDGCB pointés par 4 études NGS

Le nombre de gènes mis en évidence par Morin *et al* [260], Pasqualucci *et al* [261], Lohr *et al* [262] et Zhang *et al* [263] sont comptabilisés par un diagramme de Venn. Les comptes sont directement issus de la Figure 5 de Zhang *et al*, PNAS 2013 [263]. Le nombre total de gènes rapporté par chaque étude est rappelé entre parenthèses.

Études ultérieures

Morin *et al* [255] reprirent cette surenchère technologique un an plus tard, par le séquençage de 40 génomes complets de LDGCB additionnés de 13 lignées. Cette cohorte fut complétée à 96 patients grâce aux résultats de RNA-seq publiés précédemment [260] et à des données de puces SNP. La prévalence des mutations de *GNA13* fut précisée dans une cohorte de 279 LDGCB par reséquençage ciblé, et deux de ses partenaires (*GNAI2*, *S1PR2*) furent également retrouvés mutés dans le sous-type GCB. Si peu d'autres nouveautés furent mises en évidence du côté des mutations codantes, le WGS eut l'avantage de mettre en évidence les CNA avec une résolution de l'ordre du nucléotide, ainsi que des réarrangements survenus en dehors des régions codantes (translocations, inversions...). De nouvelles fusions impliquant *TP63* furent ainsi trouvées, conduisant à la sur-expression de cet oncogène. Des preuves de chromothripsis furent également mises en évidence, au niveau de *CDKN2A* notamment. Enfin une estimation de la chronologie de ces différents événements somatiques basée sur leur degré de clonalité fut réalisée, confirmant la survenue précoce des événements touchant *BCL2* et pointant des événements sous-clonaux pourtant jusqu'alors considérés comme *drivers* (*MYD88*, *TP53*, *CARD11*). Compte tenu de la faible profondeur de séquençage (34x) et des nombreuses faiblesses de cette approche [264], ces derniers résultats doivent cependant être relativisés.

Les mutations récurrentes étant désormais connues, les travaux ultérieurs se sont surtout focalisés sur des sous-groupes bien particuliers de patients. De Miranda *et al* [265] décrivirent ainsi dans une série de 31 exomes et 10 génomes de **patients chinois** un tableau mutationnel similaire, privilégiant la caractérisation fonctionnelle des mutations de *CD70*, *TMSB4X*, *LYN* et *DTX1*. Novak *et al* [266] ont publié tout récemment une étude portant sur 51 exomes de LDGCB traités par immunochimiothérapie comprenant 13 **patients réfractaires** (EFS < 24 mois). Ils décrivirent ainsi 374 gènes dont l'altération (CNA ou mutation) serait liée à ce phénotype, bien que l'absence de correction pour les tests multiples soit une faiblesse majeure. Des mutations dans 16 nouveaux gènes furent mises en avant, ainsi qu'une tendance à un pronostic plus sombre en cas de mutation de *FOXO1* (p=0,1) et *CIITA* (p=0,08). Des gains en 3q27 (*BCL6*), 11q23 (*CBL*) et 19q13 (*RELB*) ainsi que des délétions focalisées en 6q21 furent mises en avant comme d'autres événements plus fréquents chez ces LDGCB réfractaires. Ils mirent l'accent en 6q21 sur *SLC22A16*, malgré la présence à proximité de gènes bien connus tels que *PRDM1* ou *TNFAIP3* et de l'association établie entre délétion du bras 6q et pronostic (§3.1.1). Leurs études fonctionnelles suggèrent *SLC22A16* comme acteur de la sensibilité à la doxorubicine, composant des chimiothérapies conventionnelles, mais là encore la significativité statistique ne put être atteinte (tendance dans une lignée sur 3 testées).

Études portant sur l'évolution des tumeurs

Plusieurs équipes virent par ailleurs dans ces techniques de séquençage à haut débit une occasion de caractériser l'évolution des tumeurs, en comparant les mutations génétiques présentes à des moments différents de la maladie.

Fabbri *et al* [267] s'intéressèrent ainsi aux **syndromes de Richter** (§1.5.4), via le séquençage exomique de 9 paires de tumeurs avant (LLC) et après (LDGCB) transformation, additionné de 6 paires analysées par puces SNP. Ils observèrent dans 12 paires que la quasi-totalité des événements présents dans la LLC étaient retrouvés dans la forme transformée associée, suggérant une évolution « linéaire » de la LLC vers le LDGCB (Figure 4.3, cas 1, 2, 8 et 9). Par opposition les 3 dernières paires semblaient indiquer une évolution « branchée » où un précurseur moins muté donne naissance au LDGCB (Figure 4.3, cas 4 et 5), voire à un second clone circulant différent dans 2 cas sur 3. Aucun événement récurrent ne fut toutefois mis en évidence par cette approche parmi les mutations acquises à la transformation. L'extension à 18 exomes de Richter permis toutefois de suggérer les altérations de *TP53* et *NOTCH1* comme des événements plus fréquents que dans les LDGCB *de novo*, et les translocations de *BCL6*, délétions de *TNFAIP3* ou *PRDM1* et mutations de *CREBBP* ou *EZH2* comme moins fréquentes.

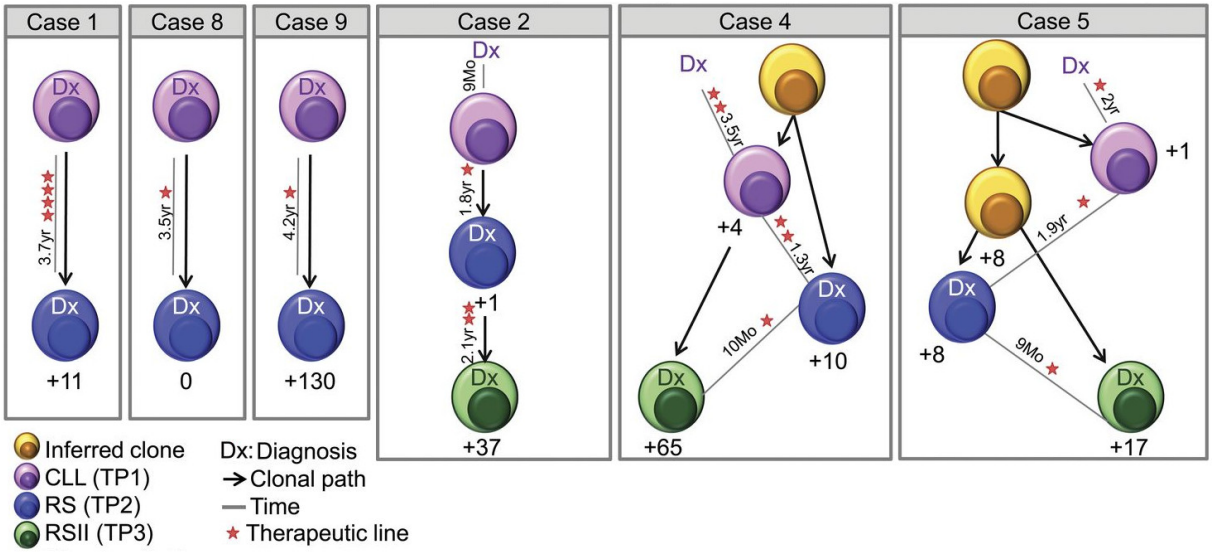


Figure 4.3 – Evolution clonale dans les syndromes de Richter

Quelques exemples d'évolution clonale inférée par séquençage exomique des tumeurs pré et post-transformation. Figure adaptée de Fabbri *et al*, JEM 2013 [267]. *CLL* = LLC, RS = *Richter syndrome*, TP = *Time Point*, yr = *years*, Mo = *months*.

Pasqualucci *et al* [268] réalisèrent un travail similaire concernant la **transformation des lymphomes folliculaires** (§1.5.4), avec le séquençage exomique de 12 paires LF / LFt. Ils établirent cette fois-ci une prédominance du modèle d'évolution

« branché » (10/12) sur le modèle « linéaire », ainsi qu'une accélération marquée de l'acquisition des mutations lors de la transformation. Étendant cette analyse à 27 exomes supplémentaires de LFt, ils firent la constatation que les mutations touchant *EZH2*, *MLL2*, *CREBBP*, *BCL2* ou *FAS* étaient des événements précoces, déjà présents dans la forme folliculaire non-transformée. Inversement la délétion de *CDKN2A* ou *TP53*, les réarrangements de *MYC* et l'emballement du SHM observé sur des gènes tels que *PIM1*, *PAX5* ou *BCL7A* survenaient plutôt après la transformation. La comparaison avec une série déjà publiée de LDGCB établit également que les événements touchants *CDKN2A*, *STAT6* ou *FAS* étaient sur-représentés dans les LFt par rapport aux LDGCB GCB génétiquement proches, expliquant en partie leur pronostic plus sombre et leur complexité génomique plus importante.

Enfin Jiang *et al* [269] mirent à profit le marqueur de clonalité naturel fourni par le réarrangement VDJ des lymphocytes B (§1.3.2) pour caractériser la parenté entre **LDGCB au diagnostic et à la rechute**. Ils appliquèrent ainsi des algorithmes de phylogénie aux régions VDJ de 14 paires de tumeurs, bénéficiant à la fois de la filiation du lymphocyte B via son CDR3 unique et d'une forte diversité évolutive liée au phénomène de SHM, toujours en activité dans ces tumeurs (§1.4.5). Ils mirent là encore en évidence plusieurs processus d'évolution : la rechute directe du clone majoritaire au diagnostic (divergence tardive, 6/14), la rechute à partir d'un précurseur moins altéré de ce clone au diagnostic (divergence précoce, 7/14) et une rechute anecdotique potentiellement liée à l'émergence d'un nouveau clone. Ces observations furent confirmées par le séquençage exomique différentiel de 7 de ces paires, mettant en évidence un nombre important de mutations gagnées à la rechute pour les divergences tardives et un équilibre entre mutations gagnées et perdues pour les divergences précoces. Parmi les altérations gagnées se trouvaient notamment des événements potentialisant les effets d'une altération présente au diagnostic, telle que l'apparition d'une *indel* dans le gène *CD58* déjà délété de manière hétérozygote ou l'amplification d'un SNV dans *B2M*. Le modèle de divergence précoce se caractérisait également par la présence au diagnostic d'un clone minoritaire portant une partie des mutations observées à la rechute, suggérant au diagnostic l'existence d'un clone proche de la tumeur mais ayant déjà commencé à diverger. Le reséquençage ciblé de quelques SNV à une profondeur de 2 000 000 dans ces tumeurs au diagnostic permit de confirmer cette observation. Finalement ils établirent que les mutations d'un groupe de gènes impliqués dans la régulation épigénétique (*MLL2*, *EP300*, *TET2*...) survenait de manière particulièrement précoce dans les LDGCB, considérant le fait qu'elles étaient présentes au diagnostic et à la rechute des divergents précoces, et donc vraisemblablement dans le précurseur dont ces tumeurs sont issues.

4.1.3 Mutations récurrentes dans les LDGCB

Vue d'ensemble

Les différentes études de séquençage à haut débit présentées précedemment ont permis de dresser un profil précis des mutations somatiques présentes dans les LDGCB, et de souligner leur grande hétérogénéité. Cette diversité dans les gènes rapportés par ces nombreuses études s'explique en partie par le taux de mutation important de ces tumeurs agressives. En se basant sur les 55 exomes publiés précédemment par leur équipe [262], Lawrence *et al* [270] placent ainsi les LDGCB au 7eme rang parmi 21 types de cancers, avec un taux médian de mutations non-synonymes par méga-bases de l'ordre de 3,3 (Figure 4.4), soit environ 125 mutations par exome. Cherchant à évaluer le degré d'avancement atteint dans la caractérisation de ces différents types de cancers, ces auteurs ont pu constater que la plupart des gènes fortement mutés ont aujourd'hui été mis en évidence, mais que le nombre de patients nécessaires à la mise en évidence des gènes moins mutés (en nombre de patients ou en nombre de mutations par patient) croit de manière exponentielle. Considérant qu'environ 300 exomes ou génomes de LDGCB ont été séquencés jusqu'à présent, il peut être dérivé de leurs calculs qu'environ 90% des gènes mutés à un taux seulement 5% supérieur à ce taux moyen sont désormais connus, mais que chaque nouvelle cohorte apportera son lot marginal de nouveaux gènes (Figure 4.4). Cette estimation est basée sur leur propre méthode d'analyse (Firehose), et il convient de considérer les différences méthodologiques (techniques de capture, profondeur de séquençage, filtre des variants...) avec prudence dans la mesure où elles peuvent énormément jouer sur le nombre de mutations rapportées.

Afin de fournir une vue d'ensemble sur les gènes mutés dans les LDGCB, les résultats des 7 études présentées précédemment ont été compilés (Figure 4.5), en se limitant aux gènes retrouvés mutés dans 3 études au moins. Les résultats d'une étude en cours de publication dans notre laboratoire [271, 272] ont également été ajoutés à titre indicatif. Dans cette dernière étude, 215 patients atteints de LDGCB ont été caractérisés par reséquençage ciblé de 34 gènes d'intérêt hématologique, dans le but de préciser leur taux de mutation dans une série conséquente de patients protocolaires. La séparation en sous-types GCB et ABC a été faite dans la mesure du possible, ces différents sous-types présentant des profils mutationnels distincts. Les LPM et LF présents dans ces études sont également rapportés compte tenu de leur fort lien de parenté avec les LDGCB (§1.6).

Le sous-type **ABC** se caractérise ainsi par des mutations fréquentes de *MYD88* (20% à 30% des cas), l'un des rares gènes à présenter un *locus* fortement récurrent (L265P). *CD79B* et *TNFAIP3* sont également des cibles récurrentes, soulignant

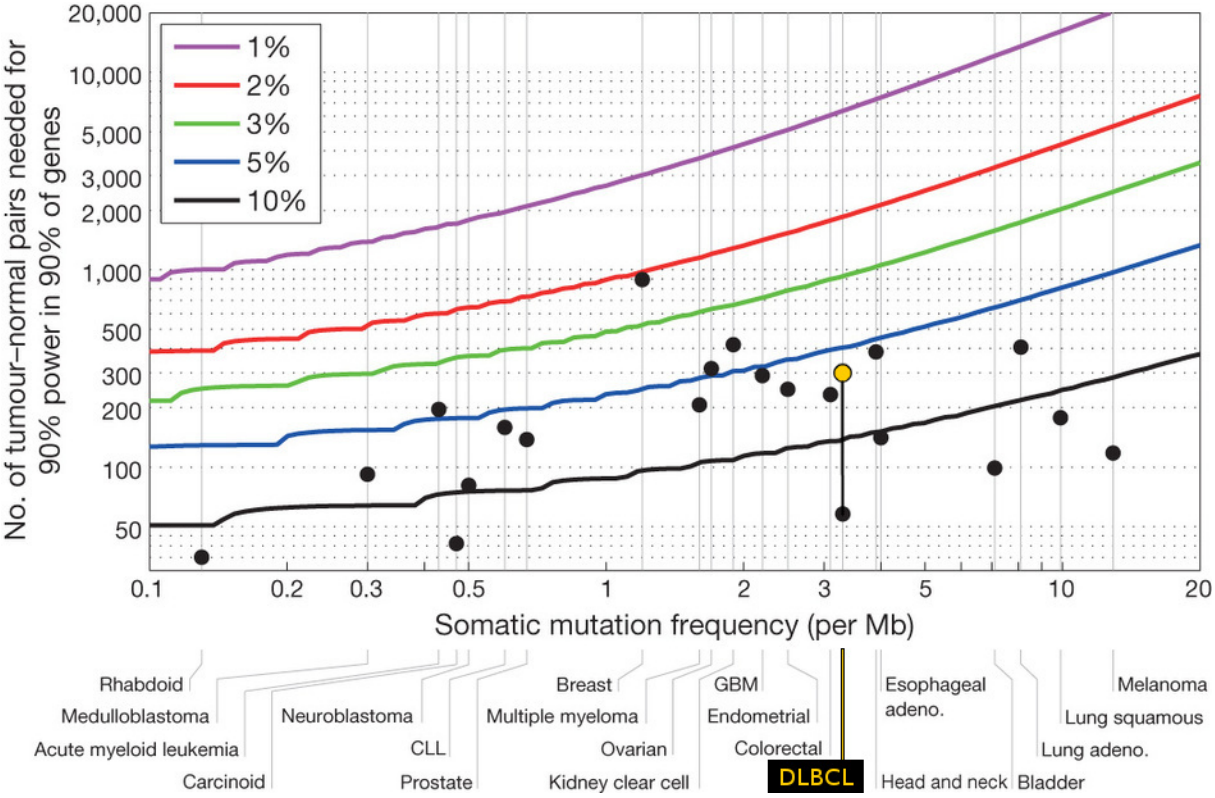


Figure 4.4 – Étude de saturation pour 21 types de cancers

L'évolution du nombre d'exomes à séquencer pour détecter 90% des gènes significativement mutés avec une puissance d'au moins 90% est présentée en fonction du taux de mutation moyen de 21 types de cancers. Les différentes courbes représentent des niveaux de mutation croissants pour les gènes à détecter (gènes mutés à un taux 1%, 2%, 3%, 5% ou 10% plus élevé que le taux moyen). Chaque point noir représente l'état des lieux dressé par les auteurs en 2014, les LDGCB étant mis à jour et en évidence en jaune. Figure adaptée de Lawrence *et al*, Nature 2014 [270].

encore une fois l'importance des voies du BCR et NFκB dans ce sous-type. L'inactivation du facteur de différenciation *PRDM1* est également un résultat attendu (§1.4.7). Si les mutations de *PIM1* sont suspectées d'être causées par AICDA (1.4.5), elles sont toutefois plus fréquentes dans le sous-type ABC. On retrouve en revanche dans le sous-type **GCB** un grand nombre de gènes impliqués dans les modifications des histones, tels qu'*EZH2*, *GNA13*, *MEF2B* ou *CREBBP*. En tant que cible de la SHM, on retrouve préférentiellement *BCL2* dans ce sous-type. *B2M* et *TNFRSF14* viennent enfin compléter ce tableau. Si moins de recul est disponible concernant les **LPM**, il semble que ces lymphomes présentent un profil mutationnel très particulier, plus proche des lymphomes de Hodgkin [283] que des LDGCB. On retrouve ainsi dans ce sous-type des mutations de *SOCS1*, *TNFAIP3*, *B2M* ou *GNA13*, en plus de *STAT6* qui semble plus spécifique aux LPM qu'aux lymphomes de Hodgkin [284]. Des mutations non spécifiques sont également observées, dans des gènes tels que *TP53*, *CARD11*, *BTG1*, *CD70*, *EP300*, *MLL2* et les monomères d'histone.

Groupe Référence n	LF A n=13	GCB A n=54	 B *	 C n=17	 D n=83	LDGCB E n=49	 F n=31	 G n=40	 H n=51	ABC A n=32	 B *	 C n=25	 D n=81	LPM D n=18	Pronostic [Réf.]	Trairements expérimentaux (ClinicalTrials.gov)	Loci préférentiellement mutés
B2M	8%	13%	7%*	6%	18%	10%	23%	8%	12%	6%	5%*	4%	9%	50%			
CARD11	15%	13%	4%*	18%	7%	20%	3%	20%	10%	6%	6%*	12%	14%				Domaine coiled-coil
CD79B	0%	2%	0%*	0%	2%	16%	10%	8%	14%	19%	12%*	16%	25%		= [I]	Everolimus (NCT01854606)	Domaine ITAM, Y196
CREBBP	46%	22%	11%*	6%	31%	16%	10%	8%	16%	13%	7%*	4%	6%	11%		Mocetinostat (NCT02282358)	R1446, Y1503, S1680–
MYD88	0%	6%	0%*	0%	10%	12%	19%	18%	12%	31%	8%*	24%	28%		– [J]	IMO-84800 (NCT02252146)	L265P
TP53	8%	33%	8%*	12%	16%	24%	16%	10%	24%	3%	9%*	0%	19%	11%	– [K]		R248
BCL2	69%	52%	8%*	12%	24%		3%	20%	20%	13%	1%*	4%	1%		= [L]		Hors domaines BH
BTG1	0%	13%		12%		16%	13%	8%	16%	19%		16%					
CD58	0%	6%	3%*		10%	10%	3%	8%	6%	0%	3%*		6%	39%			
EZH2	23%	33%	5%*	6%	18%	14%	3%	23%		0%	1%*	0%		6%	= [M]	EPZ-6438 (NCT01897571) GSK2816126 (NCT02082977)	Y646, A682, A692
GNA13	69%	37%		24%	12%	20%	3%	15%	14%	3%		0%	9%	50%	= [G]		
HIST1H1C	8%	9%		0%		14%	10%	10%	6%	9%		8%					
MEF2B	31%	26%	5%*	12%	23%	18%		13%	14%	3%	3%*	0%	12%	11%			Domaines fonctionnels
MLL2 / KMT2D	69%	43%	12%*		46%	29%	6%	23%	33%	28%	11%*		41%	17%			
PIM1			4%*	12%	8%	31%	23%	18%	12%		11%*	24%	33%			PIM inhibitors [N]	Motifs WRCY
CD70	0%	6%				10%	23%	8%	10%	6%							
PRDM1			2%*	0%	6%		6%	8%	0%		10%*	0%	16%				
SGK1	8%	28%		6%		10%	10%	10%		0%		0%					
TNFAIP3	8%	2%	4%*	12%	11%		10%		4%	6%	12%*	4%	15%	61%	= [O]		
TNFRSF14	31%	15%			17%	22%	6%	10%	14%	0%			2%		– [P]		
CCND3	8%	9%	2%*			6%			6%	9%	4%*						
HIST1H1E				12%		14%	13%	13%				4%					
KLHL6	8%	13%				8%	3%	13%		6%							
PCLO						35%	10%	15%	18%								
STAT3	0%	11%				10%	10%		4%	6%							Domaine SH2
TMEM30A	8%	0%	2%*	6%				8%		6%	4%*	4%					
BCL6				6%			10%	8%				0%					
CIITA					10%	10%	10%		8%				12%	56%			
DUSP2			2%*			6%		13%			4%*						
EP300	23%	9%	1%*		14%				2%	3%	5%*		15%	17%		Mocetinostat (NCT02282358)	
IRF8	15%	15%		12%				13%		9%		0%					
KLHL14				0%			3%	5%				24%					
LRP1				12%			13%	13%				4%					
LRP1B				24%			13%	13%				12%					
LRRN3				6%			13%	8%				0%					
MYC			5%*	0%	10%				8%		1%*	4%	5%	6%			
P2RY8						12%		13%	10%								
SOCS1				12%	16%	6%		13%				0%	6%	56%			
TBL1XR1						6%	10%	8%									
TNF			4%*			6%		10%			12%*						
ACTB						10%	10%										
BCL10								8%									
BCL2L10			0%*								2%*						
BCL7A				12%								0%					
FOXO1	23%	11%			12%				4%	0%			4%	6%	– [Q]		Domaine *Forkhead box*
IKZF3								5%									
IRF4					5%			10%					14%	11%			
KMT2C				12%			6%					12%					
LYN							10%										
MTOR				0%								4%					
NFKBIA						8%											
NOTCH1			1%*		1%						1%*		7%	6%	– [R]	RO4929097 (NCT01158274)	L2482
NOTCH2				6%	10%							0%	2%		– [R]	RO4929097 (NCT01158274)	R2400
PIK3CD				18%								0%					
STAT6					14%	4%								72%			N417
SYK						4%											

Figure 4.5 – Bilan des SNV répertoriés dans les LDGCB

Les fréquences de mutations rapportées par 8 études de séquençage à haut débit sont présentées pour 56 gènes d'intérêt, par sous-type de LDGCB ou de lymphomes apparentés lorsque l'information était donnée. Ces gènes ont été rapportés par 3 études au moins (40 premières lignes, jusqu'à *TNF* inclus), ou sélectionnés manuellement pour leur intérêt particulier. Le nombre de cas étudié dans le sous-groupe considéré est rapporté pour chaque étude. Pour la référence B (*), les pourcentages sont rapportés au nombre de cas dans la cohorte complète, et non pas dans le sous-groupe comme pour les autres références. Les gènes préférentiellement mutés dans un sous-type sont rapportés en orange, bleu ou vert selon le sous-type. Les considérations pronostiques et *loci* récurrents sont tirés de Dubois *et al* [273]. Références : **A** = Morin *et al* 2011 [260], **B** = Pasqualucci *et al* 2011 [261], **C** = Zhang *et al* 2013 [263], **D** = Dubois *et al* [271, 272], **E** = Lohr *et al* 2012 [262], **F** = de Miranda *et al* 2014 [265], **G** = Morin *et al* 2013 [255], **H** = Novak *et al* 2015 [266], **I** [274], **J** [275], **K** [276], **L** [277], **M** [278], **N** [279], **O** [280], **P** [281], **Q** [248], **R** [282].

Cinq gènes parmi les plus représentatifs des LDGCB (*MYD88*, *CARD11*, *EZH2*, *CD79A* et *CD79B*) ont fait l'objet de séquençages Sanger chez 161 LDGCB au sein de notre laboratoire [130], et d'une revue publiée en 2015 (Annexe A).

Mutations courantes

Les mutations de ***BCL2*** constituent un événement fréquent dans le sous-type GCB (15% à 25%, Figure 4.5), une observation qui peut sembler paradoxale quand on connait l'importance que cet oncogène peut avoir dans la survie des lymphocytes B (§3.1.2). Une analyse plus approfondie de ces mutations révèle cependant qu'elles n'ont généralement pas d'impact, les mutations non-synonymes étant confinées hors des domaines fonctionnels BH [262, 277]. Considérant que le fort taux de mutation pourrait être expliqué par la SHM, particulièrement active dans *BCL2* chez les LDGCB GCB du fait de la t(14;18) également fréquente dans ce sous-type [262], ces mutations ne devraient pas mettre en péril l'action de ce gène, et donc des éventuels traitements ciblés [273, 285]. Du coté ABC, la SHM semble privilégier l'anti-apoptotique ***PIM1*** (15% à 25%, Figure 4.5). Ce phénomène surprend également et reste à élucider compte tenu du fort niveau d'expression de cette protéine, particulièrement dans ces LDGCB ABC où *BCL6* est fréquemment transloqué et sur-exprimé [286].

Les mutations d'***EZH2*** sont quasiment exclusives des LDGCB GCB (environ 20%, Figure 4.5), et touchent tout particulièrement la position protéique Y646 (anciennement Y641, Annexe A). Si la protéine EZH2 est capable de catalyser l'ajout de un à trois groupes méthyls sur la lysine 27 des protéines histones H3, l'introduction de cette mutation induit un glissement dans ses préférences de la mono-méthylation vers la tri-méthylation [287]. Cela se traduit par un niveau de tri-méthylation plus élevé [131], et donc une répression transcriptomique liée à la compaction de la chromatine dans de nombreuses régions du génome. La prévalence de cette mutation dans les LF et LDGCB GCB a conduit au développement d'inhibiteurs spécifiques, qui montrent des résultats prometteurs y compris chez les patients non-mutés [273].

KMT2D (anciennement *MLL2*) est une autre histone méthyl-transférase, ciblant cette fois-ci la lysine 4 (H3K4) pour activer la transcription. Elle est retrouvée mutée dans 20% à 40% des LDGCB, indistinctement du sous-type. Si sa taille importante (5 500 acides aminés) peut de prime abord laisser douter de leur spécificité, l'impact des nombreuses mutations tronquantes et des mutations non-synonymes en C_{ter} a été démontré sur les populations cellulaires du centre germinatif [288, 289]. L'observation de ces mutations dans les lymphomes folliculaires [260] et leur théorisation dans des stades particulièrement précoces des LDGCB [269] semblent de plus indiquer que ces mutations sont parmi les premiers événements de la lymphomagénèse, facilitant son établissement par l'altération de la régulation épigénétique.

Toujours dans la famille des régulateurs épigénétiques, les paralogues ***CREBBP*** et ***EP300*** sont également la cible de mutations récurrentes dans les LDGCB, préférentiellement dans le sous-type GCB (30% et 10-15% respectivement). Ces acétylases ciblent entre autres les histones, conduisant à réguler en interaction avec d'autres partenaires l'expression des gènes à proximité. Les mutations observées dans ces deux gènes sont principalement inactivatrices, et peuvent conduire à un défaut d'inactivation de *BCL6* et d'activation de *TP53* propices à la lymphomagénèse [290].

La mutation L265P de la protéine ***MYD88*** et les mutations des domaines ITAM de ***CD79A*** et ***CD79B*** sont des acteurs emblématiques de l'activation constitutive de la voie NFκB dans le sous-type ABC, avec des fréquences d'environ 30% et 20% chacun (Annexe A). La protéine MYD88 gagne par cette mutation la faculté d'assembler ses effecteurs IRAK1 et IRAK4 en l'absence de stimulation (via les TLR en temps normal), et donc d'activer plusieurs voies métaboliques en aval, dont NFκB [42]. Les protéines CD79 sont quant à elles couplées au BCR (Figure 1.1D), et la désactivation ou le remplacement des tyrosines « clé » des motifs ITAM conduiraient au maintien de l'activation « chronique » du BCR. Si de nombreuses tyrosines kinases peuvent transmettre le signal depuis CD79, cette mutation hétérozygote empêche l'action de la tyrosine kinase *LYN*, qui contrairement aux autres est impliquée dans une boucle rétroactive atténuant l'activation du BCR. En privilégiant les autres kinases, l'activation (même légère) du BCR permet le recrutement de SYK et l'activation de la voie NFκB [42] via une cascade de transduction du signal (*BTK*, PKCβ, *CARD11*...) [291]. Ces observations ont conduit au développement d'inhibiteurs de *BTK*, afin de bloquer en aval du BCR l'activation de NFκB, une molécule qui semble également donner des résultats prometteurs chez les patients mutés *MYD88* [292].

La mutation de ***CARD11*** fournit un moyen plus efficace de garder les voies du BCR et NFκB sous activation constante, y compris en l'absence de composants essentiels du BCR [291]. On pourra s'étonner cependant que ces mutations soient retrouvées dans les deux sous-types (Annexe A), malgré les études réservant l'activation de NFκB au sous-type ABC [45]. Compte tenu du faible taux de mutation de *CARD11* et des nombreuses altérations n'impliquant pas NFκB chez les GCB, une explication possible serait que cette voie ne soit pas impliquée aussi systématiquement que chez les ABC, d'où un phénomène de « dilution » dans les études comparant les deux sous-types.

La voie NFκB semble également un acteur important dans la pathogénicité des LPM, mais son activation tiendrait plutôt de l'inactivation de son inhibiteur ***TNFAIP3***. Si sa position en 6q en fait un gène fréquemment délété dans les lymhomes en général (§3.1.1), il présente également un taux inhabituellement élevé de mutations tronquantes dans les LPM (60%).

Ces LPM se singularisent également par de nombreuses mutations centrées sur la voie de signalisation JAK-STAT, avec de fréquentes (>50%) mutations de *STAT6*, *PTPN1* et *SOCS1*. **STAT6** est un facteur de transcription, qui en réponse à l'IL4 induit l'expression de nombreux gènes, dont l'anti-apoptotique *BCL2L1*, mais semble également inhiber l'expression de *BCL6* [293]. Les mutations observées dans ce gène sont rarement tronquantes, mais touchent essentiellement son domaine de liaison à l'ADN, privilégiant l'hypothèse d'un changement d'affinité pour ses cibles [284]. La protéine **PTPN1** est quant à elle une tyrosine phosphatase, qui convertit STAT6 de sa forme activée (phosphorylée) à sa forme inactive [294]. Les nombreuses mutations tronquantes observées dans ce gène expliquent ainsi la forte activation de STAT6 dans les LPM. **SOCS1** enfin serait impliqué dans la dégradation de la protéine JAK2, dont l'importance dans les LPM est déjà soulignée par les fréquentes amplifications de son *locus* en 9p (§3.1.1). Là encore des mutations tronquantes sont fréquentes, empêchent SOCS1 de jouer son rôle et permettent donc une activité prolongée de JAK2 sur ses cibles [295].

Enfin comme dans de nombreux cancers, le gène ***TP53*** est fréquemment retrouvé muté dans les LDGCB, sans préférence de sous-type. Le séquençage de 506 LDGCB a permis d'estimer la proportion de patients mutés à 22% des LDGCB, et de démontrer une nette prédominance des mutations non-synonymes (81%). En cohérence avec son rôle de suppresseur de tumeur bien connu, la majorité de ces mutations non-synonymes démontrent un impact sur la fonctionnalité de ce gène, de même que les 10% de mutations faux-sens. L'analyse de survie réalisée sur des patients traités par R-CHOP démontre un pronostic défavorable pour ces mutations, et ce dans les deux sous-types GCB et ABC [276].

4.2 Objectifs et approche suivie

Les séquences publiées de génomes, exomes ou transcriptomes ont permis de dresser un portrait détaillé des mutations somatiques dans les LDGCB. Si les estimations de Lawrence *et al* [270] suggèrent que des mutations moins fréquentes restent encore à découvrir, les projets en cours de consortiums tels que *The Cancer Genome Atlas* (TCGA) [296] ou l'*International Cancer Genome Consortium* (ICGC) [297] sauront rapidement combler ce manque avec des moyens supérieurs aux nôtres. Afin de contribuer à cette caractérisation mutationnelle de manière efficace, nous avons donc choisi de concentrer nos efforts sur des groupes bien particuliers de LDGCB, pour lesquels peu de données sont disponibles.

Fidèles à l'intérêt que porte notre unité à la clinique de ces lymphomes, nous avons dans un premier temps concentré nos efforts de séquençage sur des patients réfractaires, dans l'optique d'identifier des anomalies génétiques qui pourraient expliquer cette résistance au traitement. Nous avons ainsi choisi 14 patients inclus dans les essais cliniques du GELA (devenu aujourd'hui LYSA), pour lesquels un suivi protocolaire avait permis de mettre en évidence une rechute dans l'année suivant le diagnostic. Les exomes normaux (cellules nucléées du sang périphérique) et tumoraux (biopsie ganglionnaire) de ces 14 patients ont été séquencés, afin de révéler les mutations somatiques et CNA récurrents dans ces tumeurs (§4.4).

Le laboratoire ne disposant pas à mon arrivée des outils et compétences nécessaires à ce projet, nous avons fait le choix d'en sous-traiter la partie « humide » à une société privée (Integragen), et de développer localement notre propre pipeline d'analyse. L'élaboration de ce pipeline a constitué une part significative de mon travail de thèse, qui ouvre la voie au séquençage d'autres LDGCB particuliers, tels que les LDGCB cutanés primitifs de la jambe (§6.3).

4.3 Méthodes

4.3.1 Séquençage exomique

Principe

Dans la mesure où l'interprétation de mutations non-codantes, situées dans les introns ou les régions inter-géniques, reste encore à l'heure actuelle une énigme, le séquençage du seul exome est un substitut acceptable au séquençage du génome complet qui permet de maitriser les coûts. En effet, l'exome humain représente avec ses 38 millions de paires de bases (Ensembl Genes 79 [298]) un peu plus de 1% des 3,23 milliards que compte le génome entier [299]. Filtrer *a priori* l'ADN cible permet donc de séquencer simultanément plus de patients, et/ou à une profondeur plus élevée, sans augmenter le nombre de *runs* à réaliser.

Limiter l'ADN séquencé a cependant ses désavantages, notamment en ce qui concerne la détection de réarrangements chromosomiques plutôt que de mutations ponctuelles. En effet, de nombreux points de cassure ont lieu en dehors des exons (dans le cadre de translocations notamment, §3.1.2), et ne pourront donc pas être mis en évidence par cette méthode. De plus, les techniques de capture utilisées ne présentent pas des efficacités identiques pour chaque exon, ce qui conduit à une couverture beaucoup plus hétérogène que lors d'un séquençage basé sur la fragmentation d'un ADN génomique complet. Ce dernier point peut poser problèmes aux

algorithmes de détection de CNA basés sur la profondeur de couverture (§4.4.1), mais également prévenir la détection de mutations dans une part modeste de l'exome (généralement moins de 5%), faute de couverture suffisante [300].

Capture de l'exome

D'un point de vue pratique, de nombreuses solutions commerciales sont actuellement disponibles pour extraire un ensemble de régions d'intérêt d'un échantillon d'ADN génomique [301]. En ce qui concerne l'exome complet, les techniques « hybrides » basées sur puces [302] sont progressivement abandonnées au profit d'un enrichissement « en solution » [303], qui permet d'obtenir des résultats comparables en limitant les appareillages spécifiques et la quantité d'ADN nécessaires [304]. Le principe reste néanmoins inchangé, l'ADN génomique étant fragmenté puis filtré par hybridation à des séquences complémentaires synthétisées au préalable à partir de banques de données d'exons. Nous l'illustrerons à travers la technologie SureSelect™, développée par *Agilent Technologies*, utilisée dans les travaux relatés dans ce chapitre.

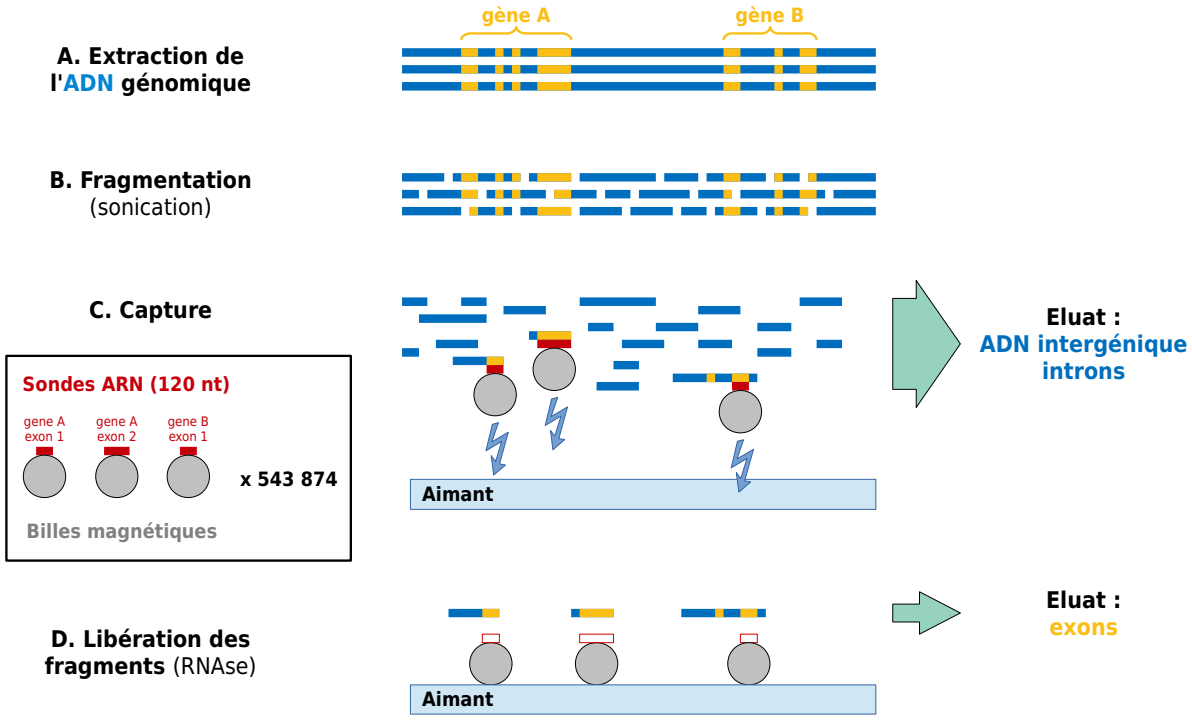


Figure 4.6 – Principe de la capture exomique SureSelect™

Les principales étapes (A à D) de la capture de l'ADN exomique par la technologie *SureSelect™ Human All Exons V5* commercialisée par *Agilent Technologies* sont présentées ici.

L'ADN génomique est ainsi extrait des cellules à séquencer selon des protocoles classiques en biologie moléculaire (Figure 4.6A), qui peuvent varier en fonction

du matériel biologique étudié (biopsie fraîche, congelée, incluse en paraffine voire échantillon de liquide biologique). L'ADN est ensuite fragmenté par sonication (Figure 4.6B), une technique physique qui permet d'induire des cassures en des points différents d'une copie du génome à une autre. Comme nous le verrons par la suite, seules les extrémités de ces fragments d'environ 1 000 paires de bases sont effectivement séquencées, il est donc primordial d'avoir une répartition aussi homogène que possible de ces points de cassure pour obtenir une séquence de l'ensemble de la région cible. La phase de capture proprement dite (Figure 4.6C) est réalisée dans le cadre du protocole *SureSelect*™ au moyen de billes magnétiques, sur lesquelles ont été fixées des sondes ARN d'environ 120 nucléotides. Chaque exon identifié dans une banque de données telles que RefSeq ou Ensembl est ainsi ciblé par un type de bille, qui présente à sa surface une sonde ARN correspondant à sa séquence complémentaire. Les fragments d'ADN génomique contenant au moins une partie d'un exon seront ainsi retenus par une bille, qui pourra être séparée des fragments en suspension à l'aide d'un aimant. Il est important de souligner que la spécificité d'un kit de capture d'exome dépend ainsi des connaissances contemporaines à sa fabrication, et que ces kits doivent être fréquemment mis à jour à mesure que ces banques de référence évoluent. Une fois les fragments exempts d'ADN exomique élués, les fragments piégés par les billes peuvent être libérés par l'action d'une RNAse (Figure 4.6D), qui dégrade les sondes ARN tout en laissant les fragments d'ADN intacts. La solution finale ne contient donc en théorie que des fragments contenant de l'ADN exomique, étendu d'un nombre de nucléotides introniques plus ou moins élevé selon la taille des fragments choisie.

Chimie de séquençage Illumina / Solexa

Une fois l'ADN cible extrait et éventuellement filtré, le séquençage proprement dit peut commencer. Si plusieurs technologies de séquençage à haut débit se livrent actuellement concurrence [301], nous nous limiterons ici à celle développée par Solexa et commercialisée par Illumina qui fut utilisée dans ces travaux [305, 306]. Comme pour la plupart de ces technologies, la première étape consiste à fixer aux extrémités de chaque fragment d'ADN une paire d'amorces commune, qui permettra le séquençage simultané de ce qu'il convient d'appeler désormais une « librairie ».

Selon un principe proche du séquençage Sanger tel qu'il est pratiqué aujourd'hui, cette technologie repose sur la détection d'une succession de signaux fluorescents, chacun des 4 nucléotides étant associé à un fluorochrome différent. Contrairement au Sanger en revanche, chaque fragment d'ADN sera la cible d'un séquençage indépendant et parallèle, et devra donc être amplifié de manière clonale pour générer un signal fluorescent d'intensité suffisante. La technologie Illumina repose ainsi sur une

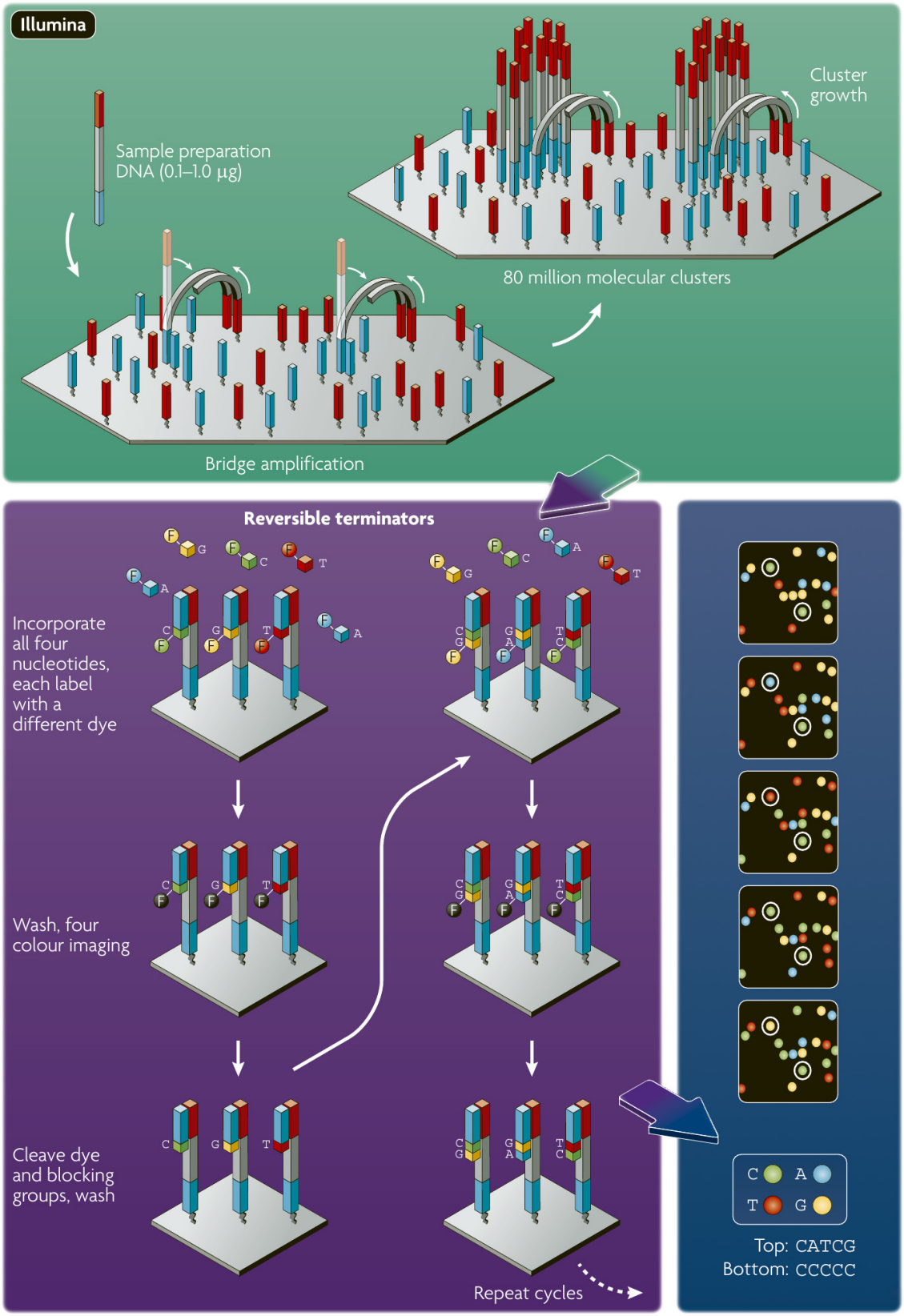


Figure 4.7 – Principe du séquençage Illumina

Les principales étapes du séquençage ADN selon la technologie développée par Illumina sont présentées ici : amplification clonale par ponts (vert), séquençage en parallèle grâce à des terminateurs réversibles (violet) et analyse conjointe des fluorescences émises (bleu). Figure adaptée de Metzker 2008 [301, 307].

première phase d'amplification en ponts (Figure 4.7, panneau vert), qui permet de copier chaque fragment d'ADN tout en gardant ses copies géographiquement proches les unes des autres. Cette amplification est réalisée sur une phase solide, la *flow cell*, sur laquelle sont fixées de manière covalente les séquences complémentaires des amorces de séquençage. Un fragment hybridé à l'une de ces amorces pourra alors former un « pont » en hybridant sa deuxième extrémité à une amorce proche, et donc générer une copie à partir de cette deuxième amorce, par PCR. Chaque fragment issu de la librairie génère ainsi un *cluster* de copies à l'endroit où il est entré en contact avec la *flow cell*, un contrôle fin de la stœchiométrie permettant d'éviter que les *clusters* ne se chevauchent.

La lecture simultanée des séquences de chacun de ces *clusters* se déroule ensuite selon un principe appelé « séquençage par synthèse » (Figure 4.7, panneau violet). Les oligonucléotides formant chaque *cluster* sont réduits à l'état simple brin, et l'une des deux amorces de séquençage est hybridée. La *flow cell* est ensuite mise en présence des 4 nucléotides, et l'élongation de l'amorce de séquençage par l'une de ces bases est détectée indépendamment pour chaque *cluster*. À chaque cycle, un unique nucléotide pourra s'hybrider à la suite de l'amorce, dans la mesure où une modification chimique de ces nucléotides en solution prévient son élongation en 3' [301]. Ces nucléotides embarquant également un fluorochrome, un lavage et la lecture des 4 fluorescences sur l'ensemble de la *flow cell* permet d'identifier le nucléotide incorporé dans chaque *cluster* (Figure 4.7, panneau bleu). Le fluorochrome et le groupe chimique qui préviennent l'élongation sont ensuite clivés, et un nouveau cycle peut commencer. En comparant la position des signaux fluorescents enregistrés à chaque cycle sur la *flow cell*, la séquence de chaque *cluster* (et donc du fragment d'ADN l'ayant formé) peut ainsi être retrouvée (Figure 4.7, panneau bleu). Chaque cycle nécessite plusieurs étapes et conduit nécessairement à l'incorporation d'un unique nucléotide, ce qui explique que les lectures générées par cette technologie soient de taille fixe et relativement modestes.

Protocole de séquençage *paired-ends*

Indépendamment des méthodes de capture et de séquençage proprement dites, il existe plusieurs façons d'aborder le séquençage de ces fragments. On distingue ainsi le séquençage *single-end* du séquençage *paired-ends*, selon que les fragments d'ADN soient séquencés dans un seul sens ou dans les deux sens, en deux étapes distinctes. Le sens en lui-même n'a que peu d'importance, la plupart des techniques de capture permettant de réaliser le séquençage *single-end* à la fois depuis les extrémités 5' et 3', chaque fragment de chaque génome étant séquencé soit dans un sens soit dans l'autre. La distinction tient essentiellement du fait que dans le cas du séquençage

paired-ends, deux lectures sont réalisées à partir du même fragment d'ADN.

Dans la mesure où les séquenceurs à haut débit les plus répandus (technologies *IonTorrent*™ et *Illumina*) tirent leur avantage d'un séquençage fortement parallélisé sur une faible longueur (quelques centaines de paires de bases), celui-ci ne peut être réalisé de bout en bout que sur des fragments courts, et seulement aux extrémités sur des fragments plus longs. Dans le cas de fragments suffisamment courts pour être séquencés de bout en bout, l'approche *paired-ends* permet de lire chaque base deux fois (une dans chaque sens), et donc de minimiser les erreurs. Pour des fragments plus longs, cette approche apporte un avantage précieux lors de la phase d'alignement, dans la mesure où elle produit deux lectures non-chevauchantes qui doivent nécessairement être alignées dans la même région du génome, et séparées d'un espace de taille connue (taille moyenne des fragments produits). C'est cette dernière approche qui a été retenue ici, avec des lectures de 76 nucléotides aux extrémités de fragments d'environ 1 000 paires de bases.

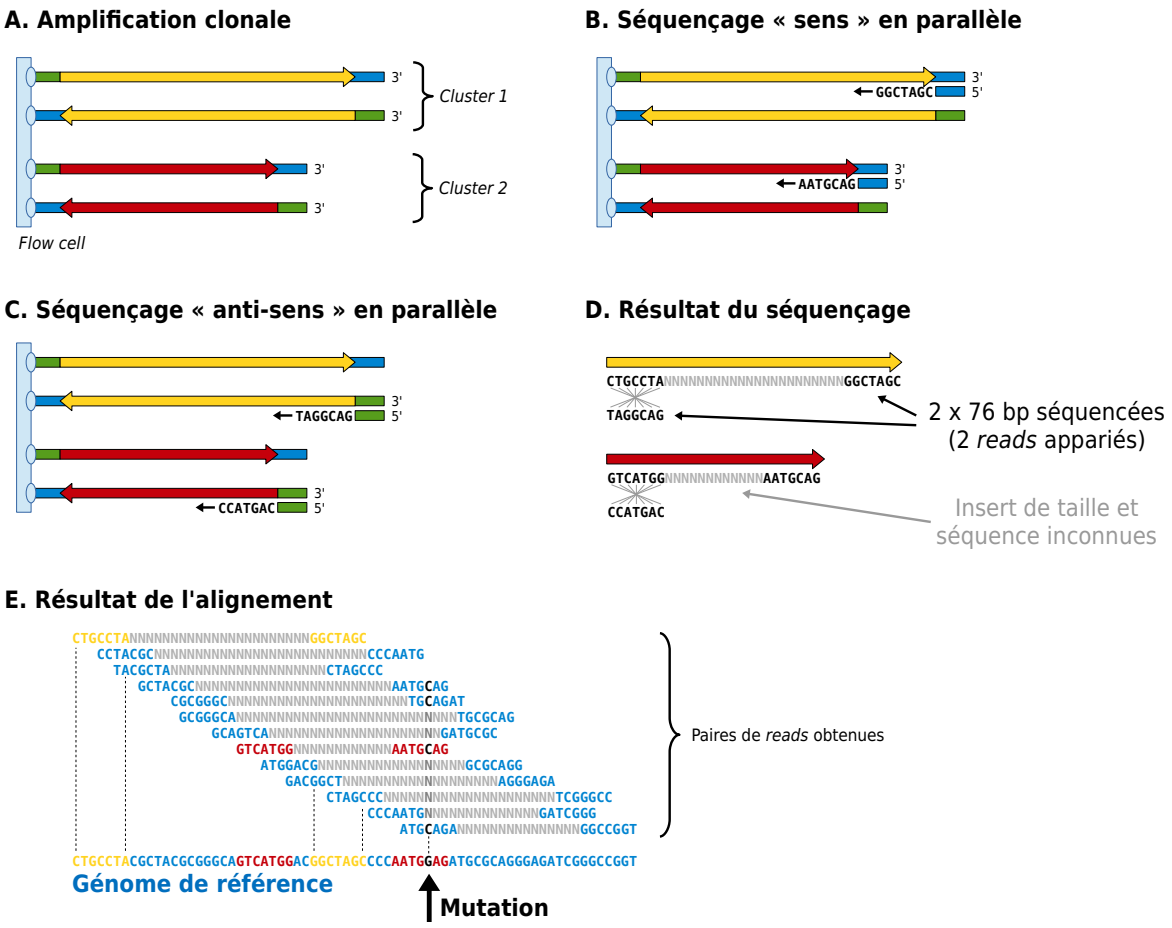


Figure 4.8 – Principe du séquençage *paired-ends*

Les principales étapes (A à E) du séquençage *paired-ends* sont présentées ici, dans le cadre de la technologie Illumina / Solexa.

D'un point de vue pratique, une première lecture est réalisée sur une longueur de 76 paires de bases, selon la chimie décrite précédemment (Figure 4.8B). Chaque *cluster* contenant à la fois la séquence correspondant au fragment d'ADN originel et son inverse complémentaire (du fait de la PCR), ce premier séquençage ne concerne en fait que la moitié des membres du *cluster*. La déshybridation du brin synthétisé et l'utilisation de la seconde amorce permet ainsi de séquencer l'autre extrémité des fragments, sans modifier la *flow cell* et selon le même principe (Figure 4.8C). La séquence des deux extrémités de chaque fragment sont ainsi connues, et séparées d'un insert de séquence et de taille inconnues, quoique cette dernière soit estimable à partir de la taille des fragments (Figure 4.8D). Pour peu que les fragments aient été réalisés sur suffisamment de génomes et avec des cassures suffisamment aléatoires, chaque position de la région cible sera ainsi couverte par plusieurs lectures, permettant ainsi la détection de mutations (Figure 4.8E).

4.4 Séquençage d'exomes de LDGCB réfractaires

4.4.1 Pipeline d'analyse

Dans la mesure où ce projet constituait la première expérience en matière de séquençage exomique au sein de notre laboratoire, la première étape a consisté à assembler un pipeline d'analyse adapté à nos besoins et au matériel informatique à notre disposition. L'analyse devait débuter des fichiers FASTQ contenant les séquences et qualités des lectures opérées par le séquençage, mettre en évidence les mutations somatiques en tenant compte de l'ADN génomique également séquencé, et retourner une liste de variants filtrés pour leur pertinence biologique.

D'un point de vue plus technique, nous disposions pour son exécution d'un serveur de calcul local doté de 12 cœurs cadencés à 2,66 GHz, 48 Go de mémoire vive et 12 To de stockage, présent dans notre laboratoire dans le cadre d'un autre projet de séquençage (*Torrent Server*™). Le développement de ce pipeline s'inscrit dans le cadre des *Best practices* publiées par le *Broad Institute* [308], et comprend uniquement des logiciels libres mis à disposition par la communauté scientifique. Chacune des étapes de ce pipeline est pilotée par un script *Bash* autonome, paramétré par divers arguments respectant la norme POSIX [309]. Le pipeline dans sa globalité est piloté par un script *Bash* gérant l'appel et la journalisation des différentes étapes, selon un fichier de configuration fourni par l'utilisateur.

La première étape de ce pipeline vise à **aligner** les paires de lectures (Figure 4.9), grâce à l'algorithme MEM intégré dans BWA [310]. Il s'agit d'une étape de grande complexité algorithmique, qui tente de retrouver pour chacune des 90 millions de

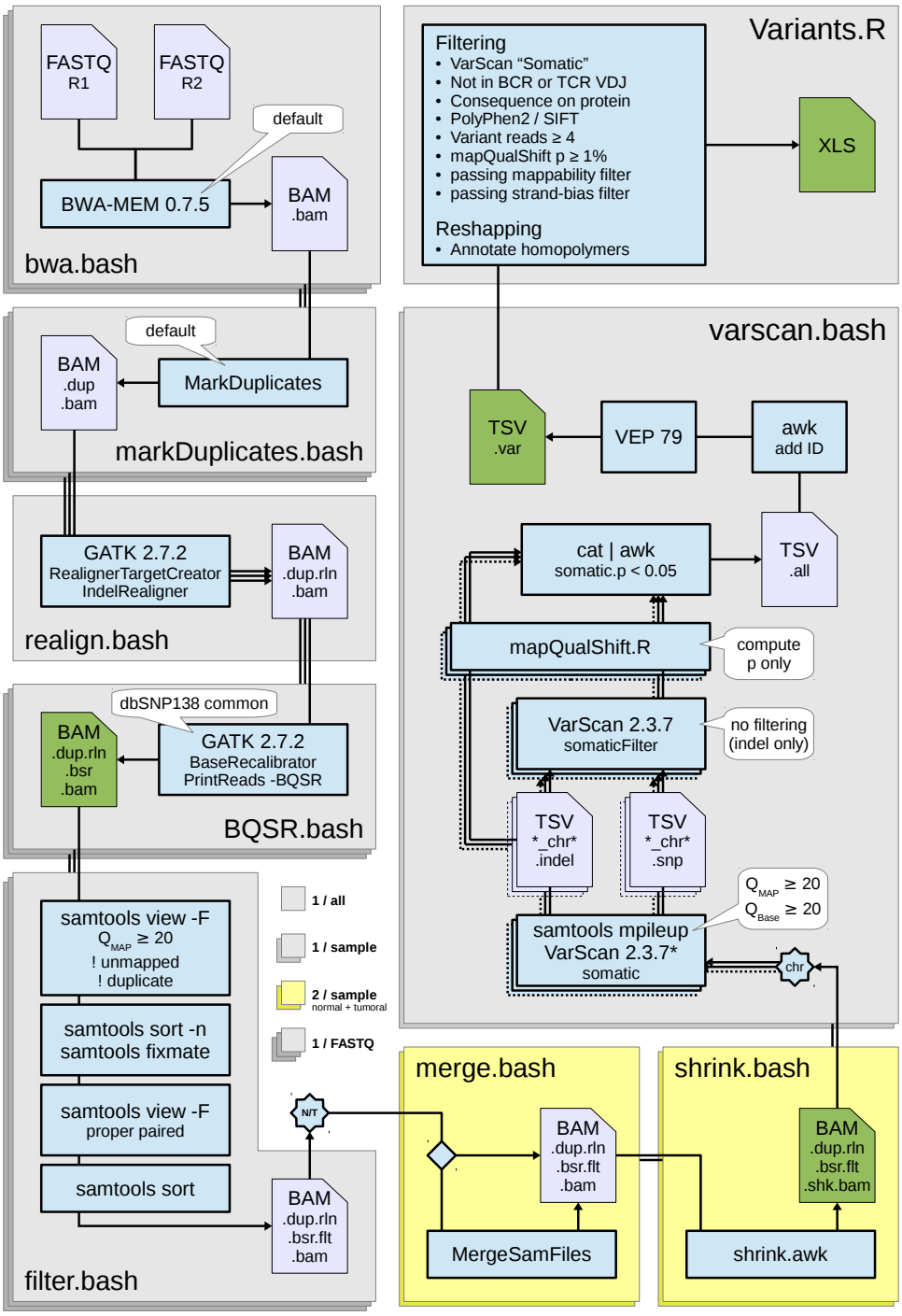


Figure 4.9 – Présentation détaillée du pipeline d'analyse d'exomes

Les différents scripts *Bash* constituant le pipeline d'analyse de données de séquençage exomique sont présentés sur cette figure, des fichiers FASTQ originaux jusqu'au tableau XLS de variants. Les fichiers temporaires sont représentés en mauve, les fichiers conservés à long terme en vert. Chaque encadré bleu correspond à une étape de traitement, basée sur un unique outil dont le numéro de version utilisé est signalé. Les bulles blanches apportent des précisions quant aux paramètres utilisés pour ces outils. Chaque script peut être lancé une ou plusieurs fois, selon le niveau auquel il fonctionne : échantillon, patient ou série complète (légende en bas à gauche sur l'image).

lectures la région du génome de référence dont elle provient, et ce pour chacun des 28 échantillons séquencés. L'objectif étant de détecter des variations vis-à-vis de cette séquence de référence, un certain niveau de mésappariement doit être toléré dans cette analyse, en minimisant toutefois le nombre de solutions. BWA repose ainsi sur l'extension d'une ou plusieurs « graines » d'une dizaine de nucléotides, pour lesquelles la lecture à aligner et le génome de référence sont strictement identiques. Si ces graines correspondent généralement à plusieurs régions du génome, elles ont la particularité d'être aisément identifiables via l'utilisation de structures de stockage des données adéquates. La combinaison de plusieurs graines sur la même lecture réduit ainsi les possibilités d'alignement, rendant réalisable une approche plus traditionnelle pour aligner les dernières bases. Plusieurs alignements sont généralement possibles pour une lecture de 76 paires de bases, et la mise en relation avec les alignements possibles de la lecture appariée (séquençage *paired-ends*) permet généralement de trancher. Un score d'alignement tenant compte de la qualité et de la quantité d'alignements trouvés est finalement attribué, laissant à l'utilisateur le soin de fixer un seuil de tolérance (score de 20 au moins dans notre cas, soit 1% de risque d'erreur selon l'échelle Phred [311]).

La seconde étape consiste à identifier les duplicats parmi les lectures (Figure 4.9), afin de s'assurer que chaque fragment d'ADN originel n'ait été lu qu'une seule fois. Des **duplicats optiques** peuvent être lus avec la technologie Illumina / Solexa, lorsqu'un *cluster* de forme inhabituelle est identifié comme deux *clusters* distincts par les algorithmes d'analyses d'image (Figure 4.7, panneau bleu). Ces duplicats sont aisément identifiables *a posteriori* par leurs séquences identiques et leurs coordonnées voisines sur la *flow cell*. Des **duplicats de PCR** peuvent également survenir, dans la mesure où quelques cycles d'amplification sont généralement appliquées aux librairies avant séquençage. Il s'agit dans ce second cas de deux fragments distants sur la *flow cell*, qui présentent des paires de lectures identiques, un phénomène improbable hors de ce contexte car il nécessiterait que les deux cassures définissant le fragment aient lieu strictement aux mêmes endroits dans deux génomes distincts. L'outil *markDuplicates* de la suite *Picard* [312] est utilisé afin de réaliser cette filtration.

La troisième étape consiste en un **réalignement local autour des *indels*** (Figure 4.9), grâce aux outils de la suite GATK [313]. L'objectif de cette analyse est d'homogénéiser l'alignement des courtes insertions et délétions, qui ont souvent plusieurs solutions équivalentes. Par exemple, la perte d'un élément répété en tandem peut être annotée de différentes manières (perte du premier élément, ou du second, etc), et les algorithmes d'alignement choisissent l'une ou l'autre selon les cas. Il convient donc d'homogénéiser cette annotation, afin de ne pas manquer d'événements récurrents annotés sous plusieurs nomenclatures différentes.

La quatrième étape, également prise en charge par GATK [313], consiste à **recalibrer les scores de qualité des bases** (BQSR, Figure 4.9). Lors de la lecture d'une base (Figure 4.7), le séquenceur tente d'évaluer le degré de certitude à placer dans le résultat sous forme d'un score. En théorie, ces scores respectent l'échelle Phred [311], et quantifient de manière absolue le taux d'erreurs commises par le séquenceur. Dans la pratique, de grandes disparités sont observées entre fabricants, pas toujours en adéquation avec le taux d'erreur réel. BQSR se propose donc de recalculer et homogénéiser ces scores, en distinguant mutations (polymorphismes présents dans des banques publiques) et erreurs de séquençage probables. On peut noter que dans le cas d'un ADN tumoral les mutations somatiques sont considérées comme des erreurs de séquençage par BQSR, néanmoins leur nombre minoritaire face aux polymorphismes ne devrait pas mettre en péril la validité du modèle [314].

Suit une étape de **filtration** réalisée grâce à SAMtools [315] (Figure 4.9), afin d'éliminer les lectures présentant un score d'alignement insuffisant (inférieur à 20) ou marquées précédemment comme duplicats. Les lectures appariées aux lectures filtrées sont également retirées, de sorte de ne garder que des paires de lectures légitimes. Un des ADN tumoraux ayant été séquencé deux fois faute de profondeur suffisante lors du premier séquençage, une étape de **fusion** basée sur Picard [312] a également été ajoutée. Enfin une étape d'**allègement**, visant à retirer la plupart des informations optionnelles ajoutées par les différents outils du pipeline, permet de diviser la taille des fichiers BAM par 2 environ.

Une fois les données nettoyées, les fichiers BAM sont utilisés pour la **recherche de mutations** (Figure 4.9). Cette tâche est confiée à VarScan 2 [316], dont le mode « somatique » est souvent considéré comme l'une des références en la matière [317], notamment en terme de sensibilité [318]. La complexité de son paramétrage en fait un outil parfois décrié à tort, et il convient d'appliquer un certain nombre de filtres en aval pour limiter les faux-positifs (voir paragraphes suivants), ce qui n'est pas réalisé systématiquement dans les comparatifs publiés [319, 320]. D'un point de vue technique, VarScan repose sur deux tests exacts de Fisher. Le premier de ces tests est réalisé indépendamment dans l'ADN normal et tumoral, afin de mettre en évidence (ou non) un nombre de lectures mutées supérieur au taux d'erreur attendu du séquenceur (estimé à 0,1% pour cette technologie). Le second test est réalisé si l'ADN tumoral passe le premier test et l'ADN normal non, afin de quantifier la significativité de l'association entre lectures portant la mutation et origine tumorale du fichier BAM considéré. Un certain nombre d'heuristiques permettent de limiter le nombre de tests réellement effectués, en instaurant notamment un minimum de profondeur et un minimum de lectures mutées (en nombre et en proportion).

Les variants somatiques mis en évidence par VarScan sont ensuite **annotés** par VEP (Figure 4.9), grâce aux données stockées dans la banque Ensembl [298]. La position génomique de chaque variant permet d'identifier le gène potentiellement impacté, et les modifications que la mutation induit au niveau de l'ARN et de la protéine codés par ce gène. Les prédictions de pathogénicité produites par les algorithmes PolyPhen 2 [321] et SIFT [322] sont également extraites de cette base, afin d'écarter les mutations probablement bénignes du fait de la substitution d'un acide aminé par un autre aux propriétés physico-chimiques équivalentes, ou à une position peu conservée au sein des espèces proches.

Divers **filtres** sont finalement appliqués à cette liste de variants (Figure 4.9), afin de concentrer les efforts sur les variants les plus à même d'expliquer la pathogénicité des tumeurs étudiées, parmi lesquels :

— Un filtre **fonctionnel** relativement conservateur, n'excluant que les variants synonymes, non-codants ou prédits bénins par SIFT **et** PolyPhen.
— Un filtre de **biais de brin** basé sur le test exact de Fisher, excluant les variants observés toujours dans le même sens de séquençage.
— Un filtre de **biais d'alignement** basé sur le test de Mann-Whitney, excluant les variants dont l'allèle muté est associé à une qualité d'alignement des lectures significativement plus faible que l'allèle sauvage.
— Un filtre d'**alignabilité**, augmentant le seuil appliqué à la p valeur calculée par VarScan dans les régions dupliquées du génome humain, dans la mesure où les artefacts d'alignement y sont plus nombreux.

4.4.2 Résultats et discussion

L'application de ce pipeline d'analyse à 14 paires de LDGCB ayant rechuté moins d'un an après leur diagnostic a permis de mettre en évidence 1 541 mutations ponctuelles et 184 *indels*, pour une moyenne de 123,2 mutations par patient. 94% des 92 millions de lectures collectées par échantillon ont passé les filtres de qualité et d'alignement, résultant en une profondeur de couverture moyenne de 79x. Compte tenu des paramètres de VarScan et des estimations de contenu en cellules tumorales, 86,8% de l'exome en moyenne est ainsi suffisamment couvert pour mettre en évidence une mutation hétérozygote clonale.

Cette étude a livré pour la première fois le séquençage exomique de 5 LPM, mettant en évidence un taux de mutation significativement plus élevé que dans les LDGCB GCB et ABC. L'application de MuSiC SMG [323] à chaque sous-type a permis de mettre en évidence 63 gènes présentant un taux de mutation supérieur à la normale, comprenant de nombreux gènes connus (*PIM1*, *TP53*, *BCL2*, *MEF2B*, *CD58*, *B2M*...

§4.1.3) mais également des pistes intéressantes qui se sont vues confirmées par la suite (*XPO1*, *MFHAS1*) [271, 272, 324, 325]. L'implication de la voie NFκB dans les 3 sous-types a pu être détectée par Path-Scan [326], et plusieurs voies susceptibles de contribuer au phénotype réfractaire ont été avancées. La détection de CNA via une adaptation du logiciel ADTEx [327] a conduit à l'identification de 23 MCR, confirmant là encore des événements connus (délétion 9p21, gains en 2p15 et 9p22... §3.1.1). Le croisement entre CNA et mutations somatiques a également permis de lister 143 événements de sélection de l'allèle muté, dont notamment l'amplification focale du gène muté *PCLO*.

Sur un plan plus bio-informatique, le contenu en cellules normales et tumorales des échantillons a été finement modélisé en se basant sur les fréquences des mutations dans les échantillons appariés, et sur les log-ratios des CNA mis en évidence. Ces estimations se sont avérées en moyenne deux fois inférieures aux estimations histologiques, ce qui suggère de considérer ces dernières avec prudence lors de l'élaboration d'un plan d'analyse de séquençage à haut débit. Une contamination tumorale des échantillons normaux (sang périphérique) a également pu être détectée chez deux patients, un phénomène qui peut fausser la détection de variants somatiques s'il n'est pas pris en compte.

4.4.3 Article accepté (GCC)

La description du pipeline d'analyse bio-informatique et les résultats collectés sur ces 14 LDGCB réfractaires ont été acceptés pour publication dans *Genes, Chromosomes and Cancer* (IF 4,3) en octobre 2015, sous forme d'un article de recherche [328]. Ces données ont également été présentées sous forme de *posters* au congrès annuel de l'association américaine de recherche sur le cancer (AACR) en avril 2015 [329], ainsi que dans des congrès régionaux [330, 331].

Whole Exome Sequencing of Relapsed / Refractory Patients Expands the Repertoire of Somatic Mutations in Diffuse Large B-Cell Lymphoma

Sylvain Mareschal [1], Sydney Dubois [1], Pierre-Julien Viailly [1,2], Philippe Bertrand [1], Elodie Bohers [1], Catherine Maingonnat [1], Jean-Philippe Jaïs [3], Bruno Tesson [4], Philippe Ruminy [1], Pauline Peyrouze [5], Christiane Copie-Bergman [6], Thierry Fest [7], Thierry Jo Molina [8], Corinne Haioun [9], Gilles Salles [10], Hervé Tilly [1], Thierry Lecroq [2], Karen Leroy [6], Fabrice Jardin [1]*

1 INSERM U918, Centre Henri Becquerel, Rouen, France
2 LITIS, INSA EA 4108, Saint-Etienne-du-Rouvray, France
3 INSERM UMRS 872, AP-HP Hôpital Necker, Paris, France
4 Institut Carnot CALYM, Pierre-Bénite, France
5 Plate-forme de génomique fonctionnelle et structurale, Université de Lille 2, Lille, France
6 INSERM U955, IMRB, AP-HP Hôpital Henri Mondor, Créteil, France
7 INSERM U917, CHU Pontchaillou, Rennes, France
8 EA 7324, AP-HP Hôpital Necker, Paris, France
9 Unité Hémopathies Lymphoïdes, AP-HP Hôpital Henri Mondor, Créteil, France
10 CNRS UMR 5239, Hospices Civils de Lyon, Lyon, France

Correspondence to:
Pr Fabrice Jardin, fabrice.jardin@chb.unicancer.fr, +33 2 32 08 29 09
INSERM U918, Centre Henri Becquerel, Rue d'Amiens, 76038 Rouen CEDEX 1

Supported by:
The Cancéropole Nord-Ouest (CNO) and the Institut National du Cancer (INCA) funded this work as part of the RELYSE and GHEDI projects, but did not participated in the design, collection, analysis, and interpretation of data. SM received a PhD grant from the Région Haute-Normandie to conduct this work.

Running title:
Exome sequencing of refractory DLBCL

Boilerplate

Abstract

Despite the many efforts already spent to enumerate somatic mutations in diffuse large B-cell lymphoma (DLBCL), previous whole-genome and whole-exome studies conducted on patients of mixed outcomes failed at characterizing the 30% of patients who will relapse or resist current immunochemotherapies. To address this issue, we performed whole-exome sequencing of normal / tumoral DNA pairs in 14 relapsed / refractory (R/R) patients subclassified by full-transcriptome arrays (6 activated B-cell like, 3 germinal center B-cell like and 5 primary mediastinal B-cell lymphomas), from the LNH-03 LYSA clinical trial program. Aside from well-known DLBCL features, gene and pathway level recurrence analyses proposed several interesting leads including *TBL1XR1* and activating mutations in *IRF4* or in the insulin regulation pathway. Sequencing-based copy number analysis defined 23 short recurrently altered regions involving genes such as *REL*, *CDKN2A*, *HYAL2* and *TP53*. Moreover, it highlighted mutations in genes such as *GNA13*, *CARD11, MFHAS1 and PCLO* as associated with secondary variant allele amplification events. The 5 primary mediastinal B-cell lymphomas (PMBL), while unexpected in a R/R cohort, showed a significantly higher mutation rate (p=0.003) and provided many insights on this classical Hodgkin lymphoma related subtype. Novel genes such as *XPO1*, *MFHAS1* and *ITPKB* were found particularly mutated, along with various cytokine-based signaling pathways. Among these analyses, somatic events in the NFκB pathway were found preponderant in the three DLBCL subtypes, confirming its major implication in DLBCL aggressiveness and pinpointing several new candidate genes.

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common form of lymphoma, accounting for 30-40% of newly diagnosed non-Hodgkin lymphomas (NHL). Despite its biological aggressiveness, it has been demonstrated that the addition of rituximab (R) to anthracycline-based regimens (mainly CHOP chemotherapy: cyclophosphamide, doxorubicin, vincristine and prednisone) can cure more than 50% of cases (Lenz and Staudt, 2010; Roschewski et al., 2014). However, one-third of patients are still considered as primarily refractory or experience short-term relapse (R/R) with current treatments (Roschewski et al., 2014).

The molecular heterogeneity of DLBCLs has been deciphered by gene expression profiling (GEP), defining three main subtypes termed activated B-cell like (ABC), germinal-center B-cell like (GCB) and primary mediastinal B-cell lymphoma (PMBL). These subtypes arise from distinct normal counterpart B cells at diverse stages of differentiation and maturation, leading to different clinical outcomes and responses to immunochemotherapy (Roschewski et al., 2014; Sehn and Gascoyne, 2015).

Over the past decade, next-generation sequencing (NGS) technologies, which allow for massively parallel high-throughput DNA sequencing, have emerged and provided new insights into the genomic characterization of DLBCL, offering new therapeutic opportunities. To date, whole-exome (WES) or whole-genome sequences (WGS) obtained from approximately 300 DLBCL tumors have been published (Jardin, 2014; Bohers et al., 2015). These studies have identified recurrent somatic mutations targeting genes involved in B-cell receptor (BCR) signaling, nuclear factor κB (NFκB) pathway, toll-like receptor (TLR) signaling, phosphatidyl inositol 3-kinase (PI3K) pathway, immunity, cell-cycle/apoptosis or epigenetic regulation (Jardin, 2014; Bohers et al., 2015). In keeping with these works, a median non-synonymous somatic mutation rate of 3.3 per megabase (Mb) was reported in DLBCL (Lawrence et al., 2014). It has been estimated that creating a comprehensive catalog of genes in which somatic mutations occur at a minimum 5% higher rate than this mutational background would require a dataset of approximately 400 DLBCL samples (Lawrence et al., 2014). This indicates that a complete compilation of DLBCL genes will be obtained in the near future. To increase the probability of detecting new relevant variants in DLBCL, WES experiments have been focused on certain molecular or clinical subtypes, including PMBL (Gunawardana et al., 2014), primary central nervous system lymphoma (PCNSL) (Bruno et al., 2014), transformed follicular lymphoma (tFL) (Okosun et al., 2014), and DLBCL emerging in a distinct genetic background (de Miranda et al., 2014), pinpointing genetic alterations likely to play a driver role during lymphomagenesis.

To date it is unclear if recurrent somatic mutations may play a role in the emergence of refractoriness or short-term relapse in DLBCL. This is suggested by targeted sequencing data, highlighting the prognostic value of certain recurrent somatic mutations in genes such as *MYD88* (Fernández-Rodríguez et al., 2014), *SOCS1* (Schif et al., 2013), *FOXO1* (Trinh et al., 2013) or *TP53* (Li et al., 2013).

With the aim of detecting genes that might impact the R/R phenotype in DLBCL, we performed WES in

Soumis

a series of R/R patients included in the Lymphoma Study Association (LYSA) clinical trial programs, belonging to the three main molecular subtypes (GCB, ABC and PMBL) and compared our results to the main published DLBCL variant datasets (Pasqualucci et al., 2011; Lohr et al., 2012; Morin et al., 2013; Zhang et al., 2013; de Miranda et al., 2014).

MATERIALS AND METHODS

Patients and Samples

75 patients with de novo DLBCL are described in this study (Table 1), at the intersection between the GOAL and GHEDI cohorts of the LNH-03 multicentric clinical trial program led by the LYSA (Fitoussi et al., 2011; Peyrade et al., 2011; Récher et al., 2011; Delarue et al., 2013; Ketterer et al., 2013; Molina et al., 2014). Patients from these trials signed informed consent forms, and these studies were approved by an independent research ethics committee. 14 patients who relapsed in a year or less under immunochemotherapy (median of 8.9 months, Table S1) were selected as the R/R cohort, based on their clinical presentations and the availability of biological samples (Figure 1). Tumoral DNA and RNA samples were extracted from frozen lymph node biopsies at diagnosis, and normal DNA was extracted from peripheral blood mononuclear cells (PBMC). The clinical presentation at diagnosis was not significantly distinct between these 14 patients and the rest of the cohort (Fisher p=0.23, Table 1). Only one patient presented simultaneous rearrangements of *MYC* and *BCL2* (Table S1), as determined by FISH.

Gene Expression Profiling

Whole transcriptomes of the 75 patients were assessed by GeneChip Human Genome U133+2 arrays following the manufacturer recommendations (Affymetrix, Santa Clara, CA). Raw feature normalization and quality checks were handled using Bioconductor software (affy, affyQCReport, GCRMA), excluding multi-gene probesets (x). ABC, GCB and “other” subtypes were determined by applying the reference Bayesian predictor (Wright et al., 2003). PMBL were subsequently pointed out by hierarchical clustering (complete distance, Ward agglomeration) of a previously published gene signature (Rosenwald et al., 2003), excluding *TCL1A* and *E2F2* that could not be reliably measured on U133+2 arrays.

Whole-Exome Sequencing

Whole-exome sequencing (WES) of the 14 normal / tumoral DNA sample pairs was subcontracted to Integragen SA (Evry, France). Exonic DNA fragments were enriched using the SureSelect Human All Exon V5 kit (Agilent Technologies, Santa Clara, CA), defining an experimental target of about 89 Mb (capture probe mapping regions extended by 100 bp on each side). 75 bp paired-end sequencing was

Soumis pour publication

then performed using a HiSeq2000 (Illumina, San Diego, CA) following the manufacturer's recommendations. The primary analysis (from signal capture to FASTQ files) was handled by the manufacturer's Real Time Analysis software (version 1.4) with default settings. The secondary analysis (from FASTQ files to analysis-ready BAM files) was managed by a custom bioinformatics pipeline (Supplemental Methods and Figure S1). Somatic variant calling was performed by VarScan (Koboldt et al., 2012) followed by various additional filters (Table 2), putting the focus on a region of interest (ROI) consisting of all coding exons (Ensembl genes 79) (Cunningham et al., 2015) extended by 8 bp on each side (Supplemental Methods). Gene and pathway enrichment analyses were performed separately using the SMG and PathScan tools from the MuSiC suite respectively (Dees et al., 2012) (Supplemental Methods). Copy number and copy-neutral loss of heterozygosity analyses were performed using an adaptation of ADTEx (Amarasinghe et al., 2014) (Supplemental Methods). The full list of somatic variants is provided as Dataset S2, and raw data are available upon request.

sous Presse

RESULTS

Molecular Characterization of the Somatic Variants

Fourteen normal / tumoral DNA pairs from R/R patients were sequenced after exonic fragment capture. An average of 92 million reads [78-133] was produced per sample, of which 94% [92.7-94.3] were retained after duplicate and unmapped read filtering (Dataset S1). Of the remaining reads, 81% [78.6-83.9] could be mapped to the experimental target (as provided by the manufacturer) and 60% [57.0-62.0] to the ROI, resulting in an average sequencing depth over the ROI of 79-fold [68-114]. Sample purity estimation based on normal / tumoral variant allele frequency (VAF) comparison revealed a tumoral contamination of normal samples in 2/14 cases (10% and 20% respectively), and a mean tumoral cellularity of 47% [20-85] (Figure S2). Of note, histological estimations of biopsy contents were approximately two times higher, averaging 83% [50-90] of tumoral cells. The VAF-based contamination level estimates were passed to VarScan to adjust the variant-calling (Supplemental Methods). According to these estimates and VarScan settings, the sequencing depth was theoretically sufficient to call clonal heterozygous mutations in 86.8% [75-93] of the ROI (Dataset S1, Supplemental Methods).

A total of 1 541 somatic single nucleotide variants (SNVs) and 184 short somatic insertions or deletions (indels) were called across the 14 samples (Dataset S2), for an average variant count of 123.2 per patient [21-197]. The average mutation rate of 3.2 per Mb [0.55-5.1] observed in these samples was similar to previous reports (Lohr et al., 2012; Lawrence et al., 2014), and significantly higher in the PMBL subtype as compared to the two other subtypes (Mann-Whitney p=0.003, Figure 2A). The variant calling accuracy was estimated by resequencing 12 patients with an orthogonal method (Life PGM), resulting in a precision of [97-100]% and a sensitivity of [80-93]% (Supplemental Methods).

As reported by others (Lohr et al., 2012; Alexandrov et al., 2013), most coding region substitutions consisted in C>T substitutions (55% [41-63], Figure 2B and Figure S3A), and more specifically in CpG deaminations (29% [22-47]). This trend is intensified by the normalization on tri-nucleotide occurrence rates (Figure S3B), but could not be linked to specific patients or subtypes.

82% [72-92] of the somatic indels detected consisted in the gain or loss of monomers from a repetitive region (Figure 2C), including a vast majority of homopolymers (68% [57-80]) of 10 repetitions on average (Figure 2E). As larger error rates were reported with the Illumina sequencing technology in long homopolymers (Minoche et al., 2011; Ross et al., 2013), we tried to differentiate reliable somatic indels from indels likely due to this bias (termed “dubious”, Supplemental Methods). While larger proportions of indels were filtered out above 10 repetitions (Figure 2E), most of the indels showed variant allele counts in tumoral samples incompatible with the error rates reported at these levels of repetition (FDR<0.001). This indel enrichment in repetitive sequences was less present in coding regions, in which it represented only 58% of the 74 reported indels, including 6 homopolymers classified as “dubious” (Figure 2D).

Variants Called in WES Data

Variants Shared across Molecular DLBCL Subtypes

15 genes were found significantly mutated in the whole cohort rather than a specific subtype (FDR<10%, Figure 3). As described in previous WES studies (Morin et al., 2011, 2013; Lohr et al., 2012; Zhang et al., 2013), many truncating events were found in *GNA13*. The Q27* variant was found twice, and altogether one GCB, one ABC (17%) and two PMBL (40%) samples were found mutated, which might suggest a bolder enrichment in PMBLs. *PIM1* was found frequently inactivated in non-PMBL samples, especially by 3 splice-site and 2 frameshift mutations (FDRs of 10^{-12} and 10^{-6} for ABC and GCB samples respectively). *TP53* and *KMT2D* were also part of these mutated genes, with a particularly high amount of strong inactivating events in the latter (3 stop-gains and 2 frameshift indels).

Finally *IGLL5* and *TMSB4X* showed a significant amount of mutations, as recently described in a WES study of chinese DLBCL patients (de Miranda et al., 2014). *IGLL5* mutations were found only in the first two exons, outside the immunoglobulin-like domain, and still need to be assessed for biological relevance. Due to its small size (44 amino acids), *TMSB4X* was easily found significantly mutated by MuSiC, and reports of its involvement in interferon γ (Lee et al., 2011) and Akt (Piao et al., 2014) signaling pathways offer several promising leads for future studies.

ABC Subtype Variants

10 genes were found significantly more mutated than expected in the 6 ABC samples (FDR<10%, Figure 3), including 4 genes previously described as frequently mutated in DLBCLs (*MYD88*, *CD58*, *TBL1XR1* and *IRF4*). *MYD88* mutation was confirmed as one of the most ABC-specific events, with two

Soumis à Presse

L265P mutations in 3 mutated cases (Bohers et al., 2015). The four *CD58*-related events were all truncating, confirming the advantage offered by the loss of this key T-cell immunity effector (Challa-Malladi et al., 2011). *TBL1XR1* mutations were only found in ABC samples (2/6), consistent with the PCNSL mutation patterns that have been shown to heavily correlate with the ABC DLBCL subtype (Bruno et al., 2014).

IRF4 was found mutated in one patient of each subtype (at the same C99R residue for the GCB and PMBL samples), with 3 mutations for the sole ABC sample. All the mutations were found in the DNA-binding domain of IRF4, which represents 25% of the protein. High *IRF4* expression, one of the most distinctive features of ABC samples (Wright et al., 2003), is also observed in this heavily mutated ABC sample (Figure 4A). Moreover, *IRF4* activity in the mutated ABC sample (evaluated by the expression of genes it regulates, Supplemental Methods) was comparable to that of the two ABC samples with evidence of *IRF4* amplification (Figure 4B), suggesting that these are activating mutations.

GCB Subtype Variants

Due to the small number of GCB samples in this series of R/R patients, only 7 genes were found significantly more mutated than expected (FDR<10%, Figure 3). *BCL2* was found mutated 6 times and rearranged (Table S1) in the same GCB sample, fitting with its identified role as a somatic hypermutation target (Burkhard et al., 2015). Two GCB samples were found mutated in *MEF2B*, including one Y69H mutation previously described as enhancing *BCL6* expression in DLBCLs (Ying et al., 2013) and one deleterious missense mutation associated with a copy-neutral loss of heterozygosity (CNLOH).

Two missense mutations in the tyrosine phosphatase domain of *DUSP2* were found in GCB samples, as well as a splice-site deletion potentially leading to the loss of this domain. This tendency toward *DUSP2* inactivation is coherent with its reported inhibitory effect on MAPK1 and MAPK3 (Ward et al., 1994), two MAP kinases controlling various proliferation and cell survival processes.

Truncating events of the NFκB inhibitor α (*NFKBIA*) were found in one GCB and one PMBL sample (double frameshift indel and stop-gain respectively). Both frameshift indels occur in the first exon, potentially leading to protein sequence jamming. Inactivating this inhibitor provides an alternative way to activate the NFκB pathway, previously reported in classical Hodgkin lymphoma (cHL) (Lake et al., 2009).

PMBL Subtype Variants

31 genes were found significantly mutated in the 6 PMBL samples (FDR<10%, Figure 3), including *SOCS1*, *STAT6* and *TNFAIP3*, already identified as involved in PMBL pathogenicity (Melzner et al., 2005; Weniger et al., 2006; Ritz et al., 2009). *SOCS1* alterations were found in all 5 PMBL samples, including frequent truncating events (4 frameshift indels, 2 stop-gain and 2 whole-gene deletions), and a recurring F79I missense mutation at the edge of the SH2 domain. Three missense *STAT6* mutations, all in the DNA binding domain, were found in 3 PMBL samples. Frameshifts, splice-site mutations and

sous presse

whole-gene deletions (including one focal homozygous deletion in a PMBL) were frequently found in *TNFAIP3*, confirming the preponderance of its inactivation in PMBL (Schmitz et al., 2009), and more generally in DLBCL. *ACTB* was also found mutated in 3 PMBL cases, preferentially in its amino terminus end (5/6 in the first 50 amino acids), as previously described in non-PMBL DLBCLs (Lohr et al., 2012).

Strong evidence of *B2M* inactivation was found in 3 PMBL samples, coupling frameshift indels (two at the start of the protein and one at the end of the immunoglobulin-like domain) and loss of heterozygosity (two localized deletions and a CNLOH of chromosome 15). This correlates with similar findings in cHL (Reichel et al., 2015), thought to be very similar to PMBL DLBCL. Moreover, the GHE0536 sample showed both *B2M* and *CD58* bi-allelic inactivations, via conjoint frameshift indels and whole gene deletions, a co-occurrence consistent with previous reports in DLBCL (Challa-Malladi et al., 2011).

ITPKB was mutated in 2 samples, with a single PMBL patient harboring 7 missense mutations (GHE0645). While never described as mutated in DLBCL patients, this second messenger kinase has been linked to B (Maréchal et al., 2007; Miller et al., 2007) and T (Pouillon et al., 2003) cell survival and differentiation and *ITPKB* mutations have been recently reported in cHL (Reichel et al., 2015).

Two PMBL samples presented an identical 4 base frameshift deletion in *NFKBIE*, a few amino acids before the stretch of ankyrin repeats mediating its import to the nucleus (Lee and Hannink, 2002). Aside from this mutation enrichment analysis, a single ABC sample showed a whole-gene deletion of *NFKBIE* that could lead to similar results. Similarly to *NFKBIA* described above, inactivation of this NFκB nucleus exportin provides an effective means to maintain NFκB pathway activation, as suggested by similar findings in cHL (Emmerich et al., 2003).

The E571K locus of *XPO1*, previously described as mutated in chronic lymphocytic leukemia (CLL) cases at a frequency <5%, was found mutated in two PMBL samples. The CRM1 protein encoded by *XPO1* is known to export various key actors in oncology, such as p53 (Yoshimura et al., 2014) and the NFκB complex (Tam et al., 2000). It is also an actionable target that has shown promising results in Mantle Cell Lymphoma (Yoshimura et al., 2014) and T-cell acute lymphoblastic leukaemia (Etchin et al., 2013).

Pathway Enrichment Analysis

As background mutation rates and tumoral mechanisms vary by subtype, each subtype was scanned separately for pathways enriched in mutations (Dataset S3, Supplemental Methods). Resulting pathways were assessed by similar enrichment analyses in two external studies (Lohr et al., 2012; Morin et al., 2013), to pinpoint pathways that could explain the R/R phenotype of our patients (Supplemental Methods and Dataset S4). As none of these external cohorts contained averred PMBL samples, comparison could only be achieved for ABC and GCB subtypes.

submitted

ABC Subtype Pathways

Several pathways were found significantly enriched in ABC samples (Figure 5A and Dataset S3) due to mutations of the histone monomers (*HISTH1B/C/D/E*), including pathways involved in apoptosis (FDR~10^{-9}) and senescence (FDR~10^{-5}).

NFκB signaling pathway was, as expected, also significantly enriched in mutations among ABC patients (FDR~10^{-4}), presenting mutations in *MYD88*, *TNFAIP3*, *CARD11*, *TAB3, BCL2A1* and *ZAP70*.

Pathways involved in sucrose degradation were significantly enriched in mutations among ABC patients (FDR~10^{-3}). These pathways were not identified as significantly mutated in a previous DLBCL WGS study (Morin et al., 2013), potentially illustrating their impact in R/R patients specifically. One ABC patient presented associated mutations in *KHK* and *SI*, both involved in sucrose degradation. Mutations of *KHK* and *SI* point to an alteration of cellular metabolism, one of cancer's hallmarks (Hanahan and Weinberg, 2011). Mutations in *SI* have recently been identified in CLL patients (Wang et al., 2011; Quesada et al., 2012; Rodríguez et al., 2013), leading to an upregulation of pro-survival BCR signaling and cancer metabolic reprogramming pathways (Rodríguez et al., 2013).

GCB Subtype Pathways

While mostly known for its involvement in ABC lymphomagenesis, evidence of NFκB disruption was also found in the GCB subtype, via mutations of genes such as *ERC1*, *BCL2*, *BCL10*, *NFKBIA* and *CARD11* (FDR~10^{-3}).

Tryptophan degradation was identified as a pathway enriched for mutations in GCB DLBCL (FDR~10^{-3}, Figure 5B and Dataset S3), with mutations in *IDO1* and *TDO2*. Both IDO1 and TDO2 initiate the first rate-limiting reaction of tryptophan degradation. This pathway was found exclusively enriched in mutations in our R/R cohort, as compared to Morin et al, indicating that they are potentially specific to the GCB R/R phenotype.

PMBL Subtype Pathways

Clustering analysis segregated pathways centered on known driver mutations of PMBL such as *SOCS1*, *STAT6* and *PTPN1 (Steidl and Gascoyne, 2011; Gunawardana et al., 2014)* (Figure 5C and Dataset S3), mainly in the context of JAK-STAT and IL4 signaling pathways (FDR~10^{-6} and 10^{-10} respectively). Associated with *IL4R*, *IRF1/4/8*, *B2M* and *CIITA*, these driver mutations also suggest the interferon α, β (FDR~10^{-6}) and γ (FDR~10^{-9}) signaling pathways as particularly mutated in PMBLs. Many genes that are both PMBL and cHL signature genes have been shown to be interferon-inducible genes, highlighting yet another molecular resemblance between these two entities (Rosenwald et al., 2003).

A large cluster centered on histone gene mutations was also identified, involved in diverse pathways such as meiosis and epigenetic regulation. Interestingly, a sub-cluster including histone gene mutations associated with *TP53, MAPK10* and *TNIK* mutations was involved in oxidative stress-induced senescence (FDR~10^{-5}). Another sub-cluster including histone gene mutations associated with *TP53* and *RAD50* mutations was linked to DNA damage and telomere stress-induced senescence (FDR~10^{-}

Nous presse

[5]), highlighting this pathway as a potential target of interest. In a recent study, *RAD50* mutations were identified in 3% of familial cases of breast cancer without *BRCA1/2* mutations, highlighting their tumorigenic capability (Aloraifi et al., 2015).

XPO1, *TNFAIP3*, *IKBKB*, *NFKBIA* and *NFKBIE* mutations clustered together, leading to NFκB pathway activation (FDR~10^{-3}). Canonical and FCERI-mediated NFκB pathway activation differed by several mutations: the canonical pathway harbored *XPO1* and *TNFAIP3* mutations, whereas the FCERI-mediated pathway exhibited *IGHE* and *IGKC* mutations. However, both pathways presented *IKBKB* and *NFKBIA* mutations, which were respectively missense and stop-gain in PMBL patients, potentially leading to constitutive NFκB pathway activation in PMBL (DiDonato et al., 1997). Furthermore, *XPO1*-encoded CRM1 has been shown to mediate the sub-cellular localization of NFKBIA (Tam et al., 2000): PMBL *XPO1* mutations might therefore also contribute to constitutive NFκB pathway activation.

GNA12, *GNA13* and *GNAQ* mutations clustered together in the S1P2 signaling pathway. This pathway was enriched in PMBL whereas S1P2-deficient mice, and more recently GNA13-deficient mice, have been shown to develop GCB subtype DLBCL (Green et al., 2011; Muppidi et al., 2014). However, in an extended cohort (Dubois et al, submitted), we also showed that *GNA13* mutations were enriched in PMBL subtype, suggesting that S1P2 pathway alterations seem to be events common to both GCB and PMBL subtypes.

Copy Number Variations

Copy number variation (CNV) detection based on sequencing data (Supplemental Methods) revealed typical hallmarks of DLBCL in the R/R cohort (Figure 6 and S6), as previously reported (Lenz et al., 2008; Tirado et al., 2012). These include partial or total monosomies of chromosome 19 and trisomies of chromosomes 3 (ABC), 12 (GCB, PMBL) and 18 (ABC). Centromere-centered loss/gain balances suggestive of isochromosome formation were observed in chromosomes 1 and 6, resulting in the loss of the 6q arm (most recurrently the 6q21 and 6q23-6q24 loci harboring *PRDM1* and *TNFAIP3* respectively). The short arm of chromosome 9 was found gained or amplified in the 5 PMBL samples, with evidence of focal amplification of *JAK2* (+13 copies) and *PDCD1LG2* (+5 and +10 copies) consistent with previous indications of these genes as key components of 9p amplification (Rui et al., 2010; Twa et al., 2014). Large amplifications on chromosomes 5, 7, 20 and 21 were more frequent than previously described, although the 7q21-22 locus was indeed the most amplified of chromosome 7.

Across the 14 R/R patients, 23 minimal common regions (MCR) were detected (Table 3). Six focal deletions of *CDKN2A/B* (including 3 ABC), four focal gains of *REL* (including 3 GCB) and four 19q13 gains (including 3 ABC) were consistent with previous DLBCL studies, although *REL* amplification seemed more frequent than the 28% reported in GCB (Lenz et al., 2008). Evidence of focal deletion of *HYAL2* (Bertrand et al., 2005), *CRIPAK*, *NFKBIL1*, *TNFAIP3*, *TP53BP1*, *B2M*, *CD19*, *TP53* and *TPTE* (*PTEN2*) and amplification of *WDR36* and *MZF1* (Deng et al., 2013) highlight several known and

Sous presse

suspected genes of interest in lymphoma that could be assessed for CNVs in larger series. Considering larger regions of recurrence, the 17q22-23 locus was found gained or amplified several times. While hundreds of genes are located in this region, a shorter region of amplification (+3 copies) suggests that *CD79B* is the key gene of this region.

Focal deletions of T-cell receptor α due to recombination were detected in the 3 samples with lowest purity estimates (20% on average) and GHE1438 (45%), suggesting T-rich tumors. Consistently, focal deletions of the IGH locus were detected in the 6 samples with highest tumor contents (from 50% to 80%) and 2 additional samples. Large regions of germline homozygosity, suggestive of uniparental isodisomy (UPD), were detected in 8 samples at various levels: 5 low (2.4 to 7.6 Mb per sample), two intermediate (18.45 and 85.9 Mb) and one with a significantly higher level of UPD (214.4 Mb split in 18 distinct loci).

Mutated Allele Selection

Of the 1 725 somatic mutations considered, 202 fell in regions also affected by a gain, 51 in regions affected by an amplification (+2 copies or more), 120 in regions affected by a deletion and 86 in regions affected by a CNLOH. The VAF observed for 143 of them was significantly higher than expected in a standard heterozygous case, suggesting that the variant allele was amplified or preserved from LOH (Supplemental Methods and Dataset S5).

Genes involved in such events for multiple samples included *GNA13* (4 mutations in 3 distinct samples), *KMT2D* (2 GCB and 1 ABC), *AASS, CARD11*, *MFHAS1*, *PRDM1* (2 distinct samples each), *PCLO, PTPRC* and *UGP2* (each mutated twice in the same patient). The two missense mutations of *PCLO* were involved in a high level amplification (+11 copies) in the same PMBL sample, with a VAF consistent with the variant allele amplification in both cases (approximately 88% with 550x depth). As highlighted by previous experiments (Lohr et al., 2012), biological explanation of *PCLO* involvement in lymphomagenesis remains to be determined, but the very focal amplification (312 kb) of these two somatic mutations suggests that *PCLO* should be a candidate for further investigations. A *CCND3* frameshift deletion in the last exon was similarly found involved in a focal multi-copy amplification (+3 copies, 934 kb) in an ABC sample.

DISCUSSION

WES of 14 R/R patients identified a large collection of somatic events, highlighting known and novel targets of interest in DLBCL. The mutational profile of R/R DLBCLs presented herein is somewhat limited by the cohort size, especially in GCBs, which is consistent with the better prognosis reported in this subtype (Roschewski et al., 2014). According to the power analysis published by PathScan developers (Wendl et al., 2011), 10% to 40% power can be achieved in the ABC and PMBL cohorts described herein and 5% to 25% in GCBs, depending on the considered pathway size. While many

Sous presse

false negatives can thus be expected, positive calls are still reliable and proved to be pertinent when compared to previous studies (Pasqualucci et al., 2011; Lohr et al., 2012; Morin et al., 2013; Zhang et al., 2013; de Miranda et al., 2014). The lack of conjointly sequenced cured patient controls is another limitation, which was addressed by applying a comparable approach to two of the aforementioned studies, in order to pinpoint potential genetic resistance mechanisms to immunochemotherapy in our R/R cohort.

While normal contamination of tumoral samples is a well-known caveat in somatic mutation calling, we also identified tumoral contamination of two normal samples (10% and 20% respectively), a phenomenon not reported in previous DLBCL studies, which may lead to missed high VAF mutations if not addressed properly. Even after contamination correction, one ABC sample (GHE0446) exhibited fewer somatic events than other ABC samples. These findings can be linked to the reported exclusion of 6/55 samples in a previous study (Lohr et al., 2012) due to abnormally low mutation counts.

Among the genes and pathways significantly mutated in our cohort, several findings extend the current knowledge of DLBCL. In the ABC subtype, *IRF4* mutations are suggested as an alternative mechanism to enhance its well-established high activity leading to NFκB activation (Yang et al., 2012). Mutations of *GNA13* and relatives (*GNA12*, *GNAQ*) were noticeable in the R/R cohort studied here, and seemed not to be limited to the GCB subtype as has been suggested (Morin et al., 2013).

Of particular interest in this study were patient features that could explain the R/R phenotype. Prognostic markers at diagnosis were not able to delineate most of these 14 patients as prone to relapse, but several somatic events we found may constitute valuable biomarkers. *TBL1XR1* inactivation, until now mainly described in PCNSLs (Bruno et al., 2014), was recently linked to glucocorticoid resistance (Jones et al., 2014), suggesting that the *TBL1XR1* mutations and indels we report may contribute to R-CHOP resistance. *PCLO*, already identified as frequently mutated in DLBCL (Lohr et al., 2012), was found to be simultaneously affected by multiple missense mutations and copy gains. Together with *ADCYAP1*, *CAMK2A*, *ITPR3* and *RYR2* which presented the same pattern, this represents an enrichment in insulin secretion pathway activating double-events. Furthermore, the significant mutation-enrichment identified in the sucrose degradation (ABC) and insulin signaling (PMBL) pathways correlates with these findings to put the focus on hyperglycemia, which was recently described as related to treatment toxicity in NHL (Brunello et al., 2011) and resistance in breast cancer (Zeng et al., 2010). Finally, mutations in the tryptophan degradation pathway, namely *IDO1* and *TDO2*, were found to be R/R specific events. Previous studies have identified tryptophan degradation as a potential immune system escape mechanism of various cancer types, via liberation of metabolites, which induce T and NK cell death (Mellor and Munn, 1999; Frumento et al., 2002; Adams et al., 2012). *IDO1* expression by itself has also been shown to have a negative survival impact in DLBCL (Ninomiya et al., 2011).

Surprisingly, the R/R cohort retrospectively selected based on prognosis was found to be enriched in PMBLs. While unexpected according to the well-established reputation of particularly good prognosis

Under Review

in this subtype (Roschewski et al., 2014), it provided a comprehensive and unprecedented picture of this DLBCL subtype. The five PMBL patients presented consistently higher somatic variant counts than ABC and GCB DLBCLs, although no specific nucleotidic pattern enrichment could be revealed. Mutations in genes involved in IL4 and JAK-STAT signaling (sequence alteration of *SOCS1*, *STAT6*, *IL4R*, amplification of *JAK2*) as well as interferon signaling pathways (*SOCS1*, *PTPN1*, *IRF1/4/8*, *B2M*) were predominant among these somatic events and consistent with previous findings (Rosenwald et al., 2003; Andersson et al., 2010; Rui et al., 2010). We also identified new genes targeted by somatic mutations that were not previously reported in PMBL, including *MFHAS1* and *ITPKB*. *ITPKB* is known to regulate B cell survival and function by interacting with the BCR, and is highly expressed in GCB DLBCLs (Maréchal et al., 2007). *MFHAS1* is a potential oncogene with highly conserved leucine-rich tandem repeats, involved in the chromosomal translocation t(8;14)(p23.1;q21) in B-cell lymphoma cell lines and has been shown to be tumorigenic in nude mice (Tagawa et al., 2004). Altogether these gene mutations, notably *B2M*, *TNFAIP3*, *ITPKB*, *SOCS1* and *BCL7A,* underline the similarity between PMBL and cHL (Reichel et al., 2015).

Mutation and copy-number events in genes such as *XPO1*, *TNFAIP3*, *IKBKB*, *NFKBIA* and *NFKBIE also* provide evidence of NFκB pathway disruption in PMBL, a feature until now considered the hallmark of the ABC subtype (Calado et al., 2010). Whether or not this enrichment is specific to the R/R PMBLs studied here could not be determined due to the unavailability of control PMBL exomes, and should be further investigated. Mutations of *XPO1* (CRM1), whose interaction with *NFKBIA* (IκBα) can lead to the nucleus translocation of NFκB (via REL) (Tam et al., 2000), were confirmed to be recurrent in a larger cohort of PMBLs (Jardin et al. submitted), along with *ITPKB* and *GNA13* (Dubois et al. submitted). Overall, our data indicates that the NFκB pathway is targeted by a wide spectrum of heterogeneous somatic mutations, among the three molecular subtypes (Figure 7). This pinpoints several putative actionable targets but also, on the other hand, potential mechanisms of escaping therapies targeting NFκB pathway components. For instance, the high rate of *IRF4* or *CARD11* somatic mutations may have an impact on lenalidomide efficacy in ABC DLBCL patients (Yang et al., 2012).

In conclusion, whole-exome sequencing of 14 R/R DLBCLs highlights new candidate gene and pathway alterations that enhance our knowledge of lymphomagenesis and offer targeted therapy opportunities. Mutations of *IRF4*, *TBL1XR1* and *GNA13* emerged as particularly interesting leads to explain the R/R phenotype. The high level of heterogeneity observed in mutated genes however underlines the need to extend recurrence analysis to the pathway level, as we performed to pinpoint NFκB and glycemia-related pathways. Assessing whether the gene and pathway mutations pointed out here are isolated incidents or constitutive features of DLBCL is the logical continuation of this work. For this purpose, targeted resequencing in a larger cohort of DLBCL patients has been initiated.

ACKNOWLEDGMENTS

The authors would like to thank Dr Hervé Ghesquières for his review of the manuscript.

Not Published

REFERENCES

Adams S, Braidy N, Bessede A, Bessesde A, Brew BJ, Grant R, Teo C, Guillemin GJ. 2012. The kynurenine pathway in brain tumor pathogenesis. Cancer Res 72:5649–5657.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. 2013. Signatures of mutational processes in human cancer. Nature 500:415–421.

Aloraifi F, McDevitt T, Martiniano R, McGreevy J, McLaughlin R, Egan CM, Cody N, Meany M, Kenny E, Green AJ, Bradley DG, Geraghty JG, Bracken AP. 2015. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. FEBS J 282:3424–3437.

Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, Halgamuge SK. 2014. Inferring copy number and genotype in tumour exome data. BMC Genomics 15:732.

Andersson E, Schain F, Sjöberg J, Björkholm M, Claesson H-E. 2010. Interleukin-13 stimulation of the mediastinal B-cell lymphoma cell line Karpas-1106P induces a phenotype resembling the Hodgkin lymphoma cell line L1236. Exp Hematol 38:116–123.

Bertrand P, Courel M-N, Maingonnat C, Jardin F, Tilly H, Bastard C. 2005. Expression of HYAL2 mRNA, hyaluronan and hyaluronidase in B-cell non-Hodgkin lymphoma: relationship with tumor aggressiveness. Int J Cancer 113:207–212.

Bohers E, Mareschal S, Bertrand P, Viailly PJ, Dubois S, Maingonnat C, Ruminy P, Tilly H, Jardin F. 2015. Activating somatic mutations in diffuse large B-cell lymphomas: lessons from next generation sequencing and key elements in the precision medicine era. Leuk Lymphoma 56:1213–1222.

Brunello A, Kapoor R, Extermann M. 2011. Hyperglycemia during chemotherapy for hematologic and solid tumors is correlated with increased toxicity. Am J Clin Oncol 34:292–296.

Bruno A, Boisselier B, Labreche K, Marie Y, Polivka M, Jouvet A, Adam C, Figarella-Branger D, Miquel C, Eimer S, Houillier C, Soussain C, Mokhtari K, Daveau R, Hoang-Xuan K. 2014. Mutational analysis of primary central nervous system lymphoma. Oncotarget 5:5065–5075.

Burkhard R, Bhagat G, Cogliatti SB, Rossi D, Gaidano G, Pasqualucci L, Novak U. 2015. BCL2 mutation spectrum in B-cell non-Hodgkin lymphomas and patterns associated with evolution of follicular lymphoma. Hematol Oncol 33:23–30.

Calado DP, Zhang B, Srinivasan L, Sasaki Y, Seagal J, Unitt C, Rodig S, Kutok J, Tarakhovsky A, Schmidt-Supprian M, Rajewsky K. 2010. Constitutive canonical NF-κB activation cooperates with disruption of BLIMP1 in the pathogenesis of activated B cell-like diffuse large cell lymphoma. Cancer Cell 18:580–589.

Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R. 2011. Combined genetic inactivation of β2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. Cancer Cell 20:728–740.

Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kähäri AK, Keenan S, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Overduin B, Parker A, Patricio M, Perry E, Pignatelli M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Aken BL, Birney E, Harrow J, Kinsella R, Muffato M, Ruffier M, Searle SMJ, Spudich G, Trevanion SJ, Yates A, Zerbino DR, Flicek P. 2015. Ensembl 2015. Nucleic Acids Res 43:D662–669.

Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, Ding L. 2012. MuSiC: identifying mutational significance in cancer genomes. Genome Res 22:1589–1598.

Delarue R, Tilly H, Mounier N, Petrella T, Salles G, Thieblemont C, Bologna S, Ghesquières H, Hacini M, Fruchart C, Ysebaert L, Fermé C, Casasnovas O, Van Hoof A, Thyss A, Delmer A, Fitoussi O, Molina TJ, Haioun C, Bosly A. 2013. Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): a randomised phase 3 trial. Lancet Oncol 14:525–533.

Deng Y, Wang J, Wang G, Jin Y, Luo X, Xia X, Gong J, Hu J. 2013. p55PIK transcriptionally activated by MZF1 promotes colorectal cancer cell proliferation. BioMed Res Int 2013:868131.

DiDonato JA, Hayakawa M, Rothwarf DM, Zandi E, Karin M. 1997. A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB. Nature 388:548–554.

eLife Preprint

Emmerich F, Theurich S, Hummel M, Haeffker A, Vry MS, Döhner K, Bommert K, Stein H, Dörken B. 2003. Inactivating I kappa B epsilon mutations in Hodgkin/Reed-Sternberg cells. J Pathol 201:413–420.

Etchin J, Sanda T, Mansour MR, Kentsis A, Montero J, Le BT, Christie AL, McCauley D, Rodig SJ, Kauffman M, Shacham S, Stone R, Letai A, Kung AL, Thomas Look A. 2013. KPT-330 inhibitor of CRM1 (XPO1)-mediated nuclear export has selective anti-leukaemic activity in preclinical models of T-cell acute lymphoblastic leukaemia and acute myeloid leukaemia. Br J Haematol 161:117–127.

Fernández-Rodríguez C, Bellosillo B, García-García M, Sánchez-González B, Gimeno E, Vela MC, Serrano S, Besses C, Salar A. 2014. MYD88 (L265P) mutation is an independent prognostic factor for outcome in patients with diffuse large B-cell lymphoma. Leukemia 28:2104–2106.

Fitoussi O, Belhadj K, Mounier N, Parrens M, Tilly H, Salles G, Feugier P, Ferme C, Ysebaert L, Gabarre J, Herbrecht R, Janvier M, Van Den Neste E, Morschhauser F, Casasnovas O, Ghesquieres H, Anglaret B, Brechignac S, Haioun C, Gisselbrecht C. 2011. Survival impact of rituximab combined with ACVBP and upfront consolidation autotransplantation in high-risk diffuse large B-cell lymphoma for GELA. Haematologica 96:1136–1143.

Frumento G, Rotondo R, Tonetti M, Damonte G, Benatti U, Ferrara GB. 2002. Tryptophan-derived catabolites are responsible for inhibition of T and natural killer cell proliferation induced by indoleamine 2,3-dioxygenase. J Exp Med 196:459–468.

Green JA, Suzuki K, Cho B, Willison LD, Palmer D, Allen CDC, Schmidt TH, Xu Y, Proia RL, Coughlin SR, Cyster JG. 2011. The sphingosine 1-phosphate receptor $S1P_2$ maintains the homeostasis of germinal center B cells and promotes niche confinement. Nat Immunol 12:672–680.

Gunawardana J, Chan FC, Telenius A, Woolcock B, Kridel R, Tan KL, Ben-Neriah S, Mottok A, Lim RS, Boyle M, Rogic S, Rimsza LM, Guiter C, Leroy K, Gaulard P, Haioun C, Marra MA, Savage KJ, Connors JM, Shah SP, Gascoyne RD, Steidl C. 2014. Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. Nat Genet 46:329–335.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. Cell 144:646–674.

Jardin F. 2014. Next generation sequencing and the management of diffuse large B-cell lymphoma: from whole exome analysis to targeted therapy. Discov Med 18:51–65.

Jones CL, Bhatla T, Blum R, Wang J, Paugh SW, Wen X, Bourgeois W, Bitterman DS, Raetz EA, Morrison DJ, Teachey DT, Evans WE, Garabedian MJ, Carroll WL. 2014. Loss of TBL1XR1 disrupts glucocorticoid receptor recruitment to chromatin and results in glucocorticoid resistance in a B-lymphoblastic leukemia model. J Biol Chem 289:20502–20515.

Ketterer N, Coiffier B, Thieblemont C, Fermé C, Brière J, Casasnovas O, Bologna S, Christian B, Connerotte T, Récher C, Bordessoule D, Fruchart C, Delarue R, Bonnet C, Morschhauser F, Anglaret B, Soussain C, Fabiani B, Tilly H, Haioun C. 2013. Phase III study of ACVBP versus ACVBP plus rituximab for patients with localized low-risk diffuse large B-cell lymphoma (LNH03-1B). Ann Oncol Off J Eur Soc Med Oncol ESMO 24:1032–1037.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

Lake A, Shield LA, Cordano P, Chui DTY, Osborne J, Crae S, Wilson KS, Tosi S, Knight SJL, Gesk S, Siebert R, Hay RT, Jarrett RF. 2009. Mutations of NFKBIA, encoding IkappaB alpha, are a recurrent finding in classical Hodgkin lymphoma but are not a unifying feature of non-EBV-associated cases. Int J Cancer 125:1334–1342.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505:495–501.

Lee S-H, Hannink M. 2002. Characterization of the nuclear import and export functions of Ikappa B(epsilon). J Biol Chem 277:23358–23366.

Lee H, Yoon SY, Song SB, Park Y, Kim TS, Kim S, Hur DY, Song HK, Park H, Cho D. 2011. Interleukin-18-mediated interferon-gamma secretion is regulated by thymosin beta 4 in human NK cells. Immunobiology 216:1155–1162.

Lenz G, Staudt LM. 2010. Aggressive lymphomas. N Engl J Med 362:1417–1429.

Lenz G, Wright GW, Emre NCT, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W, Powell J, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Campo E, Jaffe ES, Delabie J, Smeland EB, Rimsza LM, Fisher RI, Weisenburger DD, Chan WC, Staudt LM. 2008. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A 105:13520–13525.

Li Y, Gordon MW, Xu-Monette ZY, Visco C, Tzankov A, Zou D, Qiu L, Montes-Moreno S, Dybkaer K, Orazi A, Zu Y, Bhagat G, Richards KL, Hsi ED, Choi WWL, van Krieken JH, Huang Q, Ai W, Ponzoni M, Ferreri AJM, Winter JN, Go RS, Piris MA, Møller MB, Wu L, Wang M, Ramos KS, Medeiros LJ, Young KH. 2013. Single nucleotide variation in the TP53 3' untranslated region in diffuse large B-cell lymphoma treated with rituximab-CHOP: a report from the International DLBCL Rituximab-CHOP Consortium Program. Blood 121:4529–4540.

Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, Novak AJ, Dogan A, Ansell SM, Link BK, Zou L, Gould J, Saksena G, Stransky N, Rangel-Escareño C, Fernandez-Lopez JC, Hidalgo-

Post Print

Miranda A, Melendez-Zajgla J, Hernández-Lemus E, Schwarz-Cruz y Celis A, Imaz-Rosshandler I, Ojesina AI, Jung J, Pedamallu CS, Lander ES, Habermann TM, Cerhan JR, Shipp MA, Getz G, Golub TR. 2012. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A 109:3879–3884.

Maréchal Y, Pesesse X, Jia Y, Pouillon V, Pérez-Morga D, Daniel J, Izui S, Cullen PJ, Leo O, Luo HR, Erneux C, Schurmans S. 2007. Inositol 1,3,4,5-tetrakisphosphate controls proapoptotic Bim gene expression and survival in B cells. Proc Natl Acad Sci U S A 104:13978–13983.

Mellor AL, Munn DH. 1999. Tryptophan catabolism and T-cell tolerance: immunosuppression by starvation? Immunol Today 20:469–473.

Melzner I, Bucur AJ, Brüderlein S, Dorsch K, Hasel C, Barth TFE, Leithäuser F, Möller P. 2005. Biallelic mutation of SOCS-1 impairs JAK2 degradation and sustains phospho-JAK2 action in the MedB-1 mediastinal lymphoma line. Blood 105:2535–2542.

Miller AT, Sandberg M, Huang YH, Young M, Sutton S, Sauer K, Cooke MP. 2007. Production of Ins(1,3,4,5)P4 mediated by the kinase Itpkb inhibits store-operated calcium channels and regulates B cell selection and activation. Nat Immunol 8:514–521.

Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol 12:R112.

de Miranda NFCC, Georgiou K, Chen L, Wu C, Gao Z, Zaravinos A, Lisboa S, Enblad G, Teixeira MR, Zeng Y, Peng R, Pan-Hammarström Q. 2014. Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. Blood 124:2544–2553.

Molina TJ, Canioni D, Copie-Bergman C, Recher C, Brière J, Haioun C, Berger F, Fermé C, Copin M-C, Casasnovas O, Thieblemont C, Petrella T, Leroy K, Salles G, Fabiani B, Morschauser F, Mounier N, Coiffier B, Jardin F, Gaulard P, Jais J-P, Tilly H. 2014. Young patients with non-germinal center B-cell-like diffuse large B-cell lymphoma benefit from intensified chemotherapy with ACVBP plus rituximab compared with CHOP plus rituximab: analysis of data from the Groupe d'Etudes des Lymphomes de l'Adulte/lymphoma study association phase III trial LNH 03-2B. J Clin Oncol Off J Am Soc Clin Oncol 32:3996–4003.

Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJM, Connors JM, Hirst M, Gascoyne RD, Marra MA. 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 476:298–303.

Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, Trinh DL, Bolger-Munro M, Taylor G, Hadj Khodabakhshi A, Ben-Neriah S, Pon J, Meissner B, Woolcock B, Farnoud N, Rogic S, Lim EL, Johnson NA, Shah S, Jones S, Steidl C, Holt R, Birol I, Moore R, Connors JM, Gascoyne RD, Marra MA. 2013. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 122:1256–1265.

Muppidi JR, Schmitz R, Green JA, Xiao W, Larsen AB, Braun SE, An J, Xu Y, Rosenwald A, Ott G, Gascoyne RD, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Vaidehi N, Staudt LM, Cyster JG. 2014. Loss of signalling via Gα13 in germinal centre B-cell-derived lymphoma. Nature 516:254–258.

Ninomiya S, Hara T, Tsurumi H, Hoshi M, Kanemura N, Goto N, Kasahara S, Shimizu M, Ito H, Saito K, Hirose Y, Yamada T, Takahashi T, Seishima M, Takami T, Moriwaki H. 2011. Indoleamine 2,3-dioxygenase in tumor tissue indicates prognosis in patients with diffuse large B-cell lymphoma treated with R-CHOP. Ann Hematol 90:409–416.

Okosun J, Bödör C, Wang J, Araf S, Yang C-Y, Pan C, Boller S, Cittaro D, Bozek M, Iqbal S, Matthews J, Wrench D, Marzec J, Tawana K, Popov N, O'Riain C, O'Shea D, Carlotti E, Davies A, Lawrie CH, Matolcsy A, Calaminici M, Norton A, Byers RJ, Mein C, Stupka E, Lister TA, Lenz G, Montoto S, Gribben JG, Fan Y, Grosschedl R, Chelala C, Fitzgibbon J. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat Genet 46:176–181.

Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, Gaidano G, Mullighan CG, Rabadan R, Dalla-Favera R. 2011. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 43:830–837.

Peyrade F, Jardin F, Thieblemont C, Thyss A, Emile J-F, Castaigne S, Coiffier B, Haioun C, Bologna S, Fitoussi O, Lepeu G, Fruchart C, Bordessoule D, Blanc M, Delarue R, Janvier M, Salles B, André M, Fournier M, Gaulard P, Tilly H, Groupe d'Etude des Lymphomes de l'Adulte (GELA) investigators. 2011. Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma: a multicentre, single-arm, phase 2 trial. Lancet Oncol 12:460–468.

Piao Z, Hong C-S, Jung M-R, Choi C, Park Y-K. 2014. Thymosin β4 induces invasion and migration of human colorectal cancer cells through the ILK/AKT/β-catenin signaling pathway. Biochem Biophys Res Commun 452:858–864.

Pouillon V, Hascakova-Bartova R, Pajak B, Adam E, Bex F, Dewaste V, Van Lint C, Leo O, Erneux C, Schurmans S. 2003. Inositol 1,3,4,5-tetrakisphosphate is essential for T lymphocyte development. Nat Immunol 4:1136–1143.

Not to be published

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JMP, Tubío JMC, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, Valencia A, Himmelbauer H, Bayés M, Heath S, Gut M, Gut I, Estivill X, López-Guillermo A, Puente XS, Campo E, López-Otín C. 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44:47–52.

Récher C, Coiffier B, Haioun C, Molina TJ, Fermé C, Casasnovas O, Thiéblemont C, Bosly A, Laurent G, Morschhauser F, Ghesquières H, Jardin F, Bologna S, Fruchart C, Corront B, Gabarre J, Bonnet C, Janvier M, Canioni D, Jais J-P, Salles G, Tilly H, Groupe d'Etude des Lymphomes de l'Adulte. 2011. Intensified chemotherapy with ACVBP plus rituximab versus standard CHOP plus rituximab for the treatment of diffuse large B-cell lymphoma (LNH03-2B): an open-label randomised phase 3 trial. Lancet Lond Engl 378:1858–1867.

Reichel J, Chadburn A, Rubinstein PG, Giulino-Roth L, Tam W, Liu Y, Gaiolla R, Eng K, Brody J, Inghirami G, Carlo-Stella C, Santoro A, Rahal D, Totonchy J, Elemento O, Cesarman E, Roshal M. 2015. Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. Blood 125:1061–1072.

Ritz O, Guiter C, Castellano F, Dorsch K, Melzner J, Jais J-P, Dubois G, Gaulard P, Möller P, Leroy K. 2009. Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. Blood 114:1236–1242.

Rodríguez D, Ramsay AJ, Quesada V, Garabaya C, Campo E, Freije JMP, López-Otín C. 2013. Functional analysis of sucrase-isomaltase mutations from chronic lymphocytic leukemia patients. Hum Mol Genet 22:2273–2282.

Roschewski M, Staudt LM, Wilson WH. 2014. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nat Rev Clin Oncol 11:12–23.

Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, Chan WC, Zhao T, Haioun C, Greiner TC, Weisenburger DD, Lynch JC, Vose J, Armitage JO, Smeland EB, Kvaloy S, Holte H, Delabie J, Campo E, Montserrat E, Lopez-Guillermo A, Ott G, Muller-Hermelink HK, Connors JM, Braziel R, Grogan TM, Fisher RI, Miller TP, LeBlanc M, Chiorazzi M, Zhao H, Yang L, Powell J, Wilson WH, Jaffe ES, Simon R, Klausner RD, Staudt LM. 2003. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med 198:851–862.

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. 2013. Characterizing and measuring bias in sequence data. Genome Biol 14:R51.

Rui L, Emre NCT, Kruhlak MJ, Chung H-J, Steidl C, Slack G, Wright GW, Lenz G, Ngo VN, Shaffer AL, Xu W, Zhao H, Yang Y, Lamy L, Davis RE, Xiao W, Powell J, Maloney D, Thomas CJ, Möller P, Rosenwald A, Ott G, Muller-Hermelink HK, Savage K, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Weisenburger DD, Chan WC, Gascoyne RD, Levens D, Staudt LM. 2010. Cooperative epigenetic modulation by cancer amplicon genes. Cancer Cell 18:590–605.

Schif B, Lennerz JK, Kohler CW, Bentink S, Kreuz M, Melzner I, Ritz O, Trümper L, Loeffler M, Spang R, Möller P. 2013. SOCS1 mutation subtypes predict divergent outcomes in diffuse large B-Cell lymphoma (DLBCL) patients. Oncotarget 4:35–47.

Schmitz R, Hansmann M-L, Bohle V, Martin-Subero JI, Hartmann S, Mechtersheimer G, Klapper W, Vater I, Giefing M, Gesk S, Stanelle J, Siebert R, Küppers R. 2009. TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. J Exp Med 206:981–989.

Sehn LH, Gascoyne RD. 2015. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. Blood 125:22–32.

Steidl C, Gascoyne RD. 2011. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. Blood 118:2659–2669.

Tagawa H, Karnan S, Kasugai Y, Tuzuki S, Suzuki R, Hosokawa Y, Seto M. 2004. MASL1, a candidate oncogene found in amplification at 8p23.1, is translocated in immunoblastic B-cell lymphoma cell line OCI-LY8. Oncogene 23:2576–2581.

Tam WF, Lee LH, Davis L, Sen R. 2000. Cytoplasmic sequestration of rel proteins by IkappaBalpha requires CRM1-dependent nuclear export. Mol Cell Biol 20:2269–2284.

Tirado CA, Chen W, García R, Kohlman KA, Rao N. 2012. Genomic profiling using array comparative genomic hybridization define distinct subtypes of diffuse large B-cell lymphoma: a review of the literature. J Hematol OncolJ Hematol Oncol 5:54.

Trinh DL, Scott DW, Morin RD, Mendez-Lago M, An J, Jones SJM, Mungall AJ, Zhao Y, Schein J, Steidl C, Connors JM, Gascoyne RD, Marra MA. 2013. Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. Blood 121:3666–3674.

Twa DDW, Chan FC, Ben-Neriah S, Woolcock BW, Mottok A, Tan KL, Slack GW, Gunawardana J, Lim RS, McPherson AW, Kridel R, Telenius A, Scott DW, Savage KJ, Shah SP, Gascoyne RD, Steidl C. 2014. Genomic rearrangements involving programmed death ligands are recurrent in primary mediastinal large B-cell lymphoma. Blood 123:2062–2065.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365:2497–2506.

Ward Y, Gupta S, Jensen P, Wartmann M, Davis RJ, Kelly K. 1994. Control of MAP kinase activation by the mitogen-induced

For Press

threonine/tyrosine phosphatase PAC1. Nature 367:651–654.

Wendl MC, Wallis JW, Lin L, Kandoth C, Mardis ER, Wilson RK, Ding L. 2011. PathScan: a tool for discerning mutational significance in groups of putative cancer genes. Bioinforma Oxf Engl 27:1595–1602.

Weniger MA, Melzner I, Menz CK, Wegener S, Bucur AJ, Dorsch K, Mattfeldt T, Barth TFE, Möller P. 2006. Mutations of the tumor suppressor gene SOCS-1 in classical Hodgkin lymphoma are frequent and associated with nuclear phospho-STAT5 accumulation. Oncogene 25:2679–2684.

Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. 2003. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A 100:9991–9996.

Yang Y, Shaffer AL, Emre NCT, Ceribelli M, Zhang M, Wright G, Xiao W, Powell J, Platig J, Kohlhammer H, Young RM, Zhao H, Yang Y, Xu W, Buggy JJ, Balasubramanian S, Mathews LA, Shinn P, Guha R, Ferrer M, Thomas C, Waldmann TA, Staudt LM. 2012. Exploiting synthetic lethality for the therapy of ABC diffuse large B cell lymphoma. Cancer Cell 21:723–737.

Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, Califano A, Pasqualucci L, Basso K, Dalla-Favera R. 2013. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol 14:1084–1092.

Yoshimura M, Ishizawa J, Ruvolo V, Dilip A, Quintás-Cardama A, McDonnell TJ, Neelapu SS, Kwak LW, Shacham S, Kauffman M, Tabe Y, Yokoo M, Kimura S, Andreeff M, Kojima K. 2014. Induction of p53-mediated transcription and apoptosis by exportin-1 (XPO1) inhibition in mantle cell lymphoma. Cancer Sci 105:795–801.

Zeng L, Biernacka KM, Holly JMP, Jarrett C, Morrison AA, Morgan A, Winters ZE, Foulstone EJ, Shield JP, Perks CM. 2010. Hyperglycaemia confers resistance to chemotherapy on breast cancer cells: the role of fatty acid synthase. Endocr Relat Cancer 17:539–551.

Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, Bernal-Mizrachi L, Mann KP, Flowers C, Naresh K, Evens A, Gordon LI, Czader M, Gill JI, Hsi ED, Liu Q, Fan A, Walsh K, Jima D, Smith LL, Johnson AJ, Byrd JC, Luftig MA, Ni T, Zhu J, Chadburn A, Levy S, Dunson D, Dave SS. 2013. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A 110:1398–1403.

eLife's presse

TABLES

Table 1 - Clinical features of the R/R cohort

Main clinical features are compared between the 14 R/R patients that were sequenced and the 61 other patients of the considered series. The only significant difference concerns the molecular subtype (p=0.03), with an enrichment in PMBLs in the R/R cohort.

	R/R cohort n=14		Other n=61		p
Age (years)	63 [29-82]		56 [19-84]		0.62[a]
Follow-up	16.5 [7.1-46.3]		33.1 [0.1-81.3]		–
Bulky mass ≥ 10cm	2	14%	8	13%	1[b]
IPI					
0-2	6	43%	38	62%	0.23[b]
3-5	8	57%	23	38%	
Age Ajusted IPI					
0-1	5	36%	32	52%	
2	6	43%	23	38%	0.37[c]
3	3	21%	6	10%	
Ann-Harbor stage					
I	2	14%	8	13%	
II	2	14%	11	18%	0.72[c]
III	1	7%	11	18%	
IV	9	64%	31	51%	
Molecular subtype					
ABC	6	43%	25	41%	
GCB	3	21%	24	39%	0.03[c]
PMBL	5	36%	5	8%	
other	0	0%	7	11%	
Treatment					
R-CHOP	10	71%	24	39%	
R-mCHOP	1	7%	6	10%	0.09[c]
R-ACVBP	3	21%	31	51%	

[a] Mann-Whitney test
[b] Fisher's exact test
[c] Chi-squared test

sous presse

Table 2 - Somatic event filtering

Considered and validated events for each filter are presented in this table, along with the tool implementing it and a short description. Pre-filtering was performed separately on indels and SNVs, while the main filters were common to both event types.

Tool	Considered	Passed	Criterion
SNV pre-filtering			
VarScan somaticFilter	1 433 502	78 314	SNVs with somatic p value of 5% or less
VarScan somaticFilter	78 314	77 994	Not a SNV in a group of 3+ in a 21 bp wide window
VarScan somaticFilter	77 994	74 216	Not a SNV closer than 4 bp to an indel
Indel pre-filtering			
custom	159 841	8 806	Indels with somatic p value of 5% or less
SNV + Indel filtering			
custom	11 545	11 307	Not a (potentially rearranged) BCR VDJ element
custom	11 545	11 500	Not a (potentially rearranged) TCR VDJ element
VEP	11 545	3 116	Considered as one (or many) from the following :
		7	*- mature_miRNA_variant*
		9	*- coding_sequence_variant*
		3	*- incomplete_terminal_codon_variant*
		11	*- initiator_codon_variant*
		2 325	*- missense_variant*
		38	*- splice_acceptor_variant*
		50	*- splice_donor_variant*
		401	*- splice_region_variant*
		136	*- stop_gained*
		9	*- stop_lost*
		87	*- TF_binding_site_variant*
		134	*- frameshift_variant*
		29	*- inframe_deletion*
		18	*- inframe_insertion*
PolyPhen, SIFT	11 545	11 421	Not considered “benign“ by PolyPhen **and** “tolerated“ by SIFT
custom	11 545	11 245	At least 4 variant-supporting reads retained
custom	11 545	9 981	Variant supporting reads have not a significantly lower mapping quality than reference supporting reads (Mann-Whitney's test, p>1%)
custom	11 545	7 261	Somatic p value is lower than the mappability-based threshold (more stringency in repeated regions, Supplemental Methods)
custom	11 545	11 018	Variant allele found on both strands in tumor, or in a desequilibrium non-significantly distinct from the reference allele (Fisher's exact tests, p>5%)
	11 545	**1 725**	**Passing all filters**

sous presse

Table 3 - Minimal common altered regions in the R/R cohort

Recurrent copy gains, copy losses and copy-neutral heterozygosity losses are listed along with altered genes that could explain the enrichment observed according to the literature. Overlaps between at least 3 patients with compatible copy states harboring less than 20 genes were considered, as well as overlaps between 2 highly similar alterations.

Chrom	Start	End	Genes of interest	Alteration
1	16 952 953	17 046 612		LOH
2	61 000 000	61 149 677	*REL*	+
2	133 010 000	133 020 723	*CDC27P1*	LOH
3	10 088 267	10 089 851	*FANCD2, IRAK2*	LOH
3	47 364 266	50 649 088	*HYAL2*	–
4	955 469	1 388 000	*CRIPAK*	–
4	68 264 884	68 266 018		LOH
5	110 436 281	110 443 145	*WDR36*	+
6	31 496 785	31 515 798	*NFKBIL1*	–
6	137 700 000	138 202 463	*TNFAIP3*	–
7	38 313 181	38 356 944	*TRG*	–
7	142 104 421	142 480 465	*TRB*	–
9	5 520 000	5 557 759	*PDCD1LG2*	+
9	21 968 717	22 029 523	*CDKN2A*	–
11	48 346 514	48 388 898	*OR4C4P*	LOH
14	19 600 000	22 900 000	*TRA*	–
14	106 329 454	106 388 513	*IGH*	–
15	43 600 000	45 879 730	*TP53BP1, B2M*	–
16	28 850 000	28 980 000	*CD19*	–
17	7 187 050	7 750 341	*TP53*	–
19	59 000 000	59 082 763	*MZF1*	+
21	10 906 898	11 098 625	*TPTE*	–
22	22 730 788	23 040 766	*IGL*	–

Post presse

FIGURES

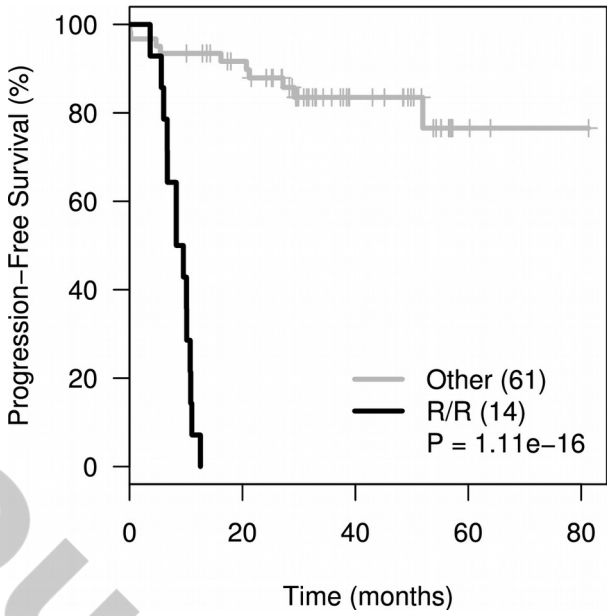


Figure 1 - Progression-Free Survival of the R/R cohort. Kaplan-Meier curves comparing the 14 R/R patients selected for sequencing (relapse or progression under treatment in less than a year) to the rest of the available cohort.

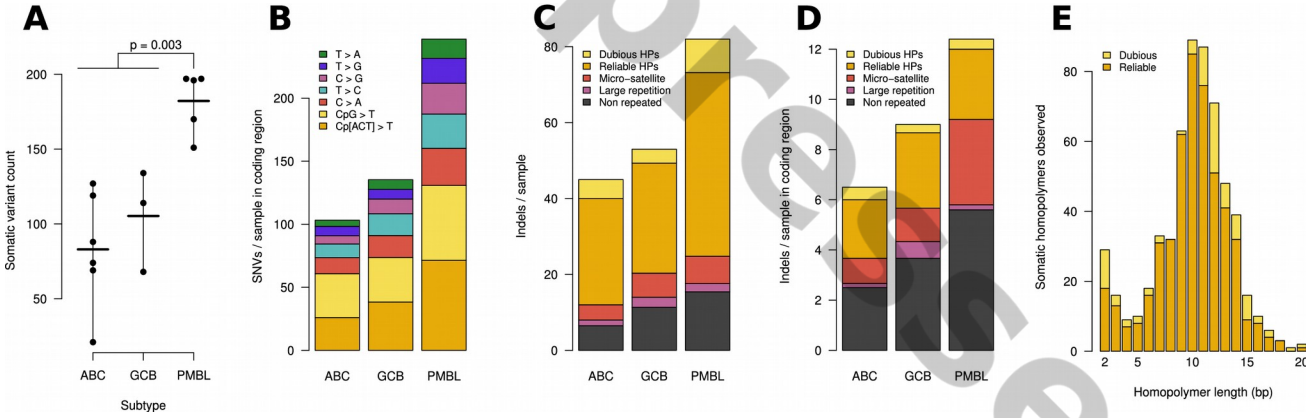


Figure 2 - Molecular characterization of somatic events. **(A)** Sums of SNVs and indels are represented for every patient (black dots), and grouped by subtypes. Horizontal bars represent the mean number of somatic events in each subtype. **(B)** Somatic SNVs are categorized according to the normal and tumoral nucleotides, and presented averaged by subtype. In **(C)** and **(D)**, somatic indels are categorized according to the homopolymer category they fall in (partial gains and losses of a monomer are considered as non-repeated events). **(C)** describes all somatic events detected, **(D)** only the events in the coding region (ROI). **(E)** Somatic homopolymer alteration counts according to the homopolymer lengths. Events evidencing significantly more mutated reads than expected by the sole sequencing error-rate are classified as “reliable”, the others as “dubious”.

sous presse

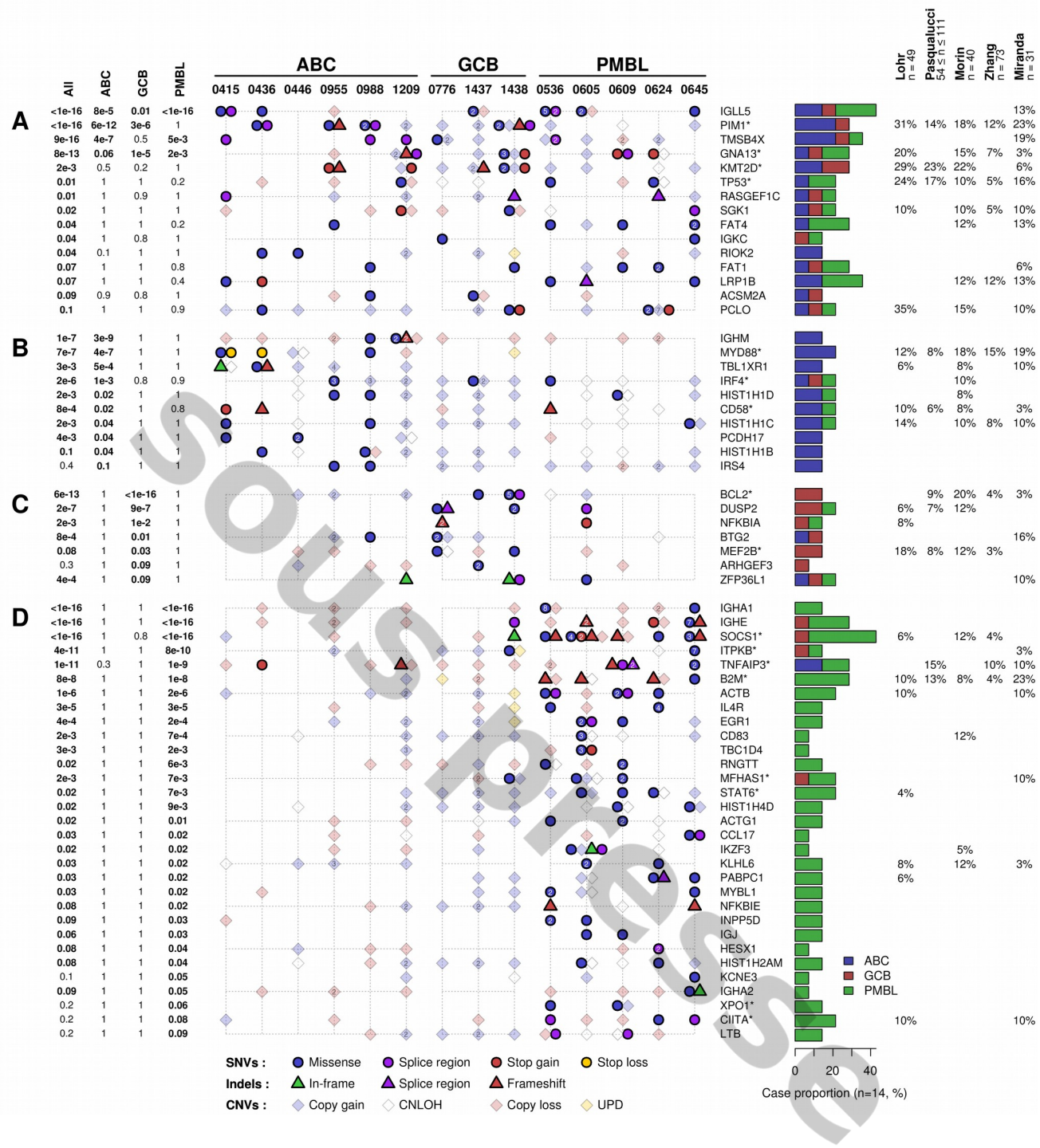


Figure 3 - Genes significantly mutated in the R/R cohort. Individual somatic events are reported as colored symbols, with an optional white number signaling multiple events of the same category. Genes are listed in rows, grouped by the DLBCL subtype they were found significant in (all patients for panel **A**, ABC for panel **B**, GCB for panel **C** and PMBL for panel **D**) and ordered by descending significance. The 14 R/R patients are described in distinct columns, grouped by their respective DLBCL subtype. The left panel presents the FDR of each gene in each subtype as computed by MuSiC SMG in separate analyses (Supplemental Methods), considering FDR<10% as significant (bold font). The right panel visually summarizes the mutated patient count for each gene and each subtype, and compares it to the proportions of mutated samples described in various whole-exome and whole-genome DLBCL studies. While CNV, LOH and UPD events are presented with diamonds and lighter colors, they were not considered in the statistical analysis.

Solis Press

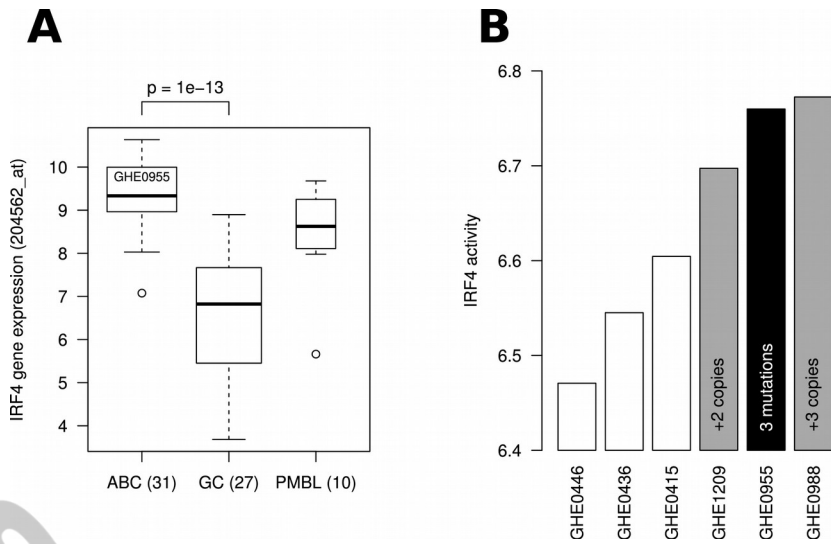


Figure 4 - Transcriptional characterization of the *IRF4* mutant. (A) The *IRF4* gene expression of the highly mutated ABC patient GHE0955 (9.78) is represented in the context of the whole series of DLBCL patients. **(B)** The activity of IRF4, measured by the mean expression of the 258 genes it is reported to upregulate, is presented for the 6 R/R ABC patients sequenced by WES. Patients with *IRF4* copy number gains are colored in gray, while the only ABC patient found mutated in *IRF4* is highlighted in black.

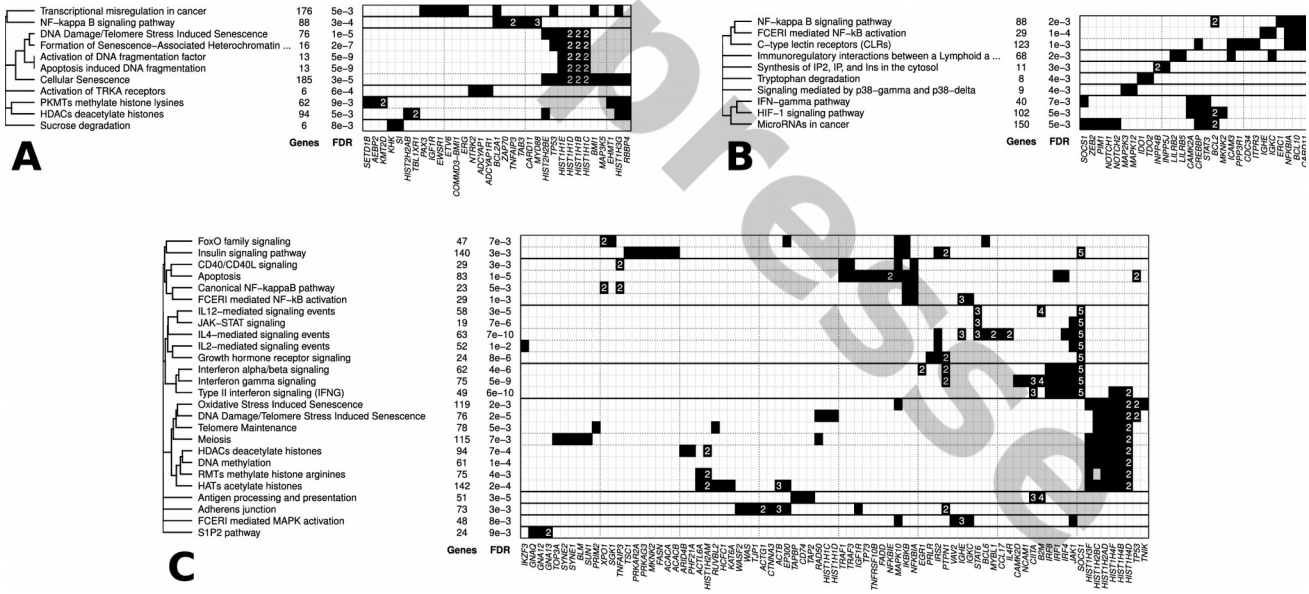
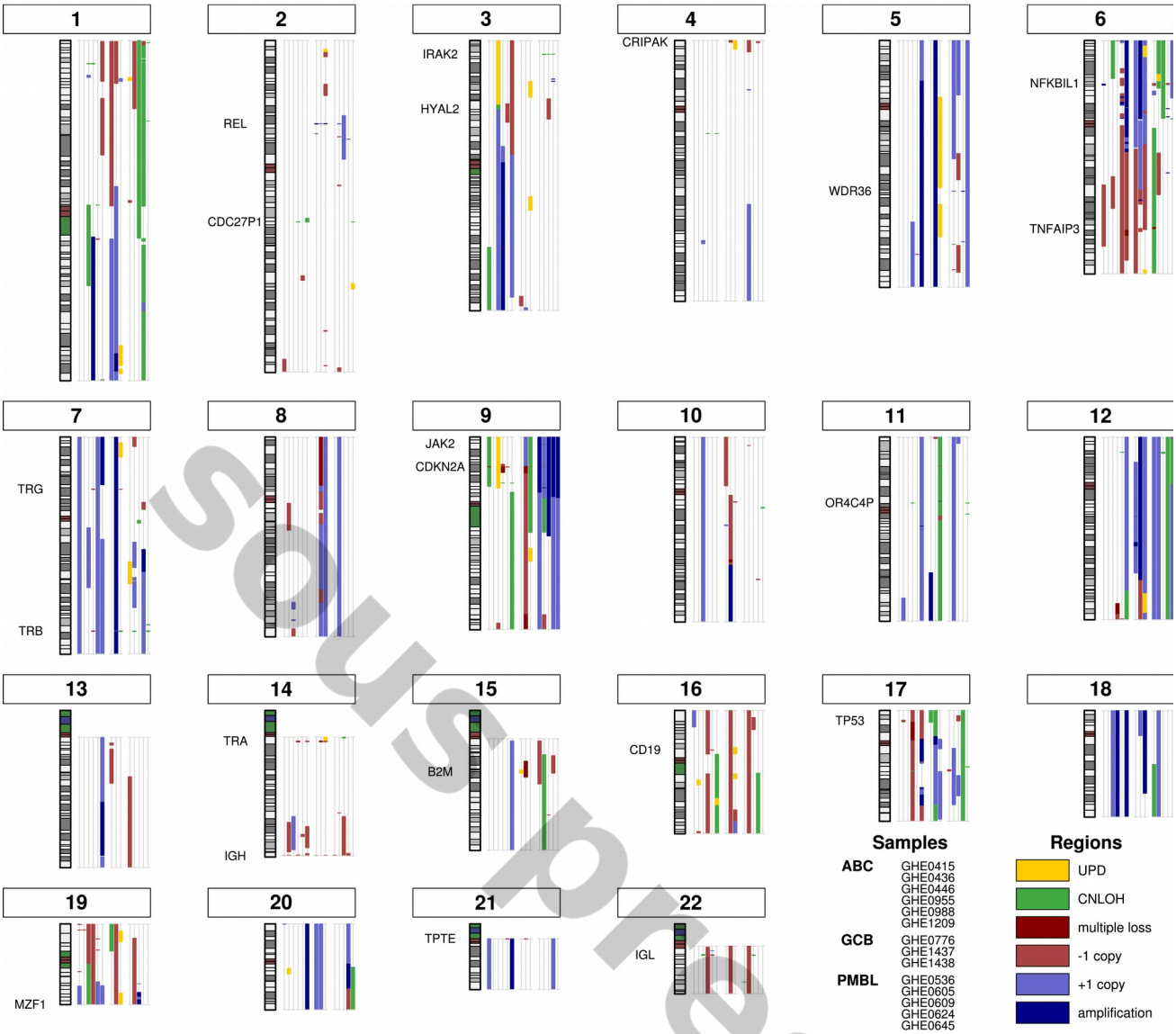


Figure 5 - Pathways significantly mutated in the R/R cohort. Pathways harboring more mutations than expected by the sole background mutation rate are presented in rows, with the corresponding mutated genes in columns. A black cell represents one or many mutations in the same patient, while a white number signals mutations in multiple patients. Hierarchical clustering was performed to identify clusters of pathways relying on the same mutations, with two arbitrary levels of decreasing similarity materialized by plain and dotted horizontal lines. The analysis was performed separately in the 3 DLBCL subtypes (ABC in panel **A**, GCB in panel **B** and PMBL in panel **C**). Pathways presented here were manually curated for relevance, the full result set (FDR<1% according to MuSiC PathScan) can be found in Supplemental Figure 4.

Soumis pour publication



sous presse

Figure 6 - Copy Number Variations in the R/R cohort. CNV and LOH events are presented as colored bands for the 14 R/R patients, grouped by chromosome and by DLBCL subtype (the figure keys are available in the bottom right corner). The assessed regions are materialized with vertical grey bars for each sample, and allosomes were not analyzed. Giemsa banding as provided by the UCSC is represented on the left of each chromosome, with centromeric regions colored in red.

Voir Figure 5.2 du manuscrit de thèse

Figure 7 - Somatic events found in the NFκB pathway. Proteins involved in the NFκB signaling pathway according to the Ingenuity Knowledge Base (Qiagen, Redwood City, CA) are represented with their published interactions. Genes found significantly mutated in one or many DLBCL subtypes are highlighted in gold (FDR < 10%), while other mutated genes are reported in light yellow. For both, the mutated patient counts are indicated by multiple stars in the protein name. Genes recurrently amplified are circled in blue, while genes recurrently deleted are circled in red. Details on altered protein families and complexes are provided in the bottom right corner.

Whole Exome Sequencing of Relapsed / Refractory Patients Expands the Repertoire of Somatic Mutations in Diffuse Large B-Cell Lymphoma

Mareschal et al.

Summary

Supplemental Methods....1
Gene-Expression Profiling (continued)....1
WES secondary analysis....2
Sample purity estimation in WES data....2
Variant calling in WES data....2
Mappability-based variant filter....3
Accuracy estimation by resequencing....3
Homopolymer indel assessment....4
Enrichment analyses....4
Pathway-enrichment comparative study....5
Copy number and heterozygosity analyses....5
Mutated Allele Selection events....6
Supplemental References....7
Supplemental Figures....8
Figure S1 - Whole-Exome Sequencing analysis pipeline....8
Figure S2 - VAF-based sample purity estimation....9
Figure S3 - Molecular signatures in the R/R cohort....10
Figure S4 - Pathways significantly mutated in the R/R cohort....12
Figure S5 - Copy number models....13
Figure S6 - Complete copy number analysis results....15
Supplemental Tables....15
Table S1 - Detailed clinical features of the R/R cohort....15
Table S2 - Somatic event orthogonal validation....16
Supplemental Datasets....16
Dataset S1 - Quality assessment of the Whole-Exome Sequencing....16
Dataset S2 - Filtered somatic variant list....16
Dataset S3 - Pathways enriched in mutations in the R/R cohort....16
Dataset S4 - Comparative enrichment analysis of significant pathways....16
Dataset S5 - Significant Mutated Allele Selection events in the R/R cohort....16

In Press

Supplemental Methods

Gene-Expression Profiling (continued)

Gene expression of single genes was represented by the average expression of a single probeset, singled out as the most representative of the gene (“best”) by the JetSet algorithm version 3.0.0 [1]. The activity of the *IRF4* transcription factor in ABC samples was estimated by the mean expression of the genes previously reported as down-regulated after *IRF4* knock-down [2]. Of the 308 genes originally described, 258 were available on the Affymetrix U133+2 array and were represented by their best probeset according to the JetSet.

WES secondary analysis

The secondary analysis (from FASTQ files to analysis-ready BAM files) was managed by a custom bioinformatics pipeline (Figure S1). Reads were first aligned to the GRCh37 assembly of the human genome using the MEM algorithm from the Burrows-Wheeler Aligner package (0.7.5) [3], discarding reads with mapping qualities lower than 20. PCR and optical duplicates were filtered out using Picard's markDuplicate tool (1.98) [4]. Local realignment of reads around indels and Base Quality Score Recalibration (BQSR) were performed by the Genome Analysis ToolKit (2.7-2) [5]. Only pairs of resulting reads were retained, using the fixmate command from SAMtools (0.1.19) [6].

Sample purity estimation in WES data

Normal cell contamination of tumoral samples (1 - tumor purity P_{tum}) and tumoral cell contamination of normal samples (1 - normal purity P_{nrm}) were estimated by the Variant Allele Frequency (VAF) modelization in paired normal and tumoral samples. Only positions covered at 80x or more with a normal / tumoral VAF difference of 10% or more were considered to fit the model. Retained positions were clustered by minimizing the euclidean distance to the following states: somatic homozygous or heterozygous mutation, Loss of Heterozygosity (LOH) preserving the reference or variant allele, and Single Nucleotide Polymorphism (SNP). The last state was defined as the $VAF_{tumor} = VAF_{normal}$ line rather than a single point to better fit the observed data. The two parameters of the model P_{nrm} and P_{tum} were manually tuned for each sample to minimize intra-cluster distances.

Variant calling in WES data

Normal / tumoral comparative variant calling was performed by the “somatic” algorithm of VarScan (2.3.7*) [7], after piling up nucleotides with base and mapping qualities of 20 or more. Minimal VAF was set to 10% as recommended with VarScan for a depth of about 100x [8]. Normal and tumor purities were set to their estimates for each sample, as described above. Resulting Single Nucleotide Variations (SNVs) and short insertions / deletions (indels) were annotated by Ensembl's Variant Effect Predictor (VEP version 79) [9], which notably embeds the SIFT [10] and PolyPhen 2 [11] pathogenicity predictors. The exhaustive list of post-calling criteria applied to select variants and the number of variants impacted are described in Table 2. All transcripts of all genes were considered during the filtering steps, thus variants meeting the above mentioned criteria for at least one transcript were retained. This approach defines a Region of Interest (ROI) of approximately 38 Mb, consisting of all coding exons (Ensembl genes GRCh37 79 [12]) extended by 8 bp on each side (the “splice region”, as considered by VEP).

Theoretical calling power according to the sequencing depth was computed considering that at least 4 mutated reads are required by VarScan. Desirable sequencing depth (D),

sous presse

depending on the tumor cellularity, was defined as `D = 4 / (cellularity * 0.5)` and ranged from 9x to 40x (Dataset S1). The proportion of the ROI covered by at least D reads was then computed for each sample (Dataset S1), and averaged at 86.8% [75-93].

VarScan “somatic” command was slightly modified to avoid over-filtering of variants in tumor-contaminated normal samples. The following replacement was operated in “Somatic.java” (lines 1799 to 1809) to better handle the highly contaminated normal samples, and VarScan was entirely recompiled using Java 5 compatibility mode:

```
if(normalFreq > normalMinVarFreq + 0.05) {
        somaticStatus = "Germline";
} else if(freqDiff >= 0.25 && diffPvalue < somaticPvalue / 10 && tumorFreq > normalFreq) {
        somaticStatus = "Somatic";
} else if(freqDiff <= -0.25 && diffPvalue < somaticPvalue / 10 && tumorFreq < normalFreq) {
        somaticStatus = "LOH";
```

Mappability-based variant filter

As low complexity regions and homologous genes lead to local surges in wrongly aligned reads and spurious mutations, we introduced a filter based on the mappability score to enforce more stringency in these regions. The genomic distribution of this score, based on the number of times a 24-mer can be mapped in the whole genome allowing up to two mutations [13], was downloaded from the University of California Santa Cruz (UCSC) data bank for the hg19 assembly. For each variant, the mappability value (Table S4) was computed as the mappability average over the variant position extended by 25 bp in both directions (from 1 for a perfectly unique 24-mer toward 0 for very low-complexity regions). The threshold applied to VarScan somatic p value was subsequently computed using the following formula :

$$threshold_{variant} = 0.05 * mappability_{variant}^{2}$$

Accuracy estimation by resequencing

Twelve R/R patients were part of a larger cohort for whom a selection of 34 genes was re-sequenced, as described elsewhere [14]. Briefly, 272 loci defining an 87.7 kb target were selected based on the results described herein and in previous studies. The sequencing was performed using the AmpliSeq™ technology on a Personal Genome Machine (PGM; Life Technologies, Carlsbad, CA), with a median depth of 225x. The whole analysis was handled by the Torrent Suite (4.2.1) provided by the manufacturer, extended with custom variant annotation and filtering.

Only positions with 20-fold sequencing depth in PGM and 8-fold in both normal and tumoral WES datasets were considered for variant comparison, representing 91% [89.5-93.4] of the 87 703 nucleotides of the PGM target. Ninety-six SNVs and short indels were

PLOS Press

thus detected by both WES and PGM, while 16 were exclusive to WES analysis and 64 exclusive to PGM analysis (Table S2). All of the 112 variants found by WES were confirmed as true positives after manual examination, while 16 of the 128 true variants were missed by WES due to insufficient sequencing depth (under the 4 variant-read threshold set by VarScan), low VAF observed in tumoral exome or lack of power in the analysis. Binomial 95% confidence intervals for WES precision and sensitivity can be estimated from these results as [97-100]% and [80-93]% respectively.

Homopolymer indel assessment

Tests were conducted to assess whether indels in homopolymer stretches were due to the documented bias of the Illumina sequencing technology [15, 16] or genuine somatic events. Such tests were performed for each indel consisting in the gain / loss of one or many monomers in a series of 2 repetitions or more of the same nucleotide. An indel was considered as “reliable” if the variant-supporting read count in the tumoral sample was significantly higher than what could be expected from the sole error-rate (FDR<0.001), while no such conclusion could be drawn in the matching normal sample (FDR>0.1). Unilateral binomial tests were used for this purpose (R software version 3.0.2), comparing to the error rates reported in indels of the same type (insertion or deletion) and size with the same sequencing technology [16]. For the two patients with evidence of tumoral contamination of the normal sample, error rates were augmented to account for the presence of heterozygously mutated cells in according proportions.

Enrichment analyses

Both gene and pathway-enrichment analyses were performed using tools from the MuSiC software suite (0.4.1-1) [17].

Gene enrichment in mutations (shorts indels and SNVs) over the ROI was performed using the Significantly Mutated Gene (SMG) tool, which highlights genes with more mutations than expected according to the background mutation rate [17]. Briefly, the somatic mutation rate is computed for each gene and for 7 mutation classes (AT / CG / CpG transitions / transversions and indels), dividing the corresponding mutation count in the studied group by the sum of the sequenced region sizes across the samples. These rates are then compared to the mean mutation rates observed throughout the whole exomes using 3 distinct tests: any combination of two tests yelding a False Discovery Rate (FDR) < 10% was considered significant (MuSiC SMG documentation). Gene enrichment analyses were conducted separately in each subtype, and in the whole cohort (“all”). In the latter case, 3 groups with comparable mutation rates were defined by the algorithm in an unsupervised fashion, separating the highly mutated PMBLs and the sparsely mutated GHE0446 from the others.

Sous presse

Pathway enrichment was assessed with the PathScan tool, which relies on the same concept of mutation rate and accounts for the sequenced length of the genes [18]. Tested gene sets consisted of 2 632 human-specific and conserved pathways retained from the NCBI BioSystems database [19]. Only pathways involving 5 to 500 ROI genes were retained, as many conserved pathways have no or few known members in the human genome. A FDR threshold of 1% was applied to consider pathways as significantly enriched in mutations. Resulting pathways were ordered by hierarchical clustering of the mutated genes (using the Jacard index as distance and the Ward algorithm for agglomeration), with arbitrary cut-offs set at heights of 0.6 and 0.9 to identify clusters of pathways relying on the same mutations.

Pathway-enrichment comparative study

To highlight pathways particularly enriched in our R/R cohort, the selected pathways were assessed in two external series of DLBCLs analyzed with similar technologies [20, 21]. Genomic coordinates of the variants originally described were converted from the hg18 human genome assembly to GRCh37 using the UCSC liftOver tool [22]. Variants were annotated and filtered in the same way as R/R data, discarding intronic or intergenic variants. Pathways selected by PathScan in our cohort were tested independently in the three cohorts, using a simplified application of PathScan concepts.

$$R_{ABC,P} = \frac{\sum_{s \in ABC} \sum_{g \in P} n_{s,g}}{|ABC| \times \sum_{g \in P} |g|}$$

The mutation rate in a pathway “P” involving multiple genes “g” to be assessed in a specific subtype (e.g. ABC) represented by multiple samples “s” was computed according to the formula presented above, as the average “n” somatic mutation count per sequenced bp. A pathway was considered as enriched in mutations if its mutation rate was found higher than the similar rate computed on the whole ROI (unilateral λ test based on the Poisson distribution, FDR<10%, R software version 3.0.2). As only 322 genes were described by Zhang et al [21], two analyses were performed: considering the remaining genes as never mutated (overly optimistic), or comparing to the mutation rate observed in the 322 described genes only (overly pessimistic).

Copy number and heterozygosity analyses

Regions affected by somatic copy number gain, loss or Copy-Neutral Loss of Heterozygosity (CN-LOH) were identified from WES data using the ADTEx [23] software (version 2.0). Depth of Coverage (DOC) was recorded for each exon in the ROI. B Allele Frequency (BAF) was computed for each position with DOC ≥ 10 in both normal and tumoral samples showing a heterozygous state in the normal sample (≥2 B allele

Nous presse

supporting reads, BAF ≥ 25%), considering the second most frequent allele as the “B” allele. Ploidy was manually fixed to 2 for all samples after visual assessment of BAF plots. Resulting regions were manually curated by browsing the normalized DOC and BAF plots with Rgb [24]. Copy calling models were tuned by visual assessment of DOC ratio distribution across the genome of each patient, estimating these distributions by kernel-based densities of segment mean DOCs weighted by the corresponding segment lengths (Figure S5). The most frequent copy state was considered as wild-type, and copy states were supposed to be separated by sample-specific DOC shifts linked to tumor purity.

Minimal Common Regions (MCR) were defined using Rgb [24], focusing on overlaps of at least 3 regions of compatible copy states harboring less than 20 genes. Overlaps of 2 highly similar alterations were also considered as MCRs.

Regions suggestive of uniparental isodisomy (UPD) were identified from the same BAF measurements generated for ADTEx. Bases in situation of heterozygosity were counted (n) in non-overlapping one Mb-wide windows (w) across samples (s from S) and chromosomes, and transformed into log-ratios (LR) according to the following formula:

$$LR_{w,s} = \log_2\left(\frac{n_{w,s}+1}{Ref_w+1}\right)$$

$$Ref_w = \frac{1}{|S|} \times \sum_{s' \in S} n_{w,s'}$$

Resulting log-ratios were segmented using the Circular Binary Segmentation algorithm [25], enforcing a minimal segment width of 3 consecutive points and using window-specific weights (W) described in the following formula. Segments with a mean log-ratio of -1 or less were considered as UPD.

$$W_w = \log\left(Ref_w + 1\right) + 1$$

Mutated Allele Selection events

Somatic mutations in regions also harboring CNV or CNLOH with evidence of variant allele enrichment were listed as “Mutated Allele Selection” events (Dataset S5). For each mutation located in such a region, a unilateral binomial test was performed to determine if the VAF observed was significantly higher than what could be expected from an heterozygous mutation with the estimated tumor purity (FDR<1%).

Sous presse

Supplemental References

1. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. BMC Bioinformatics. 2011; 12: 474.
2. Shaffer AL, Emre NCT, Lamy L, Ngo VN, Wright G, Xiao W, Powell J, Dave S, Yu X, Zhao H, Zeng Y, Chen B, Epstein J, et al. IRF4 addiction in multiple myeloma. Nature. 2008; 454: 226–231.
3. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;
4. Picard Tools - By Broad Institute. n.d.;
5. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 2011; 43: 491–498.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 2009; 25: 2078–2079.
7. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22: 568–576.
8. Stead LF, Sutton KM, Taylor GR, Quirke P, Rabbitts P. Accurately identifying low-allelic fraction variants in single samples with next-generation sequencing: applications in tumor subclone resolution. Hum. Mutat. 2013; 34: 1432–1438.
9. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinforma. Oxf. Engl. 2010; 26: 2069–2070.
10. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 2009; 4: 1073–1081.
11. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat. Methods. 2010; 7: 248–249.
12. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, et al. Ensembl 2015. Nucleic Acids Res. 2015; 43: D662–669.
13. Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. Fast computation and applications of genome mappability. PloS One. 2012; 7: e30377.
14. Dubois S, Mareschal S, Picquenot J-M, Viailly P-J, Bohers E, Cornic M, Bertrand P, Veresezan EL, Ruminy P, Maingonnat C, Marchand V, Lanic H, Penther D, et al. Immunohistochemical and genomic profiles of diffuse large B-cell lymphomas: Implications for targeted EZH2 inhibitor therapy? Oncotarget. 2015;
15. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol. 2011; 12: R112.
16. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. Characterizing and measuring bias in sequence data. Genome Biol. 2013; 14: R51.
17. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, Ding L. MuSiC: identifying mutational significance in cancer genomes. Genome Res. 2012; 22: 1589–1598.
18. Wendl MC, Wallis JW, Lin L, Kandoth C, Mardis ER, Wilson RK, Ding L. PathScan: a tool for discerning mutational significance in groups of putative cancer genes. Bioinforma. Oxf. Engl. 2011; 27: 1595–1602.
19. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI BioSystems database. Nucleic Acids Res. 2010; 38: D492–496.
20. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood. 2013; 122: 1256–1265.
21. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. 2013; 110: 1398–1403.
22. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34: D590–598.
23. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, Halgamuge SK. Inferring copy number and genotype in tumour exome data. BMC Genomics. 2014; 15: 732.
24. Mareschal S, Dubois S, Lecroq T, Jardin F. Rgb: a scriptable genome browser for R. Bioinforma. Oxf. Engl. 2014; 30: 2204–2205.
25. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinforma. Oxf. Engl. 2007; 23: 657–663.

In Press

Supplemental Figures

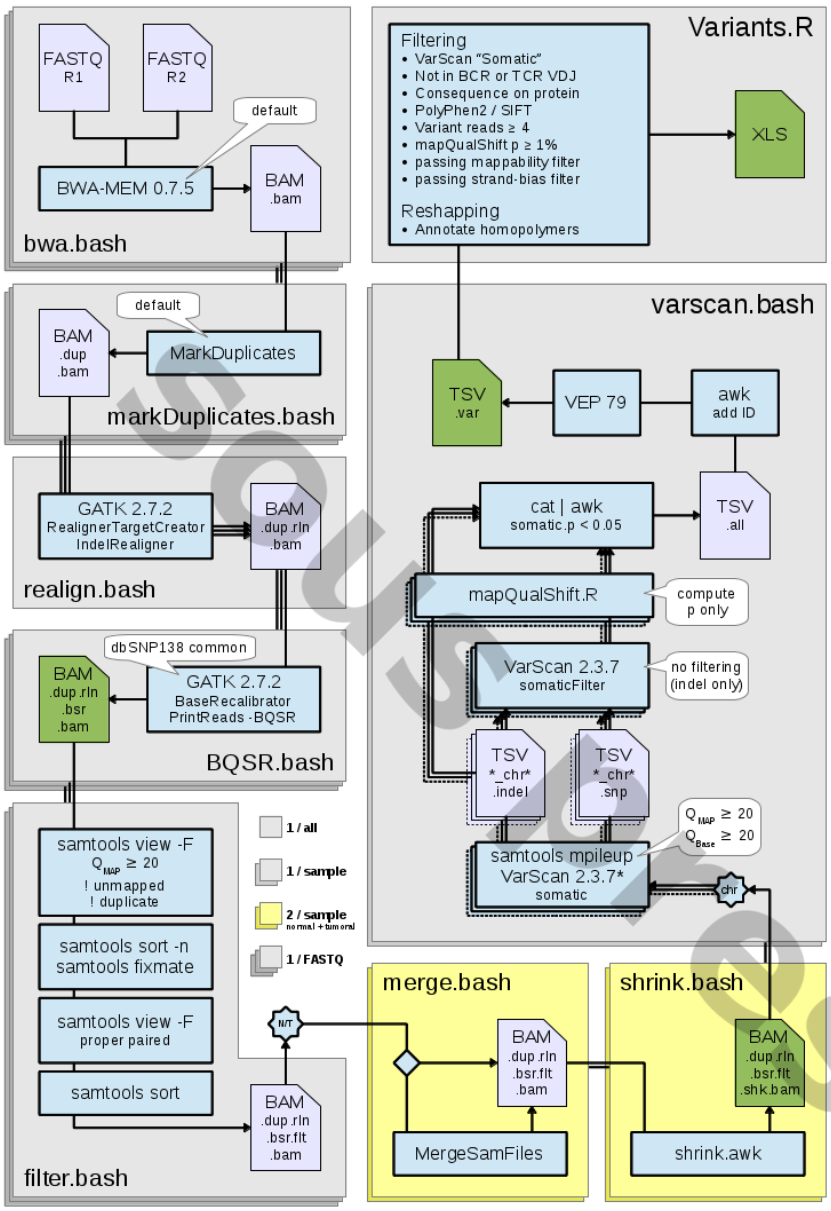


Figure S1 - Whole-Exome Sequencing analysis pipeline

Schematic of the bioinformatics pipelines developed to perform WES analysis. Box thickness represents various levels of execution (figure key in the bottom left corner): scripts are to be ran on each FASTQ bundle (possibly many for a single sample), on each sample (normal and tumoral from a single patient), on each patient or on the whole series. Files colored in gray are temporary files that are removed once the step completed, green files are preserved for subsequent analyses and quality checks. White balloons give details on the parameters used for each tool.

Soumis à Elsevier

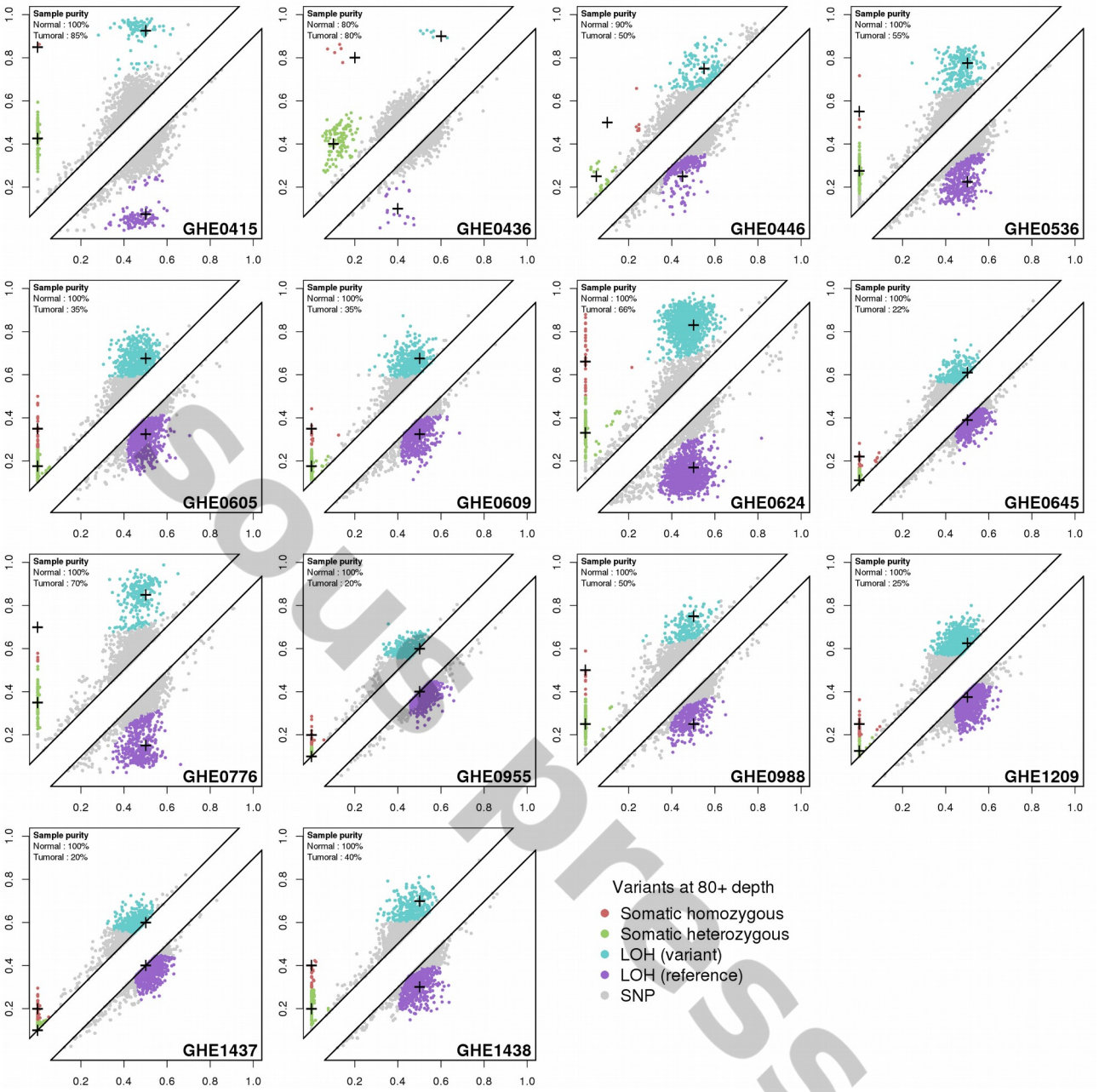


Figure S2 - VAF-based sample purity estimation.

For each patient, genomic positions covered by at least 80 reads in both normal and tumoral samples are represented as a scatter plot, with the normal and tumoral VAF on the X and Y axes respectively. A 5 state model was manually fitted on each plot (figure keys in the bottom right corner), in order to estimate the normal contamination of the tumoral samples and the tumoral contamination of the normal samples. Black crosses represent the modelized center of each state, while dots are clustered to the nearest center according to Euclidean distance (the “SNP” state is modelized by the X=Y line rather than a single point). The diagonal, corresponding to uninformative positions with a VAF difference between normal and tumoral samples lesser than 10%, was discarded to save computing time.

Soumis pour publication

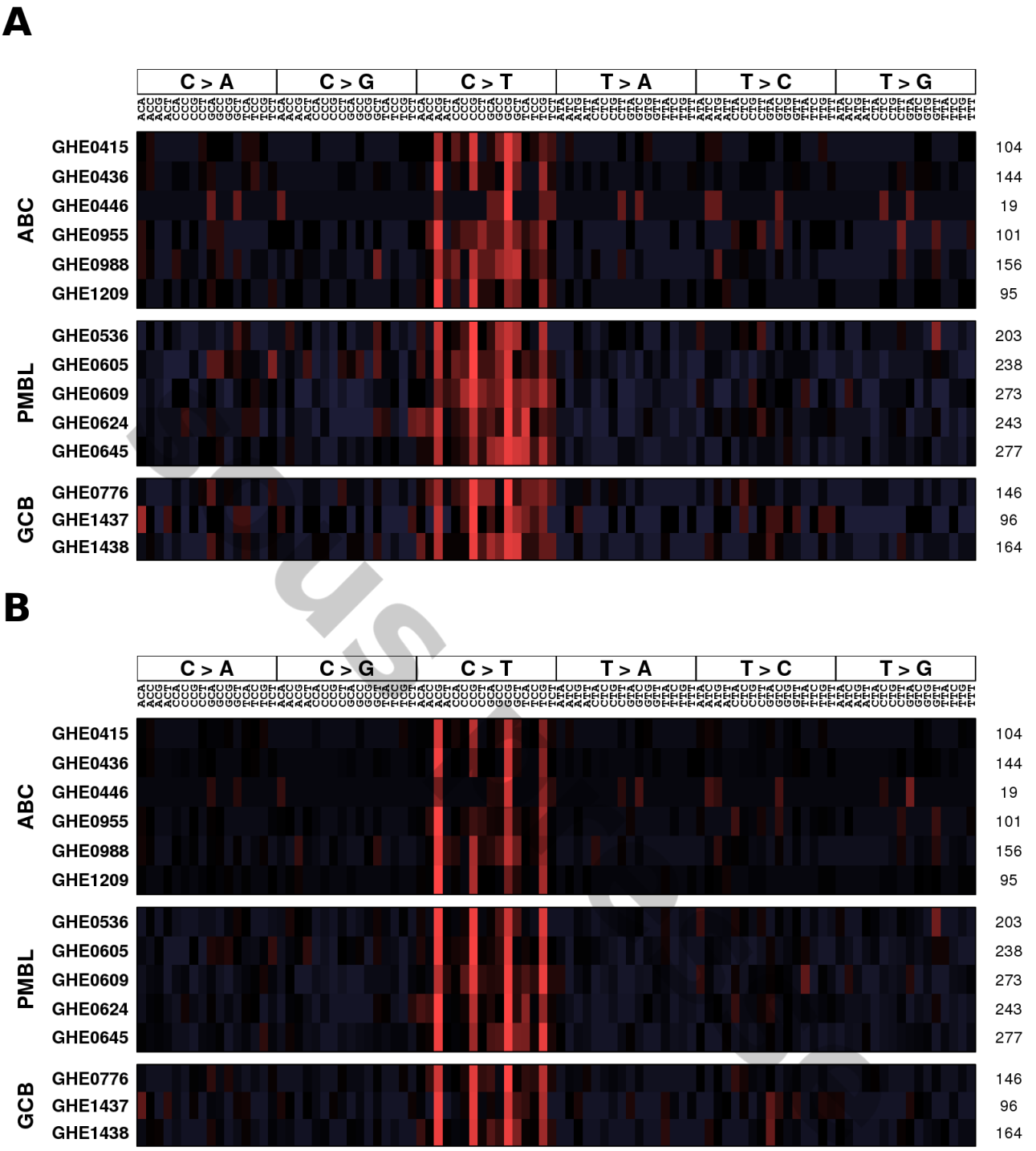
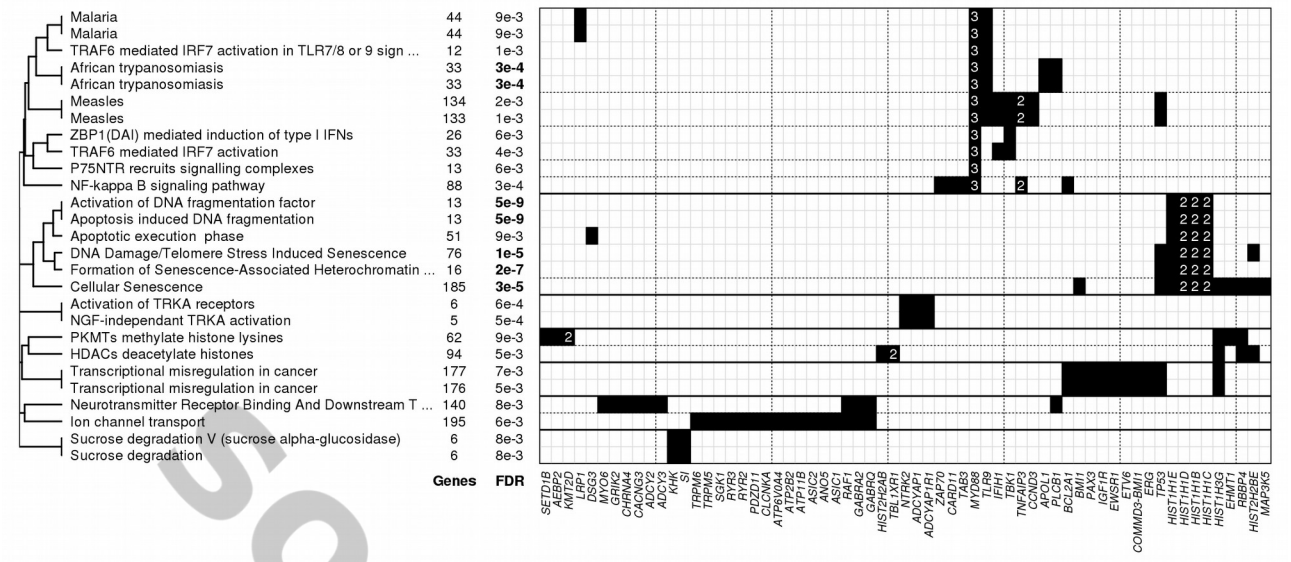


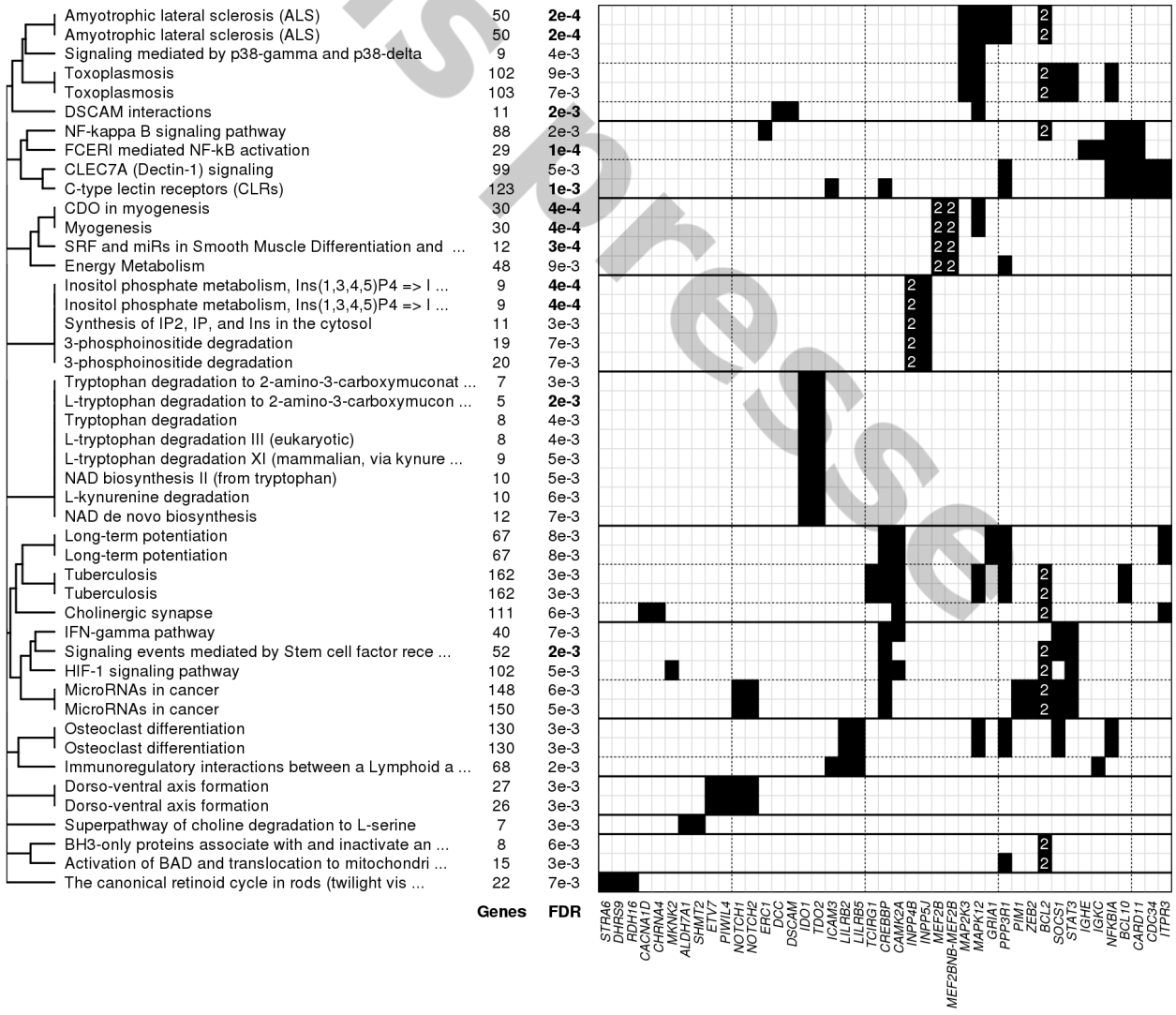
Figure S3 - Molecular signatures in the R/R cohort

Somatic SNVs are categorized according to their nucleotidic context (preceding and following base), defining 96 mutation types (regardless of the strand). Relative abundances of these 96 events for each patient are depicted as a heat map, with blue cells representing low values and red cells high values. 2 259 somatic SNVs falling in the coding region (ROI) are considered here, as synonymous and tolerated non-synonymous mutations were considered as well. The total amount of considered mutations for each patient is written on the right of each row. Panel **A** is built upon raw counts, while panel **B** is normalized by the occurrence of the corresponding tri-nucleotide pattern in the coding genome (ROI).

A



B



sous presse

C

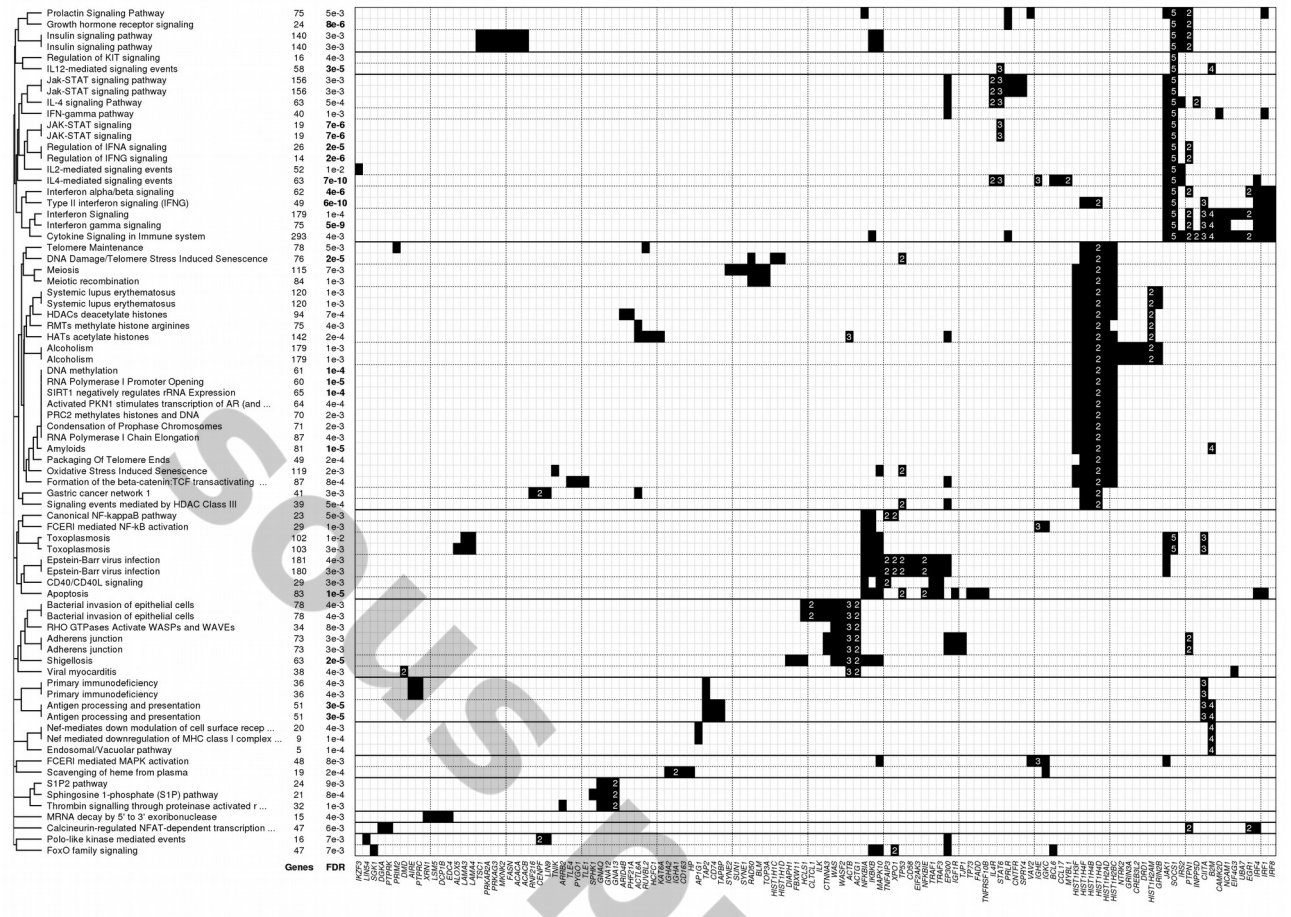
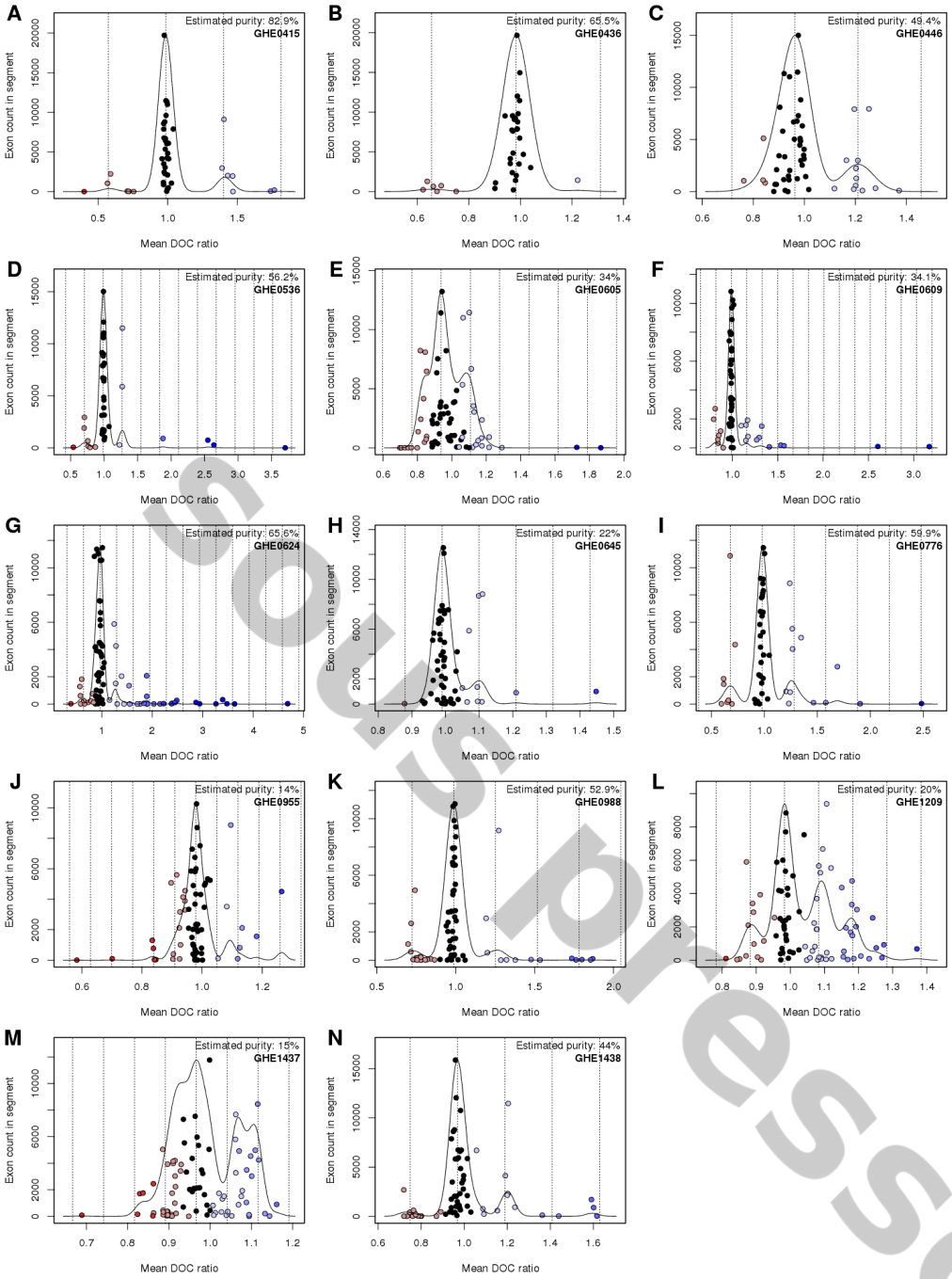


Figure S4 - Pathways significantly mutated in the R/R cohort.

Pathways harboring more mutations than expected by the sole background mutation rate (FDR<1%) are presented in rows, with the corresponding mutated genes in columns. A black cell represents one or many mutations in the same patient, while a white number signals mutations in multiple patients. Hierarchical clustering was performed to identify clusters of pathways relying on the same mutations, with two arbitrary levels of decreasing similarity materialized by plain and dotted horizontal lines. The analysis was performed separately in the 3 DLBCL subtypes (ABC in panel **A**, GCB in panel **B** and PMBL in panel **C**).

Science Direct

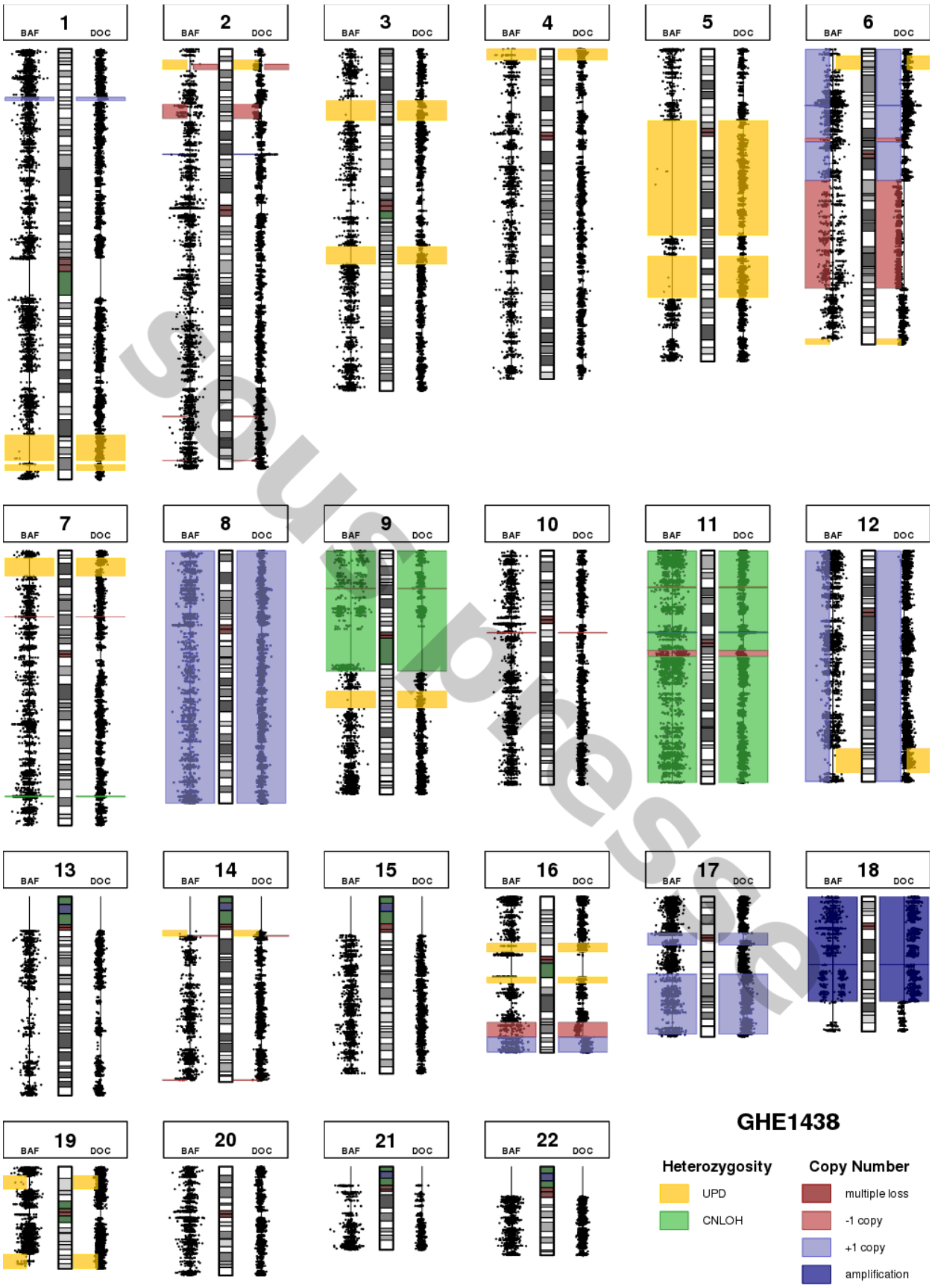


Sous presse

Figure S5 - Copy number models

For each of the 14 R/R patients (panels **A** to **N**), genomic segments are represented in a scatter plot with their mean DOC ratios on the X axis and the number of consecutive exons they consist of on the Y axis. A gaussian kernel density weighted by the segment size (in exons) was fitted independently to each sample, in order to highlight recurring DOC values corresponding to the various copy number states observed in the sample. Considering the most represented copy state as 2n (black dots), copy states were thus attributed to each segment following this model (increasing copy number gains from light to dark blue, deletions in red). The sample purity could be estimated from the shift width between consecutive DOC peaks, and is reported in the top right corner of each panel.

N



source files

Figure S6 - Complete copy number analysis results

For each of the 14 R/R patients (panels **A** to **N**), BAF and DOC ratio computed from the whole-exome sequencing data are plotted along each chromosome. CNV and LOH regions detected by the various algorithms applied are highlighted by colored bands (figure keys available in the bottom right corner). Wild-type values of 0.5 for BAF and 1 for DOC ratio are materialized by vertical lines.

[Les panneaux A à M ne sont pas inclus dans cette version]

Supplemental Tables

Patient	Sex	Age	Localization	Bulky	AAIPI	IPI	Stage	Subtype	*MYC* rear.	*BCL2* rear.	*BCL6* rear.
GHE0415	F	76	Right inguinal	No	2	3	IV	ABC	–	–	–
GHE0436	F	76	Right cervical	No	3	5	IV	ABC	–	–	–
GHE0446	M	82	Left cervical	No	2	4	IV	ABC	–	–	–
GHE0536	F	35	Mediastinum	No	1	1	I	PMBL	–	–	–
GHE0605	M	29	Mediastinum	No	1	1	I	PMBL	NA	NA	–
GHE0609	F	44	Left sus-clavicular	Yes	1	1	II	PMBL	–	–	–
GHE0624	M	35	Mediastinum	No	2	2	IV	PMBL	–	–	–
GHE0645	M	43	Left cervical	No	1	1	II	PMBL	NA	NA	–
GHE0776	M	48	Spleen. left renal. peritoneum. colon	No	1	2	IV	GCB	–	–	+
GHE0955	F	70	Left cervical	No	2	4	IV	ABC	–	–	–
GHE0988	M	75	Right inguinal	No	3	5	IV	ABC	NA	NA	NA
GHE1209	M	73	Left maxillary	Yes	2	3	III	ABC	NA	NA	NA
GHE1437	M	62	Sub-clavicular	No	2	3	IV	GCB	+	+	–
GHE1438	M	64	Left inguinal	No	3	5	IV	GCB	–	+	–

Patient	Treatment	OS (months)	OS event	PFS (months)	PFS event
GHE0415	R-CHOP21	16.7	TRUE	10.1	TRUE
GHE0436	R-CHOP21	7.1	TRUE	6.7	TRUE
GHE0446	R-Mini-CHOP21	16.4	TRUE	8.3	TRUE
GHE0536	R-CHOP21	14.4	TRUE	11.0	TRUE
GHE0605	R-ACVBP+Conso	17.9	TRUE	10.7	TRUE
GHE0609	R-CHOP21	45.5	FALSE	9.5	TRUE
GHE0624	R-ACVBP+ASCT	8.6	TRUE	5.6	TRUE
GHE0645	R-CHOP21	46.3	FALSE	10.1	TRUE
GHE0776	R-ACVBP+Conso	27.4	FALSE	3.6	TRUE
GHE0955	R-CHOP21	8.9	TRUE	6.7	TRUE
GHE0988	R-CHOP14	14.6	TRUE	12.6	TRUE
GHE1209	R-CHOP14	7.7	TRUE	6.0	TRUE
GHE1437	R-CHOP14	18.0	TRUE	10.8	TRUE
GHE1438	R-CHOP21	18.1	TRUE	8.3	TRUE

Table S1 - Detailed clinical features of the R/R cohort

Clinical description of the 14 R/R patients sequenced by WES.

Not for Use

	WES	PGM
Variants considered	**112**	**160**
- Variants found in target	131	171
- PGM depth < 20	19	8
- WES normal depth < 8	0	2
- WES tumoral depth < 8	0	3
Cross-confirmed variants	**96**	**96**
True exclusive variants	**11**	**16**
- Low depth in concurrent dataset	4	10
- Low VAF in concurrent dataset	0	1
- Missed by concurrent analysis	7	5
False exclusive variants	**0**	**41**
- Polymorphisms	–	30
- Technology-specific artefacts	0	11
Unknown status	**5**	**7**

Table S2 - Somatic event orthogonal validation

Variants (indels and SNVs) found in the 34 genes and 12 samples sequenced by both techniques are compared.

sous presse

Supplemental Datasets

Supplemental Datasets are provided as separate Excel files on the publisher website.

Dataset S1 - Quality assessment of the Whole-Exome Sequencing

Dataset S2 - Filtered somatic variant list

Dataset S3 - Pathways enriched in mutations in the R/R cohort

Dataset S4 - Comparative enrichment analysis of significant pathways

Dataset S5 - Significant Mutated Allele Selection events in the R/R cohort

5 Bilan : oncogénèse des LDGCB

5.1 Événements précoces

5.1.1 Dérégulations épigénétiques

Les LDGCB se caractérisent par des mutations d'un certain nombre de régulateurs épigénétiques, et tout particulièrement dans le sous-type GCB (§4.1.3). Les mutations du gène ***KMT2D*** / *MLL2* et de ses homologues (*KMT2C* / *MLL3*, *KMT2E* / *MLL5*), méthyltransférases ciblant la lysine 4 de la protéine histone 3 (H3K4), sont ainsi retrouvées dans plus de 30% des LDGCB GCB et 15% des ABC, et restent encore largement inexpliquées [261]. Il a au moins été clairement établi que ces mutations diminuent le niveau global de H3K4 triméthylées, en tronquant KMT2D en amont de son domaine fonctionnel localisé en C-ter, ou en l'altérant [288]. Si un parallèle intéressant a pu être établi avec l'activité d'AICDA et des protéines RAG, deux sources d'instabilité majeures dans les LDGCB, il semblerait que ces mutations conduisent en fait à réduire ces activités, en diminuant la proportion d'H3K4 triméthylées qui leur sont propices [332]. Cette observation est d'autant plus paradoxale que les mutations de *KMT2D* sont suspectées d'être un événement particulièrement précoce dans la lymphomagénèse [269], et pourrait donc jouer contre les nombreuses mutations imputées à *AICDA* dans ces lymphomes (*BCL2*, *BCL6*, *PIM1*, *MYC*...).

CREBBP et ***EP300*** sont deux gènes paralogues également inactivés par mutation de manière récurrente dans les LDGCB (§4.1.3), et ce avec une tendance à s'exclure mutuellement [261]. Là encore ces mutations conduisent à la perte d'une conformation de la chromatine favorable à la transcription (acétylation des histones), avec cette fois-ci des cibles mieux identifiées. Bcl-6 et p53 se trouvent en effet directement acétylés par ces enzymes, avec respectivement des effets inhibiteurs et activateurs. L'inactivation de *CREBBP* et *EP300* contribue donc à l'établissement d'un programme transcriptomique favorable à la tumorigénèse, avec une activité forte de Bcl-6 et une activité faible de p53 (§5.3). Bcl-6 contribuant à réprimer l'expression d'*EP300*, une

boucle de rétro-action négative s'installe dans les LDGCB exprimant fortement Bcl-6, ce qui pourrait expliquer la fréquence moindre des mutations d'*EP300* (présentant peu d'avantage pour la cellule) vis-à-vis des mutations de *CREBBP*.

Les mutations Y646 d'***EZH2*** observées fréquemment dans les LDGCB GCB contribuent également à garder la chromatine dans un état répressif, via l'accumulation de groupes tri-méthyles sur les H3K27 (§4.1.3, Annexe A). De nombreuses régions

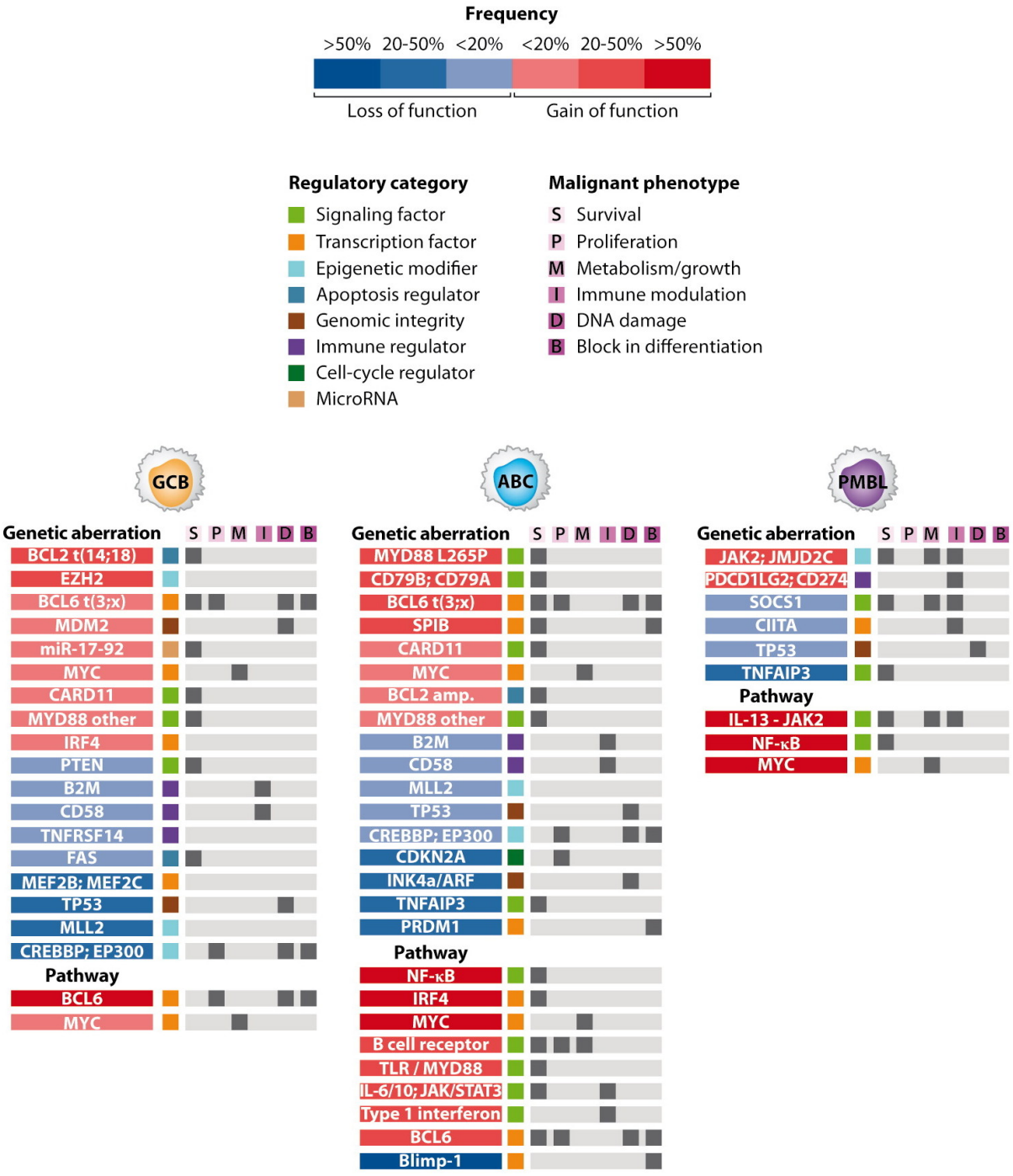


Figure 5.1 – Bilan des caractéristiques récurrentes des LDGCB ABC, GCB et LPM

Figure adaptée de Shaffer et al, ARI 2012 [42].

du génomes se trouvent dérégulées par ces changements de conformation, mais les mécanismes précis mis en jeu par cette mutation d'*EZH2* restent encore à élucider.

Globalement ces différents mécanismes tendent à réprimer l'expression de nombreux gènes, un comportement qui peut sembler curieux de la part d'une tumeur, qui se caractérise plutôt par un emballement du métabolisme. L'activité transcriptomique et plusieurs de ces modifications d'histone ont été suggérées comme favorables à l'action d'AICDA [333], on peut donc s'étonner qu'un phénomène d'apparence anti-tumorale soit aussi répandu (plus de 50% des LDGCB seraient touchés par un mutation de *CREBBP*, *EP300*, *KMT2D* ou *EZH2* [261]). Il pourrait s'agir d'un moyen pour le lymphocyte de protéger au moins une partie de ses gènes des mutations et cassures provoquées par *AICDA*, toujours en activité dans les LDGCB GCB et LF particulièrement sujets à ces mutations [43, 334]. D'autres effets fournissant un avantage supérieur à la cellule, tels que décrits pour *CREBBP* et *EP300*, pourraient également expliquer ce paradoxe, faisant de l'épigénétique un domaine particulièrement propice aux avancées concernant la mise en place de l'oncogénèse des LDGCB.

5.1.2 Translocations de *BCL2*

La translocation t(14;18) impliquant *BCL2* et *IGH* est un second événement considéré comme particulièrement précoce dans la lymphomagénèse, préférablement chez les LDGCB GCB. En effet, cette translocation, retrouvée dans 80% des LF et 22% des LDGCB GCB [38, 186], est généralement le fruit d'un dysfonctionnement survenu lors de la recombinaison VDJ (§1.3.2). Cet événement pourrait donc avoir lieu dès le stade pro-B, dans la moelle osseuse, bien en amont de ce que l'on considère improprement comme le « type cellulaire d'origine » en observant les phénotype et génotype de la tumeur (GCB et ABC). L'observation de lymphocytes B non-tumoraux arborant une t(14;18) chez des individus sains a permis d'établir que cet événement survenait chez de nombreux individus, mais dans peu de lymphocytes (10^{-5} lymphocytes B circulants ou moins [335]). S'il ne suffit pas à mettre en place l'oncogénèse, cet événement lui offre cependant un terrain propice, la prévalence des LF étant doublée chez les patients présentant ces lymphocytes pré-tumoraux [336].

Cette **translocation** offre au LDGCB une résistance à l'apoptose, qui lui permettra d'éviter les nombreux mécanismes de filtration à l'œuvre tout au long de la différenciation normale du lymphocyte B (§1.3, §1.4). Si on la retrouve plus fréquemment chez les LDGCB GCB, les LDGCB ABC semblent privilégier l'**amplification** du *locus* 18q21 qui l'héberge (31% [154]) pour arriver à cette même sur-expression de Bcl-2.

Plus tardivement, cette translocation offre un terrain particulièrement favorable aux LDGCB exprimant fortement ***BCL6*** (via un programme caractéristique du centre

germinatif ou via translocation, §5.3). En effet Bcl-6 réprime ordinairement l'expression de ce gène, et la substitution des séquences activatrices normales de *BCL2* occasionnées par la t(14;18) permet donc au lymphocyte de garder une expression forte simultanée de ces deux oncogènes (§3.1.2).

Enfin le spectre **mutationnel** observé sur *BCL2* peut être directement corrélé à cette translocation (4.1.3), l'apposition du *locus IGH* naturellement ciblé par la SHM favorisant l'action ectopique d'*AICDA* sur ce gène [262]. Le peu d'incidence des mutations observées (essentiellement synonymes dans les domaines fonctionnels) tend à confirmer que ces mutations sont un effet indésirable, qui pénalise la compétitivité des clones porteurs de mutations impactant l'activité de Bcl-2.

5.1.3 Mutations de *CD79A* et *CD79B*

Les mutations tronquantes touchant le gène ***CD79B*** (20% des LDGCB ABC, §4.1.3), et dans une moindre mesure *CD79A*, ont également été proposées comme des événements précoces dans l'oncogénèse des LDGCB, cette fois-ci du sous-type ABC [42]. En remplaçant la première tyrosine des motifs ITAM présents dans ces deux gènes (Annexe A), ces mutations sont susceptibles d'empêcher l'action des kinases effectrices de la voie du BCR (§5.2), et notamment de pénaliser l'action de LYN. Contrairement aux autres, cette kinase a également un effet inhibiteur de la voie du BCR, impliqué dans les phénomènes d'anergie pouvant survenir en fin de maturation des lymphocytes B (§1.3.4) ou plus tardivement. En gardant au moins un motif ITAM susceptible de déclencher la voie du BCR tout en minimisant l'impact de LYN, ces mutations permettraient à des lymphocytes B faiblement auto-réactifs d'éviter l'inactivation de leur BCR par anergie, tout en maintenant un niveau d'activité suffisant à leur survie.

5.2 Au cœur de l'oncogénèse

5.2.1 La voie du BCR

Les mutations de *CD79B* décrites précédemment contribuent à la mise en place d'une activation du BCR dite **chronique**, similaire à celle observée lorsqu'un lymphocyte B est activé par son antigène (§1.4). Le déclenchement de cette réaction par un auto-antigène est une des hypothèses envisagées pour expliquer cette réaction. Une seconde hypothèse suggère que des altérations du BCR en lui-même (chaines d'immunoglobulines ou co-effecteurs) suffiraient à déclencher l'agrégation

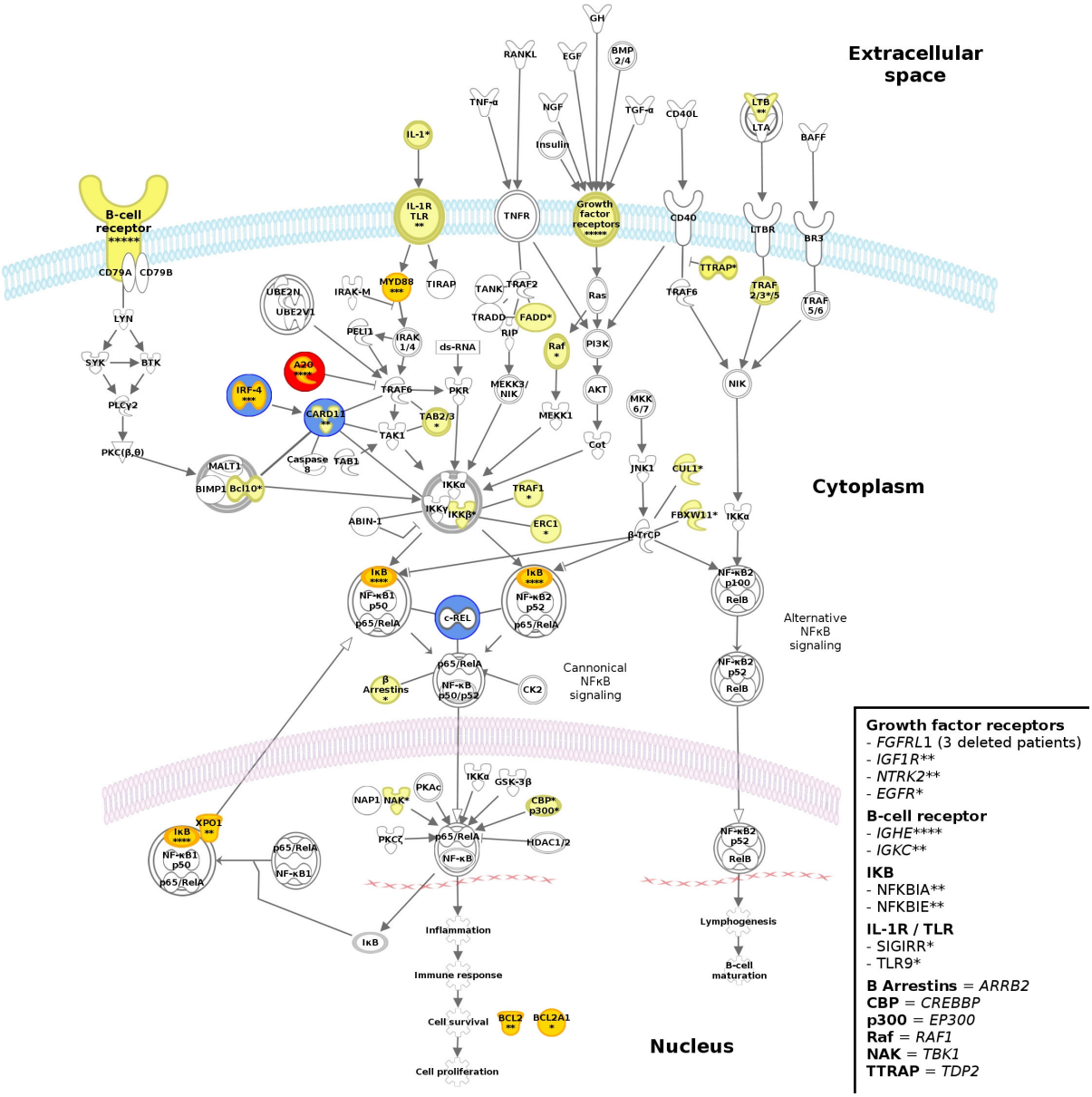


Figure 5.2 – Altérations récurrentes au sein des voies du BCR, des TLR et NFκB

Les différents composants de la voie NFκB sont présentés sur cette figure, ainsi qu'une partie des voies des TLR et du BCR conduisant à son activation. Les composants en jaune clair ont été trouvés mutés chez au moins un des 14 LDGCB réfractaires dont l'exome a été séquencé, les composants en jaune foncé ont de plus été déterminés comme significativement enrichis en mutations. Les éléments sur fond rouges sont délétés de manière récurrente, les éléments sur fond bleu amplifiés de manière récurrente. Figure extraite de Mareschal et al, GCC [328].

des BCR, et donc l'activation de la cascade de signalisation en aval [42]. Considérant que les chaines lourdes d'immunoglobulines semblent en mesure de contrôler cette agrégation [337], les mutations observées dans les exomes de nos 14 LDGCB réfractaires sont en adéquation avec cette deuxième hypothèse (§4.4). En effet, les parties constantes des chaines μ chez les ABC et des chaines α_1 et ε chez les LPM figurent parmi les gènes les plus fréquemment mutés, en accord avec l'expression préférentielle d'une IgM chez les ABC [127] et d'une immunoglobuline commutée chez les LPM [338]. Ces deux sous-types reposent fortement sur la voie NFκB [42], dont l'activation chronique du BCR pourrait être un déclencheur.

La mutation du gène *CARD11* en aval du BCR (Figure 5.2) constitue un second moyen pour les LDGCB ABC de conserver cette voie continuellement active. Cette activation se caractérise par sa constance et sa stabilité, et son indépendance vis-à-vis d'une grande partie des effecteurs en amont de *CARD11* (immunoglobulines membranaires, CD79, LYN, SYK, BTK...) [42]. Si ces mutations ont originellement été mises en évidence préférentiellement dans le sous-type ABC [339], des proportions similaires ont été depuis mises en évidence dans les LDGCB GCB (Figure 4.5, Annexe A). Quel que soit son mode d'activation, cette voie métabolique conduit à de nombreuses voies clé dans la survie des lymphocytes B (NFκB, PI3 et MAP kinases...). Les connaissances actuelles tendraient à exclure la voie NFκB des pré-requis à la survie des LDGCB GCB, néanmoins l'article fondateur de cette idée n'incluait qu'une seule lignée GCB (OCI-Ly7) [45], chez qui le gène *CARD11* n'est pas muté [263]. Nous sommes donc en droit de supposer que l'activation de NFκB via *CARD11* et le BCR est un phénomène également présent chez les LDGCB GCB, comme cela a été suggéré par l'observation d'une signature transcriptomique NFκB chez les LDGCB GCB mutés *CARD11* [339].

D'autres événements sont susceptibles de faciliter cette activation. Nous avons en effet pu mettre en évidence chez deux des 14 LDGCB réfractaires (un GCB et un ABC) un **gain de copies** focal subséquent à la mutation de *CARD11*, résultant en une amplification de l'allèle muté (§4.4). Là encore la présence d'un patient GCB muté *CARD11* confirme l'existence de ce phénomène dans ce sous-type, et l'associe à un pronostic plutôt sombre compte tenu de la rechute précoce observée chez ce patient. La sur-expression d'***IRF4***, phénomène constitutif des LDGCB ABC (§5.3), contribue également à activer *CARD11*, en augmentant son niveau de transcription [340]. Des gains de copies et des mutations d'apparence activatrice ont également pu être mises en évidence sur le gène d'*IRF4* chez nos 14 patients réfractaires, et tout particulièrement dans le sous-type ABC.

5.2.2 La voie des TLR

Les récepteurs *toll-like* (TLR) constituent un mécanisme alternatif au BCR, participant à l'immunité innée. Ces récepteurs détectent selon un principe similaire la présence d'antigènes étrangers dans l'organisme, sans toutefois la diversité caractéristique des immunoglobulines : une dizaine de récepteurs différents seulement assurent la reconnaissance d'un nombre restreint d'antigènes conservés chez les agents infectieux (lipopolysaccharides, flagélines, ARN viraux...). Dans les lymphocytes B, l'activation des TLR contribue à potentialiser les signaux obtenus via le BCR et CD40, et peut aiguiller la commutation isotypique vers l'un ou l'autre des isotypes en fonction des menaces détectées [341] (§1.4.6). L'activation des TLR débouche plus généralement sur la voie NFκB, favorisant la survie de la cellule, et sur la production d'IL-6, IL-10 et INFγ, cytokines favorables à l'évasion immunitaire (§5.3).

La mutation L265P du gène ***MYD88*** retrouvée dans 30% des LDGCB ABC (§4.1.3) constitue un moyen pour la cellule tumorale de garder cette voie en état de constante activation. En effet, cette mutation permet l'assemblage spontané d'un complexe protéique impliquant IRAK1, IRAK4 et TRAF6, dont l'interaction avec CARD11, Bcl-10 et MALT1 permet d'activer la voie NFκB [342] (Figure 5.2). Si plusieurs mutations du domaine TIR de *MYD88* semblent pouvoir générer cet effet en absence de tout stimulus issu des TLR (Annexe A), la mutation L265P présente l'intérêt supplémentaire de stabiliser IRAK1, et donc de potentialiser encore un peu plus cette activation [42]. Les gènes *MYD88*, *IRAK1* et *IRAK4* ont été démontrés comme indispensables à la survie des LDGCB ABC [343], un sous-type dans lequel une certaine complémentarité a pu être observée entre mutations de *MYD88* et activation chronique du BCR [42].

La délétion du bras 6q, incluant de manière particulièrement récurrente le gène ***TNFAIP3***, est un phénomène qui peut être directement relié à *MYD88*. La protéine A20 codée par *TNFAIP3* semble en effet inhiber l'activité de la voie NFκB en interagissant avec TRAF6 [344], sa délétion contribue donc à préserver ou potentialiser l'activation initiée en amont par *MYD88* (Figure 5.2). L'inactivation de *TNFAIP3* est un phénomène que nous avons également pu mettre en évidence dans nos travaux, via une région minimale délétée de manière homozygote d'environ 100 kb dans la série GHEDI (6% des patients, plus environ 30% de délétions hétérozygotes §3.6). Des mutations tronquantes et délétions ont également été trouvées lors du séquençage exomique des 14 LDGCB réfractaires, préférentiellement dans les sous-types ABC et LPM particulièrement sujets à une activité forte de NFκB (§4.4).

5.2.3 La voie NFκB

La voie métabolique NFκB représente un acteur majeur de la survie des lymphocytes B, et plus généralement d'un grand nombre de cellules normales et tumorales (Figure 5.2). Elle repose sur plusieurs hétérodimères formés des protéines p65 (*RELA*), RELB, c-Rel (*REL*), p50 (*NFKB1*) et p52 (*NFKB2*), retenus dans le cytoplasme par des inhibiteurs tels que IκBα (*NFKBIA*) ou IκBε (*NFKBIE*). L'activation du complexe IκB kinase (IKK) par l'un des nombreux acteurs en amont (dont le complexe protéique formé par CARD11, Bcl-10 et MALT1) permet l'adressage des inhibiteurs IκB pour dégradation dans le protéasome, et donc le transfert des différents dimères NFκB dans le noyau. Ils pourront alors réguler la transcription d'un grand nombre de gènes, dont *BCL2* et ses proches parents [42]. De la même façon une voie alternative de NFκB repose sur la kinase NIK (*MAP3K14*) et un complexe IKK légèrement différent, et conduit à des effets similaires (Figure 5.2).

Dans les LDGCB, la voie classique est fréquemment impliquée grâce à une activation pathologique en amont, dans la voie du BCR ou des TLR (*MYD88*). Plusieurs membres et régulateurs de la voie NFκB en elle-même présentent également des altérations favorisant son activité. Le *locus* 2p15-p16, abritant ***REL***, est ainsi décrit comme fréquemment amplifié dans les LDGCB GCB et LPM (35% et 50% respectivement, §3.1.1), une observation que nous avons également pu faire dans la série GHEDI (région minimale communément gagnée d'environ 100 kb, impliquant 23% des patients, §3.6). Ce gain de copies a été relié à la transformation des LF en LDGCB, ce qui pourrait expliquer qu'on la retrouve principalement dans le sous-type GCB que présentent généralement les LF transformés [166]. Si l'implication de la voie NFκB n'explique pas la lymphomagénèse chez ces patients qui présentaient déjà un LF, elle est en tout cas associée à une évolution défavorable. Ceci est cohérent avec l'observation que les LDGCB GCB ne dépendent pas autant de cette voie que les ABC pour leur survie, en soulignant toutefois son implication dans les deux sous-types.

De la même façon, le bras 17q est fréquemment amplifié dans les LDGCB (jusqu'à 25% des cas, Figure 3.5) et associé à la transformation des LF [166, 345]. Si le MCR fluctue d'une série à une autre et a conduit à présenter plusieurs gènes candidats, il est intéressant de constater la présence en 17q21 du gène ***MAP3K14*** initiant la voie NFκB alternative, à quelques dizaines de kb seulement du MCR retrouvé dans 25% des LDGCB par Scandurra *et al* [153]. Cette seconde observation soutient la thèse d'une implication relativement tardive de NFκB dans le sous-type GCB, via l'amplification de certains de ses composants ou la mutation de *CARD11* [268, 346].

L'analyse mutationnelle de 14 LDGCB réfractaires (§4.4) a permis de mettre en avant l'altération d'autres régulateurs de NFκB. Des mutations tronquantes (non-sens ou *frameshift*) ont ainsi été retrouvées dans les gènes ***NFKBIA*** et ***NFKBIE*** chez

trois LPM et un GCB, des altérations qui pourraient conduire à la relocalisation de NFκB dans le noyau en absence de tout stimulus en amont. Le gène ***XPO1***, codant l'exportine CRM1 capable de renvoyer NFκB dans le cytoplasme en interagissant avec IκBα [347], a également été retrouvé muté en un même *locus* (E571K) chez deux LPM. Si de nombreuses autres protéines d'intérêt telles que p53, CDKN1B ou E2F5 sont également exportées hors du noyau par CRM1, il s'agit d'une mutation dont nous avons pu confirmer la récurrence dans environ 40% des LPM [324, 325], et qui pourrait contribuer à la forte activité de NFκB dans ce sous-type.

5.3 Un nouveau programme transcriptomique

5.3.1 Survie cellulaire

Comme n'importe quel cancer, le LDGCB se caractérise par le maintien dans l'organisme de cellules vouées à l'apoptose, du fait d'une perte d'intégrité génomique ou de processus naturels de sélection (réarrangements non fonctionnels du BCR §1.3.2, lymphocytes auto-réactifs §1.3.4, maturation d'affinité §1.4.5 ...). Si l'activation de la voie NFκB décrite précédemment constitue un des principaux mécanismes mis en jeu dans ce processus, d'autres peuvent également contribuer.

En temps qu'effecteur de l'apoptose, ***BCL2*** constitue ainsi une cible de choix. Membre d'une famille de protéines similaires retrouvées dans la membrane externe des mitochondries, Bcl-2 est en effet impliquée dans un équilibre de forces pro (Bax, Bak) et anti-apoptotiques (Bcl-2, Bcl-X_L), garant de l'imperméabilité des mitochondries. En interagissant avec l'une ou l'autre de ces protéines membrannaires, des effecteurs tels que Puma, Noxa, Bim ou Bid sont capables de modifier cet équilibre, et d'induire le relargage dans le cytoplasme de protéines mitochondriales mortelles pour la cellule (cytochrome c, Diablo, HtrA2). L'induction par ces protéines de la caspase 9 conduit finalement à la mise en place de la mort programmée, consistant en un démantèlement systématique du noyau et de la cellule en « corps apoptotiques » [1]. L'expression de *BCL2* se retrouve ainsi fréquemment amplifiée dans les LDGCB, via la translocation t(14;18) (§3.1.2), un gain de copies sur le bras 18q (§3.1.1) ou l'activation de la voie NFκB. Les délétions en 13q14 retrouvées dans 10% à 15% des LDGCB pourraient également contribuer à l'expression du gène *BCL2*, via l'inactivation des micro-ARN miR15a et miR16-1 codés par le *locus DLEU2* [42]. Ce même *locus* a par ailleurs été retrouvé impliqué dans un MCR dans la série GHEDI, avec 5 patients de pronostic particulièrement sombre présentant une délétion homozygote (§3.6).

Le proto-oncogène ***MYC*** est également la cible de nombreuses altérations favorisant la survie de la cellule. Ce facteur de transcription potentialise l'activation

de l'expression d'un large panel de gènes, plaçant ainsi la cellule dans un état de sur-activité métabolique. *MYC* est exprimé physiologiquement de manière ponctuelle pour de courtes phases prolifératives, en réponse notamment à l'activation par les lymphocytes T (aux stades matures §1.4.3 et centrocytes §1.4.5). Son expression est toutefois rapidement réprimée par des facteurs de transcription tels que *BCL6* ou *PRDM1* durant l'immunopoïèse (§1.4). Outre son activation transcriptionnelle, on lui attribue également la répression d'un certain nombre de gènes, via des régulateurs épigénétiques tels qu'*EZH2* [193]. La translocation t(8;14) (§3.1.2) constitue un moyen pour la cellule d'exprimer *MYC* en continu, et donc de décupler les effets d'événements oncogéniques plus précoces tels que la sur-expression de *BCL2*. Cette translocation est suspectée de reposer sur un dysfonctionnement de la commutation isotypique (§1.4.6), une étape plus tardive dans le développement du lymphocyte B, ce qui est cohérent avec ce rôle d'événement secondaire amplificateur. Des gains de copies en 8q24 peuvent également être à l'origine de la sur-expression de *MYC*, dans 10% à 15% des cas (Figure 3.5). L'amplification du cluster de micro-ARN ***MIR17HG*** en 13q31-q32 (Figure 3.5) a également pu être reliée chez 13% des LDGCB GCB à une expression plus forte de *MYC*, suggérant une coopération entre ces acteurs [22]. ***PIM1*** serait un troisième acteur de ce trio [348, 349], chez qui de nombreuses mutations dues à une SHM ectopique sont mises en évidence dans les LDGCB (§4.1.3). L'impact de ces mutations de *PIM1* reste toutefois à élucider, dans la mesure où certaines de ses mutations semblent activatrices et d'autres inhibitrices [350], et que des translocations avec *BCL6* ont été rapportées [286]. ***JAK2*** et *KDM4C*, deux gènes voisins amplifiés dans plus de 50% des LPM (§3.1.1), seraient enfin impliqués dans l'activation de la transcription de *MYC*, via des mécanismes épigénétiques [42].

La protéine **p53**, suppresseur de tumeur mis en cause dans une grande majorité des cancers, ne fait pas exception dans les LDGCB. Cette protéine est garante de l'intégrité de la cellule, répondant à des stress variés (dégradation de l'ADN, stress oxydant, manque de nutriments...) par des mesures proportionnées : activation de processus de réparation, inhibition de la prolifération voire mort programmée si les dégâts sont trop importants. La réparation des dommages causés à l'ADN peut ainsi être dirigée par p53, via un arrêt temporaire du cycle cellulaire (*CDKN1A*) et l'activation d'acteurs de la réparation tels que *PARP1*. La réponse de p53 au stress oxydant est également emblématique de ce gène : cette protéine peut à la fois contribuer à diminuer un stress modéré (via de nombreux effecteurs tels que *TIGAR*, *GLS2*, *GPX1*...) ou amplifier un stress trop important, de sorte d'induire la mort cellulaire. Cette ultime solution peut également être mise en place par p53 en induisant l'expression de Bax, Noxa et Puma, qui en interagissant avec les protéines de la famille de Bcl-2 déclencheront la mort cellulaire via la mitochondrie [351]. Les cellules tumorales inactivant *TP53* disposent donc d'un avantage sélectif certain. Cette inactivation peut passer

par la mutation de ce gène, retrouvée dans 22% des LDGCB tous sous-types confondus (§4.1.3). La délétion du *locus* 17p13 qui héberge ce gène est également un phénomène fréquent et peu spécifique (10% à 15% des LDGCB, §3.1.1). Les micro-ARN codés par le *locus* ***MIR17HG*** en 13q31-q32 semblent également impliqués dans l'inhibition transcriptionnelle de *TP53* [42], leur amplification touchant 13% des LDGCB GCB pourrait donc également contribuer à la survie cellulaire via ce mécanisme (Figure 3.5). L'acétylation de p53 par ***CREBBP*** et ***EP300*** mentionnée un peu plus tôt (§5.1) peut également être mise en cause dans cette inactivation, de même que la forte expression de *BCL6* sur laquelle nous reviendrons.

Le *locus 9p21*, qui fait l'objet de fréquentes délétions dans les LDGCB et les cancers en général (§3.1.1), est également un acteur clé de la survie cellulaire. La protéine p14 codée par ***CDKN2B*** est en effet un intermédiaire entre *MYC* et p53, dont elle déclenche d'ordinaire l'activation en réponse à ces signaux d'hyper-prolifération [156]. La délétion de ce gène offre donc à la cellule la possibilité de bénéficier des avantages prolifératifs de *MYC* sans mettre en route p53 et ses nombreux acteurs suppresseurs de tumeurs. Les protéines p15 et p16 codées par ***CDKN2A*** et *CDKN2B* sont quant à elles impliquées dans un second mécanisme de contrôle, qui vise à bloquer le cycle cellulaire via *CDK4* et *CDK6* en cas de perte d'intégrité du génome [155]. Nos travaux réalisés sur la série de CGH GHEDI (§3.6) ont mis en évidence cette délétion comme un facteur pronostic majeur, y compris au sein des LDGCB ABC auxquels elle est préférentiellement associée (39/81 présentaient une délétion homozygote). Nous avons également pu mettre en évidence une signature transcriptomique de cet événement, incluant notamment les gènes *DEK*, *RDH10*, *PKN3* et *PIH1D1* dont la relevance biologique nous a paru particulièrement intéressante.

5.3.2 Blocage de la différenciation

La protéine **Bcl-6** est considérée comme un élément clé dans la formation et le maintien du centre germinatif. Il s'agit d'un facteur de transcription, dont l'effet inhibiteur repose sur le recrutement d'enzymes modifiant les histones (désacétylation, déméthylation, ubiquitination) telles que BCOR, NCOR1 et NCOR2 via son domaine BTB. Son domaine de liaison à l'ADN en C-ter lui permet de se fixer au promoteur de plusieurs milliers de gènes, dont environ un quart semblent effectivement réprimés [352]. Parmi ces gènes cibles, on retrouve notamment *IRF4* et *PRDM1* nécessaires à la différenciation des lymphocytes du centre germinatif, l'expression de *BCL6* favorisant donc un maintien à ce stade. De nombreux composants des voies CD40, TLR et BCR sont également ciblés, ce qui peut être vu comme une manière d'étouffer l'activité de ces voies lors de la maturation d'affinité (§1.4.5), et donc de ne permettre la survie que des lymphocytes B recevant un signal d'activation significatif. De la même

façon, l'anti-apoptotique *BCL2* fait partie des cibles notables de la répression transcriptomique opérée par *BCL6*. Des gènes impliqués dans le contrôle de l'intégrité du génome, tels que *TP53* ou *CDKN1A*, se retrouvent réprimés par ailleurs. Compte tenu des processus mutagènes à l'œuvre dans le centre germinatif (SHM §1.4.5, commutation isotypique §1.4.6), ceci peut s'expliquer par la nécessité pour la cellule d'être temporairement plus tolérante vis-à-vis de ces modifications génétiques nécessaires. L'avantage que tirent les LDGCB de la sur-expression de *BCL6* (notamment les GCB [40]) reste donc complexe à évaluer, des effets favorables et défavorables à l'oncogénèse étant mis en jeu. Certains de ces effets défavorables peuvent être contre-balancés, notamment par les translocations qui permettent d'extraire *BCL2* et *MYC* de son contrôle transcriptomique ou par des signaux d'activations plus forts encore, comme c'est vraisemblablement le cas pour la voie NFκB [42]. Quoi qu'il en soit, *BCL6* est fréquemment la cible de **translocations** dans les LDGCB (§3.1.2), augmentant son expression grâce à un promoteur plus fort et exempt des sites de fixation qui l'obligent d'ordinaire à réprimer sa propre transcription. Ces sites peuvent également être altérés par des **mutations**, qui concernent environ 15% des LDGCB sans translocation [353]. Plus classiquement, *BCL6* est lui-même la cible d'autres répresseurs de la transcription, tels que ***CREBBP*** et ***EP300*** qui sont eux-même la cible de mutations inactivatrices récurrentes (§5.1).

La protéine MUM1 codée par ***IRF4*** est un second facteur de transcription, occupant la même place centrale dans les LDGCB ABC que Bcl-6 dans les LDGCB GCB. C'est d'ailleurs essentiellement sur ces deux facteurs que repose l'algorithme IHC de Hans, qui vise à distinguer ces deux sous-types (2.1.1). L'expression d'*IRF4*, déclenchée dans les centrocytes par l'activation de la voie NFκB en réponse aux stimuli du BCR et de CD40, serait ainsi l'une des premières étapes de la différenciation plasmocytaire. Son activation permet la répression de l'expression de Bcl-6 et l'activation de celle de Blimp-1, passant d'un programme transcriptomique visant la maturation d'affinité et la prolifération à un programme de différenciation cellulaire [354]. La répression de la voie NFκB opérée par Bcl-6 serait alors levée, conduisant à sa forte activation caractéristique des LDGCB ABC. En outre *IRF4* stimule l'expression d'*AICDA* et la commutation isotypique (§1.4.6), une observation qui peut sembler surprenante quand on sait que les LDGCB ABC exprimant fortement *IRF4* présentent essentiellement une IgM [127]. Une étude plus attentive révèle que les LDGCB ABC présentent justement de nombreuses délétions des régions *switch*, cicatrices d'une commutation isotypique toujours à l'œuvre, mais qu'exprimer une IgM plutôt qu'un autre isotype constitue un avantage sélectif pour la cellule tumorale [355]. Le gène *IRF4* en lui-même n'est que très rarement altéré, suggérant que son activité physiologique est un atout suffisant pour la survie des cellules tumorales. Sa position en 6p25 en fait un gène occasionnellement **amplifié** par la formation d'iso-chromosomes 6 (Figure

3.5), et une **translocation** t(6;14) avec le *locus IGH* peut être observée dans 5% des LDGCB, définissant un sous-ensemble peu caractérisé de LDGCB GCB avec un profil transcriptomique hybride [89]. Lors du séquençage de 14 LDGCB réfractaires (§4.4), nous avons pu mettre en évidence de rares **mutations** potentiellement activatrices du gène *IRF4*, associées à un niveau d'expression de ses cibles transcriptomiques équivalent à celui observé chez des patients présentant un gain de copies d'*IRF4*. L'amplification du gène ***SPIB*** en 19q13 observée dans 25% à 30% des LDGCB ABC (Figure 3.6) peut également être reliée à *IRF4*, ces deux gènes interagissant pour activer *CARD11* et former ainsi une boucle de rétro-activation de la voie NFκB [42].

La protéine Blimp-1 codée par ***PRDM1*** est à la base de la différenciation en plasmocyte, qui constitue la phase suivante dans la différenciation normale du lymphocyte B. Sa seule présence a en effet été démontrée comme suffisante pour déclencher cette différenciation, et conduit à la mise en place d'un phénotype centré sur la sécrétion d'immunoglobulines (expression accrue des chaînes lourdes et légères) et une prolifération atténuée (inhibition de *MYC*). Des facteurs de transcription tels que *BCL6* et *PAX5* sont également réprimés par Blimp-1, marquant la fin du programme transcriptomique caractéristique du centre germinatif (réponse au BCR, *AICDA*, MHC de classe II, *CD40*...) [356]. Ce gène est la cible dans les LDGCB d'une répression efficace, permettant un maintien des lymphocytes B à un stade de différenciation pré-plasmocytaire propice à l'accumulation d'altérations génétiques (*AICDA*) et à la prolifération (BCR, TLR, NFκB...). Si son expression est activée par la forte présence d'*IRF4* chez les LDGCB ABC, il s'avère que le gène *PRDM1* n'est plus fonctionnel dans près de la moitié de ces lymphomes [357]. Cette inactivation peut prendre la forme d'une délétion (environ 30% des LDGCB présentent une perte de son *locus* en 6q21, §3.1.1), ou de mutations tronquantes (10% à 15% des LDGCB ABC, Figure 4.5). La sur-expression de *BCL6*, notamment par le biais de sa translocation avec divers partenaires (30% des LDGCB ABC, §3.1.2), ou encore l'amplification de *SPIB* (gains de copies en 19q13 dans 25% à 30% des LDGCB ABC, §3.1.1) peuvent mener au même résultat, dans la mesure où ces gènes répriment l'expression de *PRDM1*.

5.3.3 Évasion immunitaire

L'organisme met naturellement en place des mesures immunologiques pour contrer le développement de cellules tumorales, il est donc naturel d'observer chez les LDGCB des phénomènes d'échappement à ces mécanismes. Le MHC de classe I constitue l'un de ces mécanismes, qui permet l'activation de lymphocytes T cytotoxiques en réponse à la présentation d'un antigène intra-cellulaire, et donc la destruction de la cellule présentatrice. Le gène ***B2M*** codant un de ses composants est ainsi fréquemment inactivé dans les LDGCB, soit par mutation (10% à 20% des

LDGCB et 50% des LPM, Figure 4.5), soit par délétion de son *locus* en 15q21 (jusqu'à 25% des LDGCB, dont 40% de délétions homozygotes focales manquées en CGH, Figure 3.6 [358]). Le gène ***CD58*** codant une protéine de surface impliquée dans la coopération avec les lymphocytes T et NK est également la cible d'inactivations récurrentes, souvent associées à l'inactivation de *B2M*. Des mutations tronquantes sont ainsi observées dans un peu moins de 10% des LDGCB et 40% des LPM (Figure 4.5), de même qu'environ 20% de délétions [358]. Environ 60% des LDGCB perdraient ainsi l'expression membranaire du MHC de classe I et de *CD58*, échappant ainsi à la surveillance des lymphocytes T cytotoxiques et NK [358].

Le MHC de classe II, présentant cette fois-ci des antigènes aux lymphocytes T auxiliaires, est également la cible d'altérations dans les LDGCB. La perte de ce composant a d'ailleurs pu être corrélée à un pronostic plus sombre, qui s'expliquerait par une moindre activation des lymphocytes T cytotoxiques par ces lymphocytes T auxiliaires [359]. Le gène ***CIITA***, qui co-active l'expression de la plupart des composants du MHC de classe II, est ainsi la cible de translocations récurrentes dans 40% des LPM, perdant sa séquence promotrice naturellement active dans les lymphocytes B au profit d'un partenaire moins exprimé. Les protéine chimères ainsi obtenues auraient de plus un effet répresseur de l'activité de la protéine *CIITA* native codée par le second allèle [42]. Des mutations de ce gène peuvent également être observées dans 10% des LDGCB et 56% des LPM, contribuant à son inactivation particulièrement récurrente dans ce dernier sous-type (Figure 4.5). On pourra s'étonner tout de même d'observer des gains plutôt que des pertes de copies en son *locus* 16p13 (12% des LDGCB ABC et 7% des LPM, Figure 3.6), dans une proportion qui reste marginale vis-à-vis des autres mécanismes décrits.

6 Discussion

6.1 Transfert de la classification GCB / ABC

6.1.1 Contexte

La grande disparité biologique entre LDGCB ABC, GCB et LPM a été illustrée tout au long des différents chapitres du présent mémoire, et conduit à l'heure actuelle à une disparité pronostique d'intérêt majeur pour les cliniciens (§2.1.1). Si l'importance de cette classification et la performance des puces pan-transcriptomiques pour l'établir sont admises de tous [38, 40, 41], la question du transfert en routine clinique est cependant toujours ouverte, 15 ans après sa découverte [36]. Cantonné aux laboratoires de recherche, ce « *gold standard* » basé sur une technique relativement couteuse et complexe a permis de mieux comprendre la biologie de ces sous-types, et répondait tout à fait aux exigences de chercheurs désireux d'explorer également l'intégralité du transcriptome. De nouvelles approches thérapeutiques ont cependant découlé de ces recherches, notamment pour les LDGCB ABC qui présentent les survies les plus modestes avec le traitement de référence actuel (R-CHOP) [84, 292], et le besoin d'une technique plus simple se fait désormais sentir dans les laboratoires de diagnostic. La compatibilité avec des échantillons conservés en paraffine (FFPE) plutôt que congelés est également un prérequis majeur à ce transfert, la cryoconservation étant surtout l'apanage des laboratoires de recherche. Si des protocoles ont bel et bien été proposés pour analyser des échantillons FFPE sur puces Affymetrix [70], leur utilisation reste marginale.

6.1.2 Solution *NanoString Technologies*

La société *NanoString Technologies* propose ainsi un nCounter™ qui semble correspondre à ces attentes [361], et pratique un lobbying vigoureux pour se placer comme la technologie de référence en ce domaine [83]. Cette technique repose sur

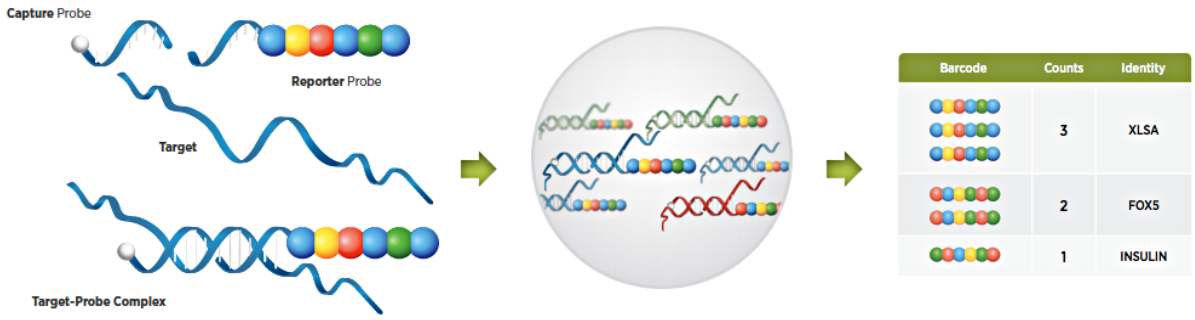


Figure 6.1 – Principe de la technologie nCounter™

Les principales caractéristiques de la technologie nCounter™ développée par *NanoString Technologies* sont illustrées ici, à savoir l'hybridation de la sonde de capture biotinylée et de la sonde marquée d'un code-barre fluorescent (panneau de gauche), la lecture multiplexe des code-barres dans un échantillon (panneau du milieu) et le comptage qui s'en suit, permettant de mesurer le niveau d'expression de chaque transcrit ciblé (panneau de droite). Figure adaptée du site internet officiel de *NanoString Technologies* [360].

l'hybridation de deux types de sondes, l'une portant une biotine permettant de capturer les transcrits ciblés au sein d'un échantillon d'ARN, l'autre portant un code-barre fluorescent permettant son identification (Figure 6.1). Ces code-barres sont composés d'une succession de 6 macro-fluorochromes, qui permettent d'attribuer environ 800 identifiants uniques au sein d'une même expérience, et donc potentiellement de mesurer l'expression d'autant de gènes, voire de protéines [362]. Son applicabilité à des échantillons FFPE et sa simplicité d'utilisation (15 minutes de manipulation sur 2 jours de protocole [360]) en font une technique adaptée aux contraintes actuelles des laboratoires de diagnostic, d'autant plus séduisante que son approche s'affranchit de la PCR et se ses biais potentiels pour compter directement les molécules présentes.

La première application de cette technologie a été développée dans les cancers du sein HR+, où existent également plusieurs sous-types requérant des approches thérapeutiques distinctes (Luminal A et B, HER2+ et *basal-like*, [363]). Initialement réalisée sur puces transcriptomiques, cette classification a pu être rapportée à 50 gènes (PAM50), dont l'expression était initialement mesurée par PCR quantitative en temps réel [364]. Le test Prosigna™ commercialisé par *NanoString Technologies* [365] propose ainsi le même modèle et la même signature avec une certification IVD en Europe et aux États-Unis [366, 367], et rejoint des tests tels que Oncotype DX, MammaPrint et Endopredict dans les dernières recommandations de l'ESMO [368].

Un scénario très similaire semble se mettre en place pour la classification GCB / ABC des LDGCB. Également découverte au début des années 2000 durant l'âge d'or des puces transcriptomiques [36], cette signature a rapidement été réduite à une vingtaine de gènes [40] dont la mesure en PCR quantitative ou toute autre méthode moléculaire peine à s'imposer dans les laboratoires de diagnostic [74]. Si plusieurs

équipes se sont proposées d'appliquer la technologie nCounter™ à ce problème [80–82], l'équipe du NCI à l'origine du modèle simplifié (G.W. Wright, L.M. Staudt *et al*) semble aujourd'hui concrétiser les choses auprès de *NanoString Technologies*. Une première **preuve de concept** a ainsi fait l'objet d'une lettre dans le journal Blood en février 2014 [81], établissant le test « Lymph2Cx » basé sur l'expression de 20 gènes et un modèle LPS classique en la matière (§2.3.3). 15 de ces gènes furent sélectionnés parmi 93 issus de la signature établie précédemment sur puces transcriptomiques [41], auxquels furent ajoutés 5 gènes de ménage. Le modèle fut entrainé dans une série de 51 biopsies FFPE par comparaison à une classification réalisée sur puces Affymetrix et échantillons congelés appariés, dont seulement 20 GCB et 19 ABC participèrent effectivement au modèle (12 indéterminés). Ils ne rapportèrent qu'une seule discordance avec le « *gold standard* » dans une série de validation comprenant 58 biopsies FFPE de types GCB ou ABC (excluant les indéterminés), et une corrélation quasi-parfaite entre les résultats obtenus dans deux laboratoires différents. Une **application** est ensuite parue dans JCO en septembre 2015 [369], reprenant ce même test dans une cohorte plus conséquente. Outre des essais satisfaisants de reproductibilité et de stabilité sur différents lots de réactifs, cette étude permit d'assigner un des trois sous-types à 335 des 344 patients rassemblés, excluant 5 tumeurs à faible contenu tumoral et 4 échecs. La différence pronostique entre LDGCB ABC et GCB sous R-CHOP fut une fois de plus démontrée, et ce indépendamment de l'IPI et de l'hyperexpression des gènes *MYC* et *BCL2* déterminée en IHC [108]. Cette technologie prend pied depuis peu dans le secteur diagnostique, d'importants moyens financiers ayant été déployés par la société Celgene pour en faire le test compagnon du lenalidomide, un traitement qui pourrait bénéficier au sous-type ABC [83, 84].

6.1.3 Solution RT-MLPA

Le test RT-MLPA que nous avons développé avec le même objectif (§2.5) se place ainsi en concurrence directe avec le test américain porté par *NanoString Technologies*, et ce depuis les toutes premières heures. Ces deux techniques ont en effet été présentées lors de la même session au congrès annuel de l'ASH à la Nouvelle-Orléans en 2013 [139, 370], et les publications qui en découlent affichent des résultats et des applications très similaires. Après un peu moins d'un an de soumissions à divers journaux, nous avons finalement pu publier cette technique dans JMD en mai 2015 [137].

En réduisant l'hybridation à une cinquantaine de nucléotides consécutifs (§2.3.2), la RT-MLPA permet d'afficher une efficacité comparable au nCounter™ sur des **échantillons FFPE** réputés pour leur importante fragmentation [69] (9 échecs sur 344 rapportés par Scott *et al* [369], 2 échecs d'extraction sur 30 échantillons FFPE dans

notre analyse). Outre ces échecs en amont de la classification, des taux similaires de patients inclassables (10% à 15%) sont rapportés par les deux techniques.

En terme de **validité de la classification**, chacune des deux techniques montre une concordance quasi-parfaite avec le « *gold standard* » Affymetrix réalisé sur échantillons congelés (1/58 et 0/53 discordances rapportées respectivement par Scott *et al* [81] et nous-même). Si la comparaison directe sur une série de 173 échantillons FFPE de la série GHEDI semble confirmer cette équivalence, les termes de l'accord passé avec *NanoString Technologies* pour accéder à ce test (encore indisponible dans le commerce) ne nous permettent cependant pas d'en publier les résultats.

La RT-MLPA souffre par comparaison aux puces Affymetrix d'une même faiblesse vis-à-vis des **gènes exprimés à des niveaux très différents**, les gènes les plus exprimés ayant tendance à écraser le signal issu des autres gènes. L'ajout de sondes 3' exemptes de séquence leste et d'amorce (compétiteurs) nous a permis de régler ce problème, en diminuant spécifiquement l'amplification de certains gènes (*IGHM* et *NEK6*). Si ce problème n'est pas signalé par Scott *et al*, il est cependant probable compte tenu de son principe que la technologie nCounter™ souffre de la même faiblesse, et que cette dernière ait influencée le choix des 15 gènes retenus dans le modèle, dont *IGHM* et *NEK6* sont d'ailleurs absents.

En terme de **débit**, la RT-MLPA repose sur un protocole plus court et flexible que la technologie de *NanoString Technologies*, quoiqu'impliquant un peu plus de manipulations manuelles. Nous avons en effet pu démontrer la facilité d'analyser indistinctement de 1 à 40 échantillons en parallèle sur une demi-journée, alors que la preuve de concept publiée par Scott *et al* fait état de plusieurs *runs* de 24h impliquant des cartouches de 12 ARN, dont un ARN de référence [81]. Si les derniers modèles de nCounter™ proposent d'analyser plusieurs ARN par pistes et de monter jusqu'à 96 échantillons simultanés via l'acquisition d'appareils supplémentaires, il convient de préciser que le nCounter™ Dx certifié IVD n'autorise que 12 échantillons avec le test Prosigna™ qui servira vraisemblablement de modèle au futur test LDGCB [371].

Le principal atout de la RT-MLPA consiste en son **coût de revient** dérisoire vis-à-vis de son concurrent. La RT-MLPA repose en effet sur des outils déjà présents dans la plupart des laboratoires de diagnostic moléculaire (thermocycleur et appareil d'électrophorèse capillaire), sur un mélange d'amorces permettant d'analyser plusieurs millions d'échantillons pour quelques centaines d'euros tout en s'affranchissant des biais de lots, et environ 5 euros de réactifs de base par échantillon. Par opposition, la solution proposée par *NanoString Technologies* représente un investissement initial d'environ 250 000 euros, pour des tests commercialisés à l'heure actuelle à environ 1 500 euros par patient (devis réalisé pour le test Prosigna™).

6.1.4 Perspectives

Ces disparités illustrent bien les difficultés du transfert des découvertes réalisées dans les laboratoires de recherche pour application dans les laboratoires de diagnostic. Si la France autorise encore l'usage de kits de tests dont l'usage diagnostique n'a pas été certifié, la loi HPST de 2009 impose aux laboratoires de biologie moléculaire d'obtenir d'ici 2016 la certification ISO-15189 auprès d'organismes tels que le COFRAC, ce qui peut être perçu comme un premier pas vers un contrôle plus poussé des outils diagnostiques. Dans ce contexte, des sociétés privées telles que *NanoString Technologies* sont plus à même de fournir les moyens financiers, légaux et humains nécessaires à l'obtention de telles certifications, ainsi que des systèmes fermés et automatisés limitant les erreurs humaines. Il convient toutefois de s'interroger sur le prix que représentent ces technologies pour la société, quand des alternatives fonctionnelles et moins coûteuses existent et peuvent être distribuées librement.

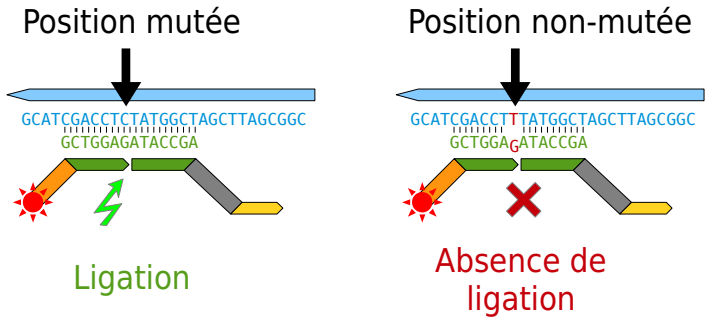


Figure 6.2 – Application de la RT-MLPA à la recherche de mutations

Phase de ligation dans le cas d'une recherche de mutation récurrente en RT-MLPA. Les autres étapes présentées en Figure 2.4 restent inchangées.

Sur un plan plus technique, le test RT-MLPA que nous avons développé fait l'objet d'un développement continu au sein de notre laboratoire, qui a depuis conduit à des améliorations notables. Outre le remplacement et l'ajout de plusieurs gènes qui nous font entrevoir la possibilité d'identifier également les LPM avec ce même test, des mutations ponctuelles récurrentes peuvent également être ciblées avec cette technologie. Il semblerait en effet qu'un mésappariement à proximité du site de ligation soit suffisant pour prévenir cette dernière, et donc de n'enregistrer un signal que dans le cas où l'ARN étudié est effectivement muté en cette position (Figures 6.2 et 6.3). Nous serions ainsi en mesure d'identifier les mutations contribuant à l'établissement du sous-type (mutation L265P de *MYD88* chez les LDGCB ABC, mutations Y646 d'*EZH2* chez les LDGCB GCB ou encore E571K de *XPO1* chez les LPM, §4.1.3 et §4.4), avec la seule contrainte que le *locus* interrogé soit la cible d'une mutation récurrente toujours identique. Cette technique est également applicable à la recherche

de transcrits de fusions [372], bien que ceux-ci soient relativement marginaux dans les LDGCB (§3.1.2). Enfin l'apparition de kits d'extraction simultanée des ADN et ARN ouvre à plus long terme la possibilité d'un test incluant également les nombreux CNA récurrents dans les LDGCB (§3.1.1), pour peu que les couples d'amorces ciblant l'ARN soient élaborés sur des jonctions d'exons et les couples d'amorces ciblant l'ADN sur des introns ou des régions intergéniques à proximité des gènes d'intérêt.

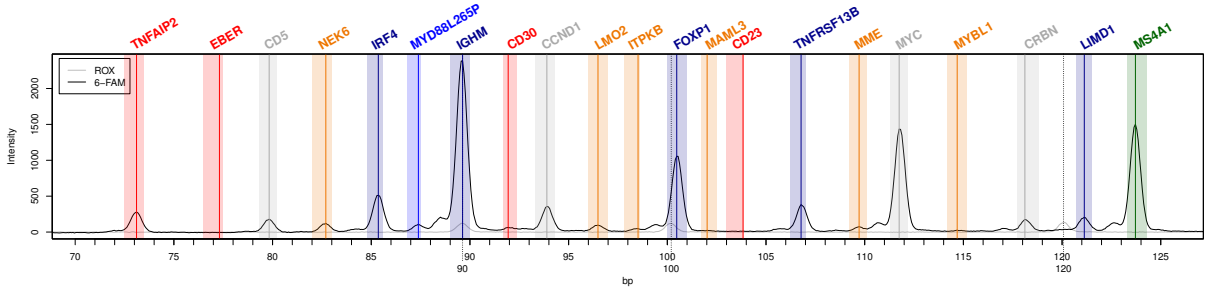


Figure 6.3 – Exemple du profil RT-MLPA en développement

Chaque pic correspond à l'expression d'un gène, habituellement surexprimé dans les ABC (bleu) ou les GCB (orange). Les gènes en rouge, gris et vert correspondent à des signatures en développement, caractéristiques des LPM (*TNFAIP2*, *CD30*, *CD23*) ou présentant un intérêt clinique. Le patient présenté ici est un ABC typique, exprimant fortement les gènes en bleu et faiblement les gènes en orange. La mutation L265P de *MYD88* est également présente.

Sur le plan biologique, les applications des techniques discutées ici vont bien au delà de la classification GCB / ABC. Outre les nombreuses autres signatures moléculaires d'intérêt pronostique survolées dans ce mémoire (§2.1.2), Keane *et al* [373] ont démontré récemment l'intérêt d'une nouvelle signature aux implications pronostiques et thérapeutiques non-redondantes avec la classification GCB / ABC. Consciente de l'importance des phénomènes d'évasion immunitaire dans les LDGCB (§5.3.3), cette équipe choisit de quantifier par la technologie *NanoString* la présence et l'activité des cellules immunitaires normales infiltrant la tumeur (CD4, CD8, CD137 et CD56 pour les lymphocytes T et NK, CD68 et CD163 pour les macrophages). Les protéines PD-1 et PD-L1 exprimées respectivement par les lymphocytes T et B furent également ciblées pour leur fort potentiel d'inhibition de la réponse T. Combinant plusieurs de ces gènes au sein d'un même score, ils établirent ainsi une classification en deux groupes de pronostics différents, et ce indépendamment du R-IPI et de la classification GCB / ABC. Si les méthodes statistiques employées laissent craindre l'*over-fitting*, l'application aux puces Affymetrix publiées par Lenz *et al* [41] semblent confirmer cet effet dans une série indépendante, avec des écarts plus modestes (survies à 4 ans de 92% et 47% dans la série d'entrainement, contre 76% et 64% dans la série de validation). Cet outil ouvre la voie aux inhibiteurs de l'axe PD-1 et PD-L1 actuellement en essais cliniques [374–376], et plus généralement à des traitements visant non plus la tumeur en elle-même mais son micro-environnement.

6.2 Quels outils pour détecter les CNA ?

6.2.1 Contexte

Comme de nombreuses équipes avant la nôtre (§3.1.1), nous avons réalisé un portrait précis des CNA récurrents dans les LDGCB, soulignant les disparités biologiques entre LDGCB ABC et GCB et des altérations d'intérêt pronostique (§3.6). Lors de l'élaboration de cette étude il y a un peu plus de 6 ans, les puces CGH mises en œuvre constituaient un outil de choix pour ce travail, et leur analyse a fait l'objet d'un développement continu tout au long de ma présence dans cette unité de recherche (§3.4, §3.5). Il convient toutefois de s'interroger sur la place que cette technique peut occuper aujourd'hui dans les laboratoires, face à l'émergence d'outils plus récents et potentiellement plus performants.

6.2.2 Puces SNP

Les puces SNP ont représenté une première évolution du concept de la CGH (§3.3.1), apportant des informations supplémentaires sans changer fondamentalement le principe. Si plusieurs modèles sont actuellement disponibles, cette technique n'implique généralement plus une compétition directe entre deux ADN marqués de couleurs différentes sur une même puce, mais plutôt la compétition entre deux allèles au sein d'un même échantillon d'ADN. Constituées de sondes couvrant les SNP de fréquence élevée, ces puces sont ainsi à même de quantifier l'hétérozygotie naturellement présente tout au long d'un génome normal diploïde, et donc de compter le nombre d'allèles en cas de CNA (Figure 6.4, AD). La comparaison à des puces réalisées sur ADN témoins permet également d'obtenir un log-ratio dépendant directement du nombre de copies, comme pour les puces CGH (Figure 6.4, LR). Si le calcul d'un log-ratio à partir de puces différentes implique pour les puces SNP un bruit généralement plus important, la mesure supplémentaire de l'hétérozygotie permet de construire des modèles plus robustes, et au final d'identifier avec beaucoup plus de précision le nombre de copies (Figure 6.4, panneau central). Il est de plus possible avec cette technique d'identifier les pertes d'hétérozygoties sans altération du nombre de copies (CNLOH), un phénomène peu étudié aujourd'hui qui touche pourtant une part significative des génomes tumoraux.

Compte tenu de ces ressemblances, de nombreux développements réalisés pour les puces CGH restent applicables aux données issues de puces SNP. Si le modèle cghRA de prédiction du nombre de copies (§3.5) est dans ce cas supplanté par des solutions tenant compte de l'hétérozygotie [377, 378], la bibliothèque Rgb sous-jacente

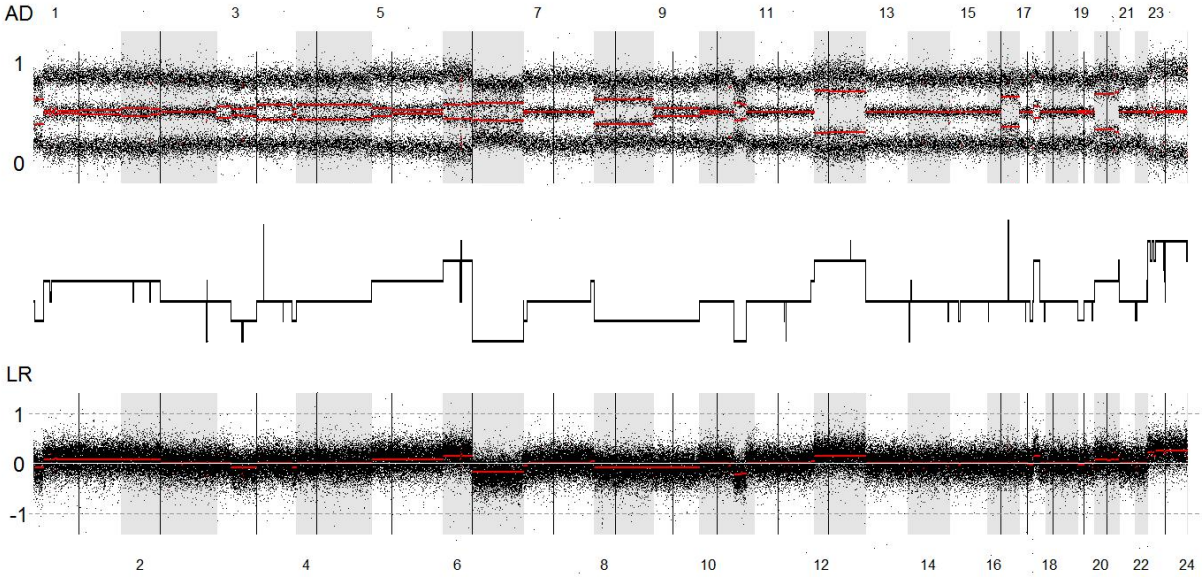


Figure 6.4 – Exemple de profil génomique par puce SNP

Les CNA au sein d'un génome de LF analysé dans le cadre du projet PRIMA-Bio sont représentés ici en fonction de leur coordonnées dans le génome normal, chromosomes mis bout à bout. Le panneau supérieur décrit la différence allélique (AD) en chaque point du génome, traduisant d'un déséquilibre d'hétérozygotie mesuré sur des SNP. Le panneau inférieur décrit le log-ratio (LR) entre un ADN normal et l'ADN tumoral, les gains de copies conduisant à un log-ratio positif. Le panneau central correspond à une interprétation de ces données par le logiciel GAP [377], le génome étant segmenté en régions chromosomiques présentant un même nombre de copies.

reste précieuse pour la représentation et la manipulation des données (§3.4). Cela a d'ailleurs pu être illustré lors du projet PRIMA-Bio, incluant 156 puces SNP CytoScanHD réalisées sur des LF au diagnostic, qui a beaucoup contribué à l'amélioration des algorithmes d'identification des CNA constitutionnels et des MCR décrits précédemment (§3.5).

Dans la littérature, les puces SNP remplacent progressivement les puces CGH, et ce d'autant plus rapidement que le nombre de sondes analysables sur une même puce augmente [22, 154]. Si comme nous le verrons les techniques de séquençage à haut débit commencent à prendre le pas sur les puces SNP, cette dernière technique reste encore une solution valable pour les années à venir compte tenu de son coût plus réduit et adapté à l'étude de grandes séries de patients. Une différence de coûts qui se retrouve au niveau des appareils de lecture (environ 190 000 € [379]) et des réactifs (environ 380 € par puce CytoScanHD [379]), mais également au niveau de l'analyse, qui est à l'heure actuelle mieux maitrisée et bien moins gourmande en ressources bio-informatiques dans le cas des puces.

6.2.3 Séquençage à haut débit

Comme cela a pu être illustré lors du séquençage exomique des 14 LDGCB réfractaires (§4.4), le séquençage à haut débit de larges portions génomiques génère également des données adaptées à la détection de CNA. Celles ci peuvent en effet être réduites à un jeu de données similaire à celui obtenu d'une puce SNP sans coût matériel supplémentaire, et faire l'objet d'un traitement équivalent (Figure 6.5). Le logiciel ADTEx utilisé dans ce projet [327] attribue à chaque exon un log-ratio calculé à partir des profondeurs moyennes observées dans l'échantillon tumoral et dans l'échantillon normal, réduisant le génome à environ 200 000 mesures pouvant s'apparenter à autant de sondes d'une puce CGH ou SNP. En s'intéressant aux séquences proprement dites, nous sommes également en mesure d'ajouter environ 50 000 positions hétérozygotes dans l'ADN normal, dont la mesure dans l'ADN tumoral pourra nous renseigner quant au nombre d'allèles et donc de copies dans la tumeur, similairement aux mesures de différence allélique des puces SNP. Incidemment, nous avons également identifié des événements de disomie uniparentale chez ces patients, qui se matérialisent par une perte plus ou moins étendue de l'hétérozygotie dans l'ADN constitutionnel (et par extension dans l'ADN tumoral, en jaune sur la Figure 6.5).

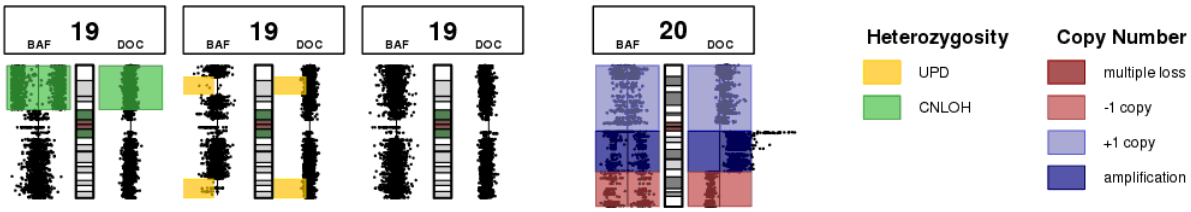


Figure 6.5 – Exemple de CNA détectés par séquençage exomique

Trois chromosomes 19 et un chromosome 20 issus de LDGCB différents sont présentés, afin d'illustrer différents types d'anomalies. Pour chaque chromosome, la piste de gauche représente la fréquence de l'allèle B (BAF) dans l'ADN tumoral, en chaque position hétérozygote dans l'ADN normal. Similairement, la piste de droite représente le ratio des profondeurs de séquençage (DOC), pour chaque exon séquencé.

Le séquençage de génomes complets plutôt que d'exomes constitue un raffinement supplémentaire dans cette analyse, la constance de la profondeur de couverture obtenue permettant d'obtenir des résultats beaucoup plus précis pour la détection de CNA [300]. Le calcul d'un log-ratio par fenêtres glissantes permet également d'obtenir des mesures réparties de manière homogène sur la quasi-totalité du génome et non plus ses régions codantes, tandis que le séquençage des régions non codantes plus riches en mutations constitutionnelles est à même de fournir une meilleur résolution en terme de suivi de l'hétérozygotie. Séquencer deux génomes pour chaque patient représente cependant un coût prohibitif pour la plupart des la-

boratoires de recherche, et ne pourra être réellement envisagé que dans plusieurs années, en admettant que le coût de ces techniques continue à baisser.

6.2.4 Techniques de transfert

Si l'application de ces méthodes à des données de reséquençage ciblé présente des difficultés supplémentaires, plusieurs équipes dont la nôtre se sont intéressées à ce problème, conscientes de la difficulté d'implémenter un séquençage d'exomes complets en routine diagnostique. Nous avons ainsi initié le développement d'un pipeline d'analyse dans le cadre du Lymphopanel, afin de détecter les altérations présentes dans les 34 gènes et 87,7 kb ciblés par ce kit de capture Ampliseq™ [271, 272]. Compte tenu du faible nombre de régions capturées par rapport à des données d'exomes, ce genre d'approche nécessite qu'un soin tout particulier soit apporté à la soustraction des différents biais de capture, d'autant plus forts qu'il s'agit ici d'une capture par amplification multiplexe et non plus par hybridation. De même cette réduction d'échelle rend les modèles de prédiction du nombre de copies basés sur la distribution des log-ratios (§3.5) ou les pertes d'hétérozygotie (§4.4.3) difficilement applicables, et appellent d'autres approches statistiques. Seul l'ADN tumoral étant généralement séquencé, il convient enfin d'établir un profil de référence à partir d'un pool d'ADN normaux, et de veiller à ce que la variabilité génétique inter-individus ne perturbe pas les résultats comme cela peut être observé en CGH (§3.5.3).

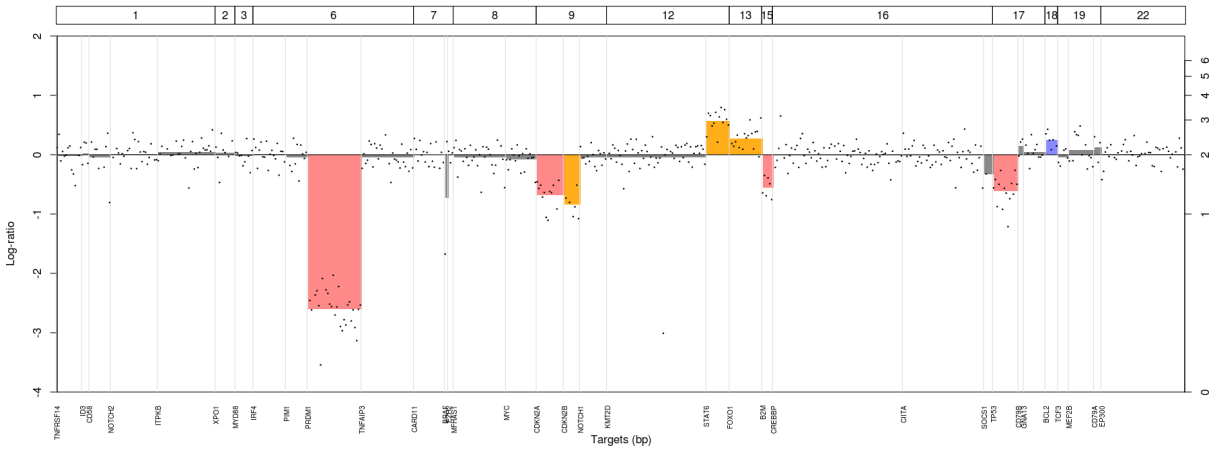


Figure 6.6 – Exemple de CNA détectés par reséquençage ciblé

Les 34 gènes du Lymphopanel sont représentés en abscisse, par ordre génomique (numéros de chromosome dans le bandeau supérieur). Chaque amplicon reséquencé est représenté par un point, dont l'ordonnée dépend du log-ratio des profondeurs observées dans l'échantillon et dans un profil normal de référence. Les délétions sont mises en évidence en rouge, les gains en bleu, les anomalies proches de la significativité en orange. Le nombre de copies théorique est présenté sur l'axe de droite. Les délétions en 6q (*PRDM1*) et 9p21 (*CDKN2A*, *CDKN2B*) sont particulièrement évidentes chez ce patient.

C'est ce que se propose de faire le logiciel ONCOCNV [380], que nous avons choisi d'implémenter compte tenu de son adéquation avec la technologie IonTorrent™ déployée pour le séquençage du Lymphopanel. Ce panel de gènes ayant été conçu pour la recherche de mutations parfois très localisées (points chauds couverts par quelques amplicons seulement), certains des gènes ciblés n'offrent pas une couverture suffisamment étendue pour assurer une sensibilité de détection des CNA optimale. La comparaison à des données de CGH semble toutefois confirmer la spécificité de ces détections, et l'adaptation du Lymphopanel et des méthodes d'analyse associées fait actuellement l'objet de développements au sein de notre laboratoire.

D'autres approches telles que la PCR quantitative [230] ou la MLPA sont également envisageables pour mettre en évidence des CNA dans un contexte diagnostique, néanmoins l'engouement actuel pour les techniques de séquençage à haut débit semble les destiner à cet usage. L'obtention conjointe de mutations et de CNA offre en effet un tableau bien plus relevant du génome tumoral, et les problèmes de débit qui limitent actuellement l'efficacité de la détection de CNA devraient s'estomper à mesure que ces techniques évoluent et que leur coût baisse.

6.3 Identifier les tumeurs réfractaires

6.3.1 Contexte

L'apparition du rituximab en clinique [381] a permis de limiter le taux de rechute des LDGCB à environ 30%, mais les facteurs cliniques et somatiques à l'origine de ces rechutes restent encore largement méconnus. C'est dans cette optique que nous avons choisi de concentrer nos efforts de séquençage sur 14 patients réfractaires ou en rechute précoce (§4.4), en rupture avec les précédentes études de séquençage à haut débit des LDGCB qui se contentaient d'un profilage génomique de cohortes hétéorogènes sur le plan clinique (§4.1.2). Si l'étude de larges séries de LDGCB par puces CGH ou SNP a permis d'identifier plusieurs CNA associés à un pronostic plus sombre (§3.1.1), il nous paraissait important de compléter ce tableau en explorant le profil mutationnel de ces tumeurs réfractaires.

6.3.2 Séquençage exomique

Novak *et al* ont réalisé un travail similaire au nôtre durant la conduite de cette étude, publié en août 2015 [266]. Cette équipe entreprit le séquençage exomique de 51 LDGCB traités par R-CHOP ou immuno-chimiothérapies équivalentes, mettant l'accent sur 13 patients présentant une EFS inférieure à 2 ans (progression, rechute,

nouveau traitement ou décès durant ce délai). Ils s'appuyèrent pour cette définition du groupe réfractaire sur les conclusions d'une étude menée sur plus de 700 LDGCB, qui démontre que les patients passant ce cap présentent une survie équivalente à la population normale [31]. Nous nous sommes limités pour notre part au délai plus resserré d'une rechute ou progression durant la première année, une différence de seuil marginale comme on peut le constater dans leur série (11/51 patients avec une EFS inférieure à 1 an, contre 13 à 2 ans).

Si aucun détail n'est donné quant aux modalités du séquençage qu'ils rapportent (profondeur, taux de mutation observé...), une citation de l'étude publiée par Lohr *et al* en 2012 [262] laisse supposer un séquençage de plus grande profondeur que le nôtre (150x contre 80x), avec une technologie similaire (Illumina HiSeq). Une sensibilité moindre étant de ce fait attendue dans notre étude, nous avons pu démontrer que cette profondeur nous permettait d'identifier des mutations clonales dans plus de 85% de l'exome, et donc que les profils mutationnels ne devraient pas différer significativement. L'accent est surtout porté dans leur étude sur la détection de CNA par puces SNP, un choix discutable quand de larges séries de LDGCB cliniquement caractérisés ont déjà été décrites [22, 153, 154]. Novak *et al* démontrent cependant une très bonne corrélation entre CNA prédits par puces SNP et par analyse de la profondeur de séquençage (PatternCNV [382]), ce qui valide notre choix de nous contenter de ce dernier dans notre série.

Ces auteurs se sont attachés à mettre en évidence une signature génomique propre au groupe de patients réfractaires, dont la robustesse reste à démontrer. Ils décrivirent ainsi 8 CNA totalisant 374 gènes associés au phénotype réfractaire, concédant toutefois le manque de correction pour les tests multiples (p valeurs brutes de tests du χ^2 comprises entre 0,003 et 0,05). À ces CNA s'ajoutèrent 16 gènes dont la mutation semblait associée à la rechute, pour la plupart inconnus dans le contexte du lymphome et là encore sans correction des p valeurs. Combinant 4 événements dont l'union couvrait 10 des 13 patients réfractaires (mutation de *FOXO1*, gains en 3q27, 11q23 et 19q13), ils proposèrent ainsi un modèle fortement lié à l'EFS, dont la validité aurait nécessité d'être vérifiée dans une série indépendante.

Nous avons également pu constater la difficulté de mettre en évidence des marqueurs génomiques propres aux LDGCB réfractaires. Faute de moyens suffisants pour séquencer un groupe de patients témoins, nous avons choisi de dresser le profil génomique de nos patients réfractaires pour le comparer ensuite aux résultats publiés dans des cohortes de LDGCB « tout-venant ». Nous avons mis en évidence essentiellement les mêmes gènes et voies métaboliques mutées que dans des séries non réfractaires, en notant tout de même la prépondérance de la voie NFκB, et en ajoutant quelques gènes d'intérêt à la liste des gènes mutés dans les LDGCB. Par ailleurs

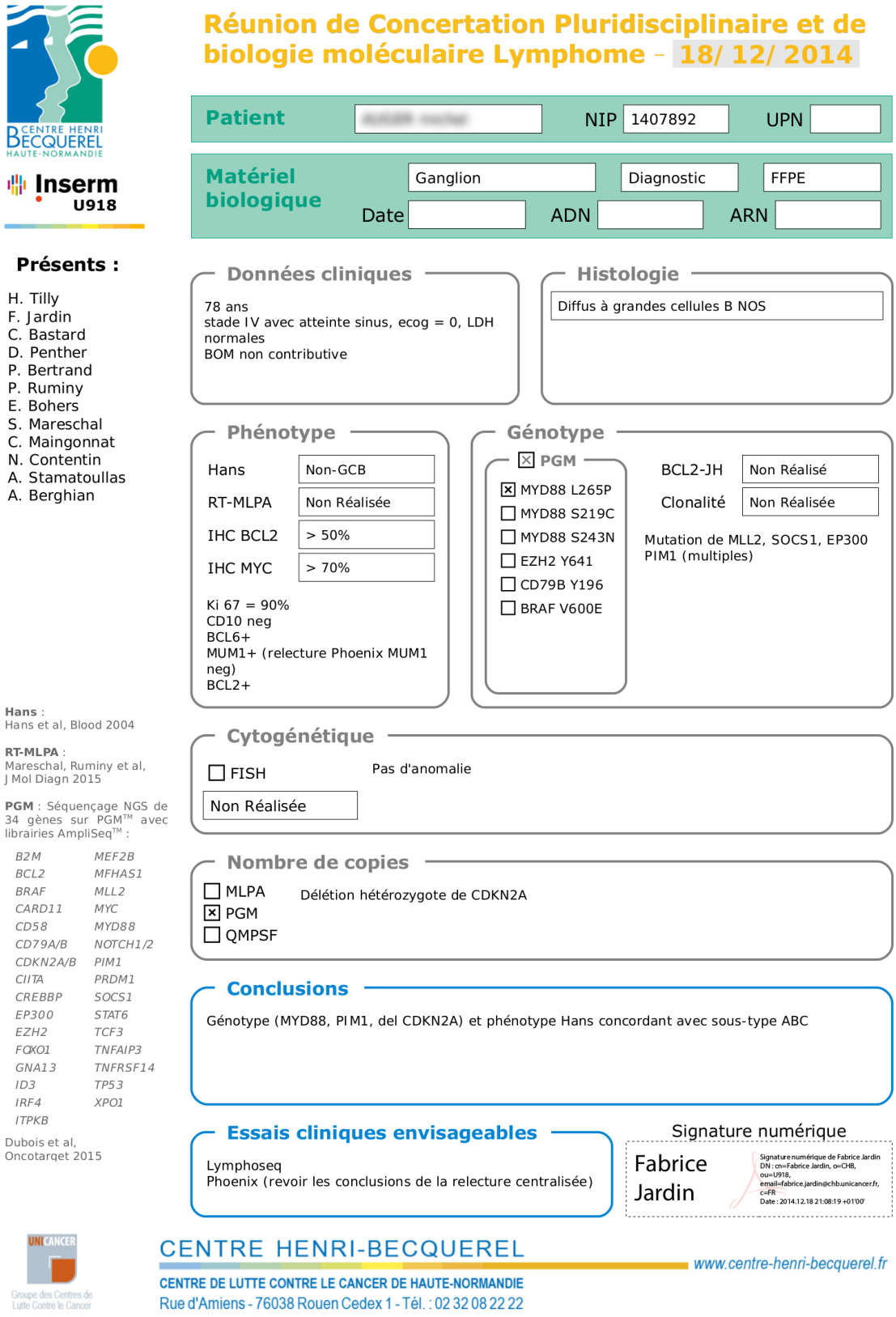
l'extension des analyses de récurrence au niveau des voies métaboliques plutôt que des seuls gènes nous offre une puissance supérieure à Novak *et al* pour effectuer la comparaison entre LDGCB réfractaires ou non. Cependant nous avons fait le choix contrairement à ces auteurs d'analyser séparément les différents sous-types (GCB, ABC, LPM), conscients que le fond biologique différent de ces tumeurs pourrait impliquer des mécanismes de résistance différents.

Globalement ces deux études ont échoué dans leur but d'expliquer le phénotype réfractaire par un nombre limité d'événements somatiques particulièrement récurrents. Ces tumeurs semblent en effet présenter les mêmes altérations que la plupart des LDGCB, selon des combinaisons très variables qui laissent supposer que chaque patient réfractaire ait sa propre histoire. Si certains événements (notamment ceux associés au groupe ABC de mauvais pronostic) semblent plus fréquents chez ces patients, établir avec précision la liste des événements significativement enrichis dans ce phénotype requiert un nombre de patients beaucoup plus important. Réanalyser de manière homogène les différentes séries d'exomes et de génomes publiées ces dernière années (§4.1.2) pourrait potentiellement permettre d'obtenir une telle liste, en les combinant aux données en cours de collecte par des consortiums tels que le TCGA (48 exomes de LDGCB disponibles [296]) et l'ICGC (100 exomes de LDGCB prévus [297]). Il conviendrait toutefois d'obtenir les données cliniques associées à ces patients, et de sélectionner avec soin une cohorte traitée de manière homogène.

6.3.3 Transfert à la pratique clinique

Compte tenu de ces résultats, la prise en charge des patients atteints de LDGCB ne peut se réduire à l'étude d'une poignée de marqueurs somatiques, mais il convient au contraire d'offrir au clinicien un point de vue aussi large que possible sur le génotype de la tumeur. C'est dans cette optique que nous développons au Centre Henri Becquerel un panel de 34 gènes retrouvés mutés ou délétés de manière récurrente dans ces lymphomes (Lymphopanel), et que nous travaillons à son intégration dans la routine clinique du Centre. Le séquençage ciblé peut ainsi être réalisé dès l'obtention d'un diagnostic de LDGCB ou à sa rechute, et conduire dans le mois suivant à l'étude collégiale du cas lors d'une Réunion de Concertation Pluridisciplinaire impliquant cliniciens, anatomopathologistes, biologistes et bioinformaticiens. La confrontation des informations collectées par ces diverses disciplines conduit à la rédaction d'une fiche de synthèse (Figure 6.7), qui permet de pointer les altérations suggestives d'une évolution favorable ou défavorable et les opportunités thérapeutiques qu'elles ouvrent.

La capacité du Lymphopanel d'identifier les événements somatiques relevants pour l'élaboration d'une réponse thérapeutique adaptée dépend directement des



CENTRE HENRI BECQUEREL HAUTE-NORMANDIE

Inserm U918

Présents :

H. Tilly
F. Jardin
C. Bastard
D. Penther
P. Bertrand
P. Ruminy
E. Bohers
S. Mareschal
C. Maingonnat
N. Contentin
A. Stamatoullas
A. Berghian

Réunion de Concertation Pluridisciplinaire et de biologie moléculaire Lymphome – 18/ 12/ 2014

Patient [redacted] NIP 1407892 UPN

Matériel biologique Ganglion | Diagnostic | FFPE
Date | ADN | ARN

Données cliniques

78 ans
stade IV avec atteinte sinus, ecog = 0, LDH normales
BOM non contributive

Histologie

Diffus à grandes cellules B NOS

Phénotype

Hans	Non-GCB
RT-MLPA	Non Réalisée
IHC BCL2	> 50%
IHC MYC	> 70%

Ki 67 = 90%
CD10 neg
BCL6+
MUM1+ (relecture Phoenix MUM1 neg)
BCL2+

Génotype

☒ PGM

- ☒ MYD88 L265P
- ☐ MYD88 S219C
- ☐ MYD88 S243N
- ☐ EZH2 Y641
- ☐ CD79B Y196
- ☐ BRAF V600E

BCL2-JH : Non Réalisé
Clonalité : Non Réalisée

Mutation de MLL2, SOCS1, EP300
PIM1 (multiples)

Cytogénétique

☐ FISH — Pas d'anomalie

Non Réalisée

Nombre de copies

- ☐ MLPA
- ☒ PGM — Délétion hétérozygote de CDKN2A
- ☐ QMPSF

Conclusions

Génotype (MYD88, PIM1, del CDKN2A) et phénotype Hans concordant avec sous-type ABC

Essais cliniques envisageables

Lymphoseq
Phoenix (revoir les conclusions de la relecture centralisée)

Signature numérique
Fabrice Jardin
Signature numérique de Fabrice Jardin
DN : cn=Fabrice Jardin, o=CHB, ou=U918, email=fabrice.jardin@chb.unicancer.fr, c=FR
Date : 2014.12.18 21:08:19 +01'00'

Hans :
Hans et al, Blood 2004

RT-MLPA :
Mareschal, Ruminy et al, J Mol Diagn 2015

PGM : Séquençage NGS de 34 gènes sur PGM™ avec librairies AmpliSeq™ :

B2M	*MEF2B*
BCL2	*MFHAS1*
BRAF	*MLL2*
CARD11	*MYC*
CD58	*MYD88*
CD79A/B	*NOTCH1/2*
CDKN2A/B	*PIM1*
CIITA	*PRDM1*
CREBBP	*SOCS1*
EP300	*STAT6*
EZH2	*TCF3*
FOXO1	*TNFAIP3*
GNA13	*TNFRSF14*
ID3	*TP53*
IRF4	*XPO1*
ITPKB	

Dubois et al, Oncotarget 2015

UNICANCER Groupe des Centres de Lutte Contre le Cancer

CENTRE HENRI-BECQUEREL
www.centre-henri-becquerel.fr
CENTRE DE LUTTE CONTRE LE CANCER DE HAUTE-NORMANDIE
Rue d'Amiens - 76038 Rouen Cedex 1 - Tél. : 02 32 08 22 22

Figure 6.7 – Exemple de fiche de synthèse d'une RCP moléculaire

gènes qu'il cible, et de notre capacité d'associer leurs mutations à un phénotype particulier. C'est donc dans l'optique de parfaire cette liste de cibles que plusieurs projets de séquençage centrés sur des types de lymphomes proches des LDGCB ont été initiés par notre équipe. Le séquençage exomique de 10 LDGCB cutanés primitifs de la jambe a ainsi été réalisé selon les mêmes modalités que pour les LDGCB réfractaires (§4.4), et devrait faire l'objet d'une publication dans les mois à venir. La liste de mutations obtenue souligne la forte ressemblance de ces lymphomes avec les LDGCB ABC, comprenant parmi les événements récurrents des mutations des gènes *MYD88*, *PIM1* ou *CD79B* (Figure 6.8).

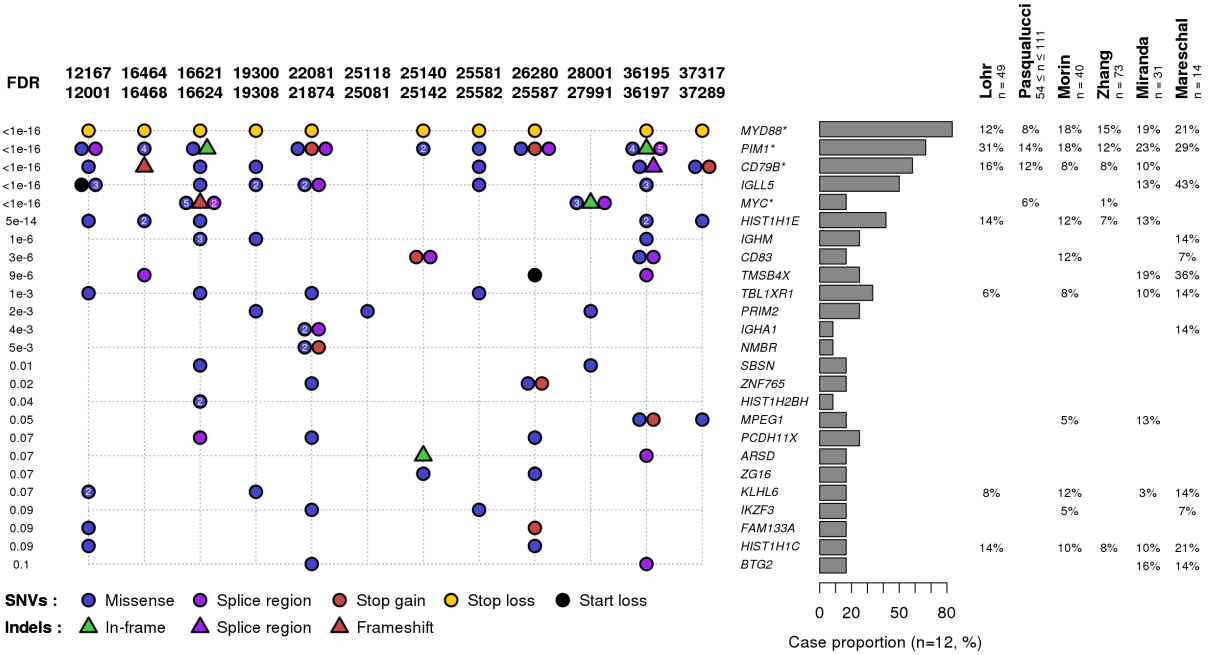


Figure 6.8 – Gènes mutés dans les LDGCB cutanés primitifs de la jambe

Les 10 patients dont les exomes normaux et tumoraux ont été séquencées sont présentés en colonnes, avec les différents gènes retrouvés significativement mutés en lignes. La colonne de gauche présente la significativité du gène (*False Discovery Rate*), les colonnes de droite décrivent les taux de mutation observés dans cette série (histogramme) et dans différentes séries publiées d'exomes de LDGCB ganglionnaires. Chaque type de mutation est représenté par un pictogramme, incluant éventuellement un nombre dans le cas où un même type de mutation a été retrouvé plusieurs fois.

L'étude des LDGCB primitifs du SNC a également été réalisée, avec le Lymphopanel cette fois-ci. Les tumeurs primitives de 30 patients ont ainsi fait l'objet d'un séquençage ciblé [383], révélant là encore un sous-type fortement ABC : mutations de *MYD88* (77%), *PIM1* (37%), CD79B (33%) et *TNFAIP3* (20%), délétions de *CDKN2A/B* (77%). La plupart de ces mutations ayant été retrouvées dans l'ADN circulant du sang périphérique, ces résultats ouvrent la voie à une caractérisation génomique et un diagnostic moins invasifs de ces tumeurs, dont la localisation cérébrale n'autorise pas toujours l'obtention d'une biopsie exploitable.

Le séquençage de cette « biopsie liquide » présente également des intérêts dans les LDGCB ganglionnaires, comme nous avons pu le souligner grâce au Lymphopanel [384]. La qualité du séquençage reste supérieure dans la biopsie primaire, et il serait dommage de s'en priver dans les cas où l'exérèse est partie intégrante de la prise en charge thérapeutique, néanmoins le suivi d'un petit groupe de mutations caractéristiques de la tumeur par séquençage « ultra-profond » de l'ADN circulant pourrait devenir une méthode de suivi de la maladie résiduelle surpassant en sensibilité la plupart des techniques actuelles [385]. A plus long terme, l'application du Lymphopanel complet sur ADN plasmatique pourrait permettre d'obtenir une vision d'ensemble chez un patient, particulièrement précieuse pour les cas multi-métastasés où plusieurs sites distants acquièrent des mutations indépendamment.

A Annexes

A.1 Revue publiée (Leuk. Lymphoma 2015)

Suite à la publication d'une série de 161 LDGCB ayant fait l'objet d'un séquençage Sanger de 5 gènes [130], une revue a été commandée à notre laboratoire par le journal *Leukemia & Lymphoma* (IF 2,89) [386]. Cette revue se focalise sur les mutations activatrices récurrentes dans les LDGCB (*MYD88*, *CARD11*, *EZH2*, *CD79A* et *CD79B*), et les opportunités thérapeutiques qui ont été explorées ou restent encore à l'être concernant ces gènes et les mécanismes sous-jacents. Elle se conclue par la recommandation d'un séquençage ciblé systématique dans cette pathologie, dans une optique de médecine personnalisée, qui a été implémentée dans notre laboratoire sous forme du « Lymphopanel » [271, 272].

Leukemia & Lymphoma, May 2015; 56(5): 1213–1222
© 2014 Informa UK, Ltd.
ISSN: 1042-8194 print / 1029-2403 online
DOI: 10.3109/10428194.2014.941836

informa
healthcare

REVIEW

Activating somatic mutations in diffuse large B-cell lymphomas: lessons from next generation sequencing and key elements in the precision medicine era

Elodie Bohers[1], Sylvain Mareschal[1], Philippe Bertrand[1], Pierre Julien Viailly[1], Sydney Dubois[1], Catherine Maingonnat[1], Philippe Ruminy[1], Hervé Tilly[1,2] & Fabrice Jardin[1,2]

[1]*INSERM U918, Centre Henri Becquerel, Rouen University and IRIB, Rouen, France and* [2]*Department of Hematology, Centre Henri Becquerel, Rouen, France*

Downloaded by [85.171.81.87] at 00:43 22 October 2015

Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common form of lymphoma, accounting for 30–40% of newly diagnosed non-Hodgkin lymphomas. Historically, DLBCL has been thought to involve recurrent translocations of the immunoglobulin heavy (IGH) locus and the deregulation of rearranged oncogenes. Whole exome sequencing (WES) of more than 200 DLBCLs has completely redefined the genetic landscape of the disease by identifying recurrent single nucleotide variants and providing new therapeutic opportunities in DLBCL molecular subtypes. Some of these somatic mutations target genes that play a crucial role in B-cell function (B cell receptor [BCR] signaling, nuclear factor κB [NF-κB] pathway, Toll-like receptor [TLR] signaling and phosphatidylinositol 3-kinase [PI3K] pathway), immunity, cell cycle/apoptosis or chromatin modification. In this review, following an overview of the somatic mutations reported in DLBCL, we focus on activating and clustered mutations targeting genes including *MYD88*, *CD79A/B*, *EZH2* and *CARD11* and discuss their clinical and therapeutic relevance in the precision medicine era.

Keywords: NGS, diffuse large B-cell lymphoma, somatic mutations, targeted therapy

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common form of lymphoma, accounting for 30–40% of newly diagnosed non-Hodgkin lymphomas (NHLs). The molecular heterogeneity of DLBCLs has been deciphered by gene expression profiling (GEP), and DLBCLs are currently divided into three main molecular subtypes: the germinal center B-cell like (GCB) subtype, the activated B-cell like (ABC) subtype and the primary mediastinal B-cell lymphoma (PMBL) subtype. These subtypes arise from distinct B cells at separate stages of differentiation and maturation, leading to well-defined GEPs and different clinical outcomes and responses to immunochemotherapy. The GCB subtype is characterized by a more favorable outcome and a spectrum of genetic aberrations, which include the t(14;18)(q32;q21) translocation, deletion of *PTEN*, amplification of the miR-17-92 cluster or *TP53* mutations. By contrast, the ABC subtype has a less favorable outcome and is characterized by a distinct genetic background that includes the t(3;14)(q27;q32) translocation, trisomy 3, deletion of the *INK4A-ARF* locus or *BCL2* amplification [1]. PMBL displays a strong molecular relationship with Hodgkin lymphoma (HL) and is characterized by the amplification of *JAK2* in nearly half of cases and the recurrent deletion of *SOCS1*, a suppressor of JAK signaling, both of which contribute to activation of the JAK2–STAT6 (Janus kinase 2–signal transducer and activator of transcription 6) signaling pathway [2].

Next generation sequencing (NGS) technologies, which allow for massively parallel, high-throughput DNA sequencing, have emerged over the past decade and have provided new insights into the genomic characterization of DLBCL and mature B-cell malignancies in general [3–7]. To date, publication of the whole exome sequences (WESs) of approximately 220 DLBCLs and related subtypes has completely redefined the DLBCL genetic landscape by identifying recurrent single nucleotide variants (SNVs). Despite the tremendous genetic heterogeneity that underlies DLBCL, a few recurrent mutations are of especial interest because they occur in "druggable" targets and/or are strongly correlated with antitumoral response to targeted therapies [8]. These recurrent mutations are activating, and lead directly or indirectly to pathway activation, including B-cell receptor (BCR) and Toll-like receptor (TLR) signaling, the nuclear factor κB (NF-κB) pathway, phosphatidylinositol 3-kinase (PI3K) and mitogen activated protein (MAP) kinase signaling or more global epigenetic modifications. In this article we give a brief overview of the SNVs recently

Correspondence: Fabrice Jardin, Centre Henri Becquerel, Rouen, France. E-mail: fabrice.jardin@chb.unicancer.fr
Received 29 May 2014; revised 19 June 2014; accepted 30 June 2014

discovered by NGS in DLBCL. Regarding the great diversity of somatic mutations in DLBCL, we have chosen to focus our analysis on five genes recurrently activated by somatic mutations, namely *MYD88*, *CARD11*, *EZH2* and *CD79A/CD79B*. Mutations targeting these genes occur in approximately 40% of cases of DLBCL, cluster in functional domain coding regions and are considered drivers during lymphomagenesis. Finally we discuss their relevance in the current or future daily practice in the precision medicine era.

The landscape of somatic mutations in diffuse large B-cell lymphoma: an overview

To date, more than 200 DLBCL exomes have been sequenced and published. Data have mainly been obtained by HiSeq, the most widely adopted platform (Illumina, San Diego, CA), and include DNA samples with matched non-tumoral DNA, multiple cell lines and the targeted resequencing of larger sample collections (Table I). The Cancer Genome Atlas (TCGA) project was performed across 21 tumor types, including 58 DLBCL tumors, and indicated that the median somatic mutation frequency in DLBCL is 3.3 mutations per Mb (for comparison, melanoma, the most mutated tumor type, displayed 12.9 mutations/Mb) [9]. The number of samples needed to detect significantly mutated genes can be calculated as a function of a tumor's median background mutation frequency. The current sample size of published WES for DLBCL reliably detects genes mutated at 5% or more above background. According to this projection, creating a comprehensive catalog of genes in which somatic mutations occur at a low frequency ($<5\%$) will require analyzing an average of approximately 500 DLBCL samples. This indicates that while we have yet to reach near-saturation, a complete comprehensive catalog of DLBCL genes will be obtained in the near future. However, the pattern of mutations in DLBCL is highly heterogeneous from patient to patient and also between published studies, as shown by a relatively modest overlap of 10–20% among four different WES studies [10]. As mentioned above, compared with cancers in which environmental mutagens are known to increase the mutation burden, such as melanoma and lung cancer, the rate of somatic mutations reported in DLBCL is relatively low. Mutation feature analyses indicate that some recurrent genes are targeted as a result of the somatic mutation process driven by activation-induced cytidine deaminase (AICD) and the failure of the high-fidelity DNA repair system [11,12]. More than 40 recurrently mutated genes displayed hallmarks of AICD mutation activity, including an excess of transitions over transversions, an excess of C/T$>$T/A transitions and an enrichment in some targeted motifs [13]. It is therefore challenging in this somatic mutation background to distinguish driver from passenger mutations. Several hotspot mutations and clustered mutations are strong indicators of positive selection. Some of these somatic mutation target genes play a crucial role in B-cell function (BCR signaling, NF-κB pathway, TLR signaling), immunity, cell cycle/apoptosis, NOTCH signaling or epigenetic regulation [9,10,14–22]. It is possible to group or summarize most genetic alterations into fewer than 20 oncogenic signaling pathways [10]. Figure 1 summarizes the distribution of selected somatic mutations recently identified by NGS in DLBCL subtypes [9,10,14–16,18–26]. Of note, rare mutations are, to date, only reported in a particular DLBCL subtype, including mutations targeting *PTPN1* in PMBL and *GNA13* in GCB-DLBCL, and may be used as diagnostic markers. Recently we performed Sanger sequencing of five genes, namely *MYD88*, *CD79A/CD79B*, *CARD11* and *EZH2*, in a large cohort of DLBCL [27]. Mutation characteristics targeting these genes provide evidence of driver mutations including hotspot, clustering in functional domains and change of protein structure. Their biological and clinical relevance is detailed in the following section.

MYD88 mutations in DLBCL

The most frequent SNVs found in ABC-DLBCL are somatic mutations in *MYD88*, which occurs in approximately one-third of cases [18,27]. A low mutation rate observed in a Korean-based population of DLBCL suggests a possible inter-ethnic rate variation [28]. MYD88 activates the NF-κB and JAK–STAT3 signaling pathways after stimulation of TLRs by their ligands [18]. The *MYD88* gene encodes an adaptor protein consisting of an N-terminal Death domain, a linker region and a C-terminal Toll-interleukin 1 receptor (TIR)

Table I. NGS studies published in DLBCL and different DLBCL subtypes.

Reference	Cohort	Key reported variants	Sequencing method/NGS platform	Mutation recurrence testing cohort
[16]	31	*EZH2* mutations	WGS, RNA-Seq/Illumina GAIIx instrument	320 DLBCL
[21]	Two cell lines	*CIITA* fusion genes	RNA-Seq/Illumina GAIIx instrument	263 NHL
[22]	7	*CREBBP/EP300* mutations	WES/Genome Sequencer FLX instrument; 454 Life Sciences	134 DLBCL
[18]	4	*MYD88* mutations	RNA-Seq/Illumina GAIIx instrument	382 DLBCL
[15]	14	*MLL2–MEF2B, CREBBP*, *EP300* mutations	WES–WGS–RNA-Seq/Illumina GAIIx, HiSeq 2000 instrument	83 DLBCL
[20]	49	*PCLO*, *TNFRSF14* mutations	WES/Illumina GAIIx instrument	—
[19]	6	*MLL2* mutations	WES/Genome Sequencer FLX Instrument; 454 Life Sciences	115 DLBCL
[26]	4	*TBLXR1*, *MYD88, PIM1*	WES/Illumina GAIIx instrument	29 PCNSL
[10]	34	*PI3KCD* mutations	WES–WGS/Illumina GAIIx instrument	39 DLBCL
[14]	40	*GNA13* mutations	WGS–RNA-Seq/Illumina HiSequation platform	279 DLBCL
[9]	58	~5000 WES from 21 various cancers	WES/Illumina HiSequation platform	—
[25]	2	*PTPN1* mutations	WGS/Illumina HiSequation platform	220 PMBL

NGS, next generation sequencing; DLBCL, diffuse large B-cell lymphoma; NHL, non-Hodgkin lymphoma; PCNSL, primary central nervous system lymphoma; PMBL, primary mediastinal B-cell lymphoma; WES, whole exome sequencing; WGS, whole genome sequencing.

Downloaded by [85.171.81.87] at 00:43 22 October 2015

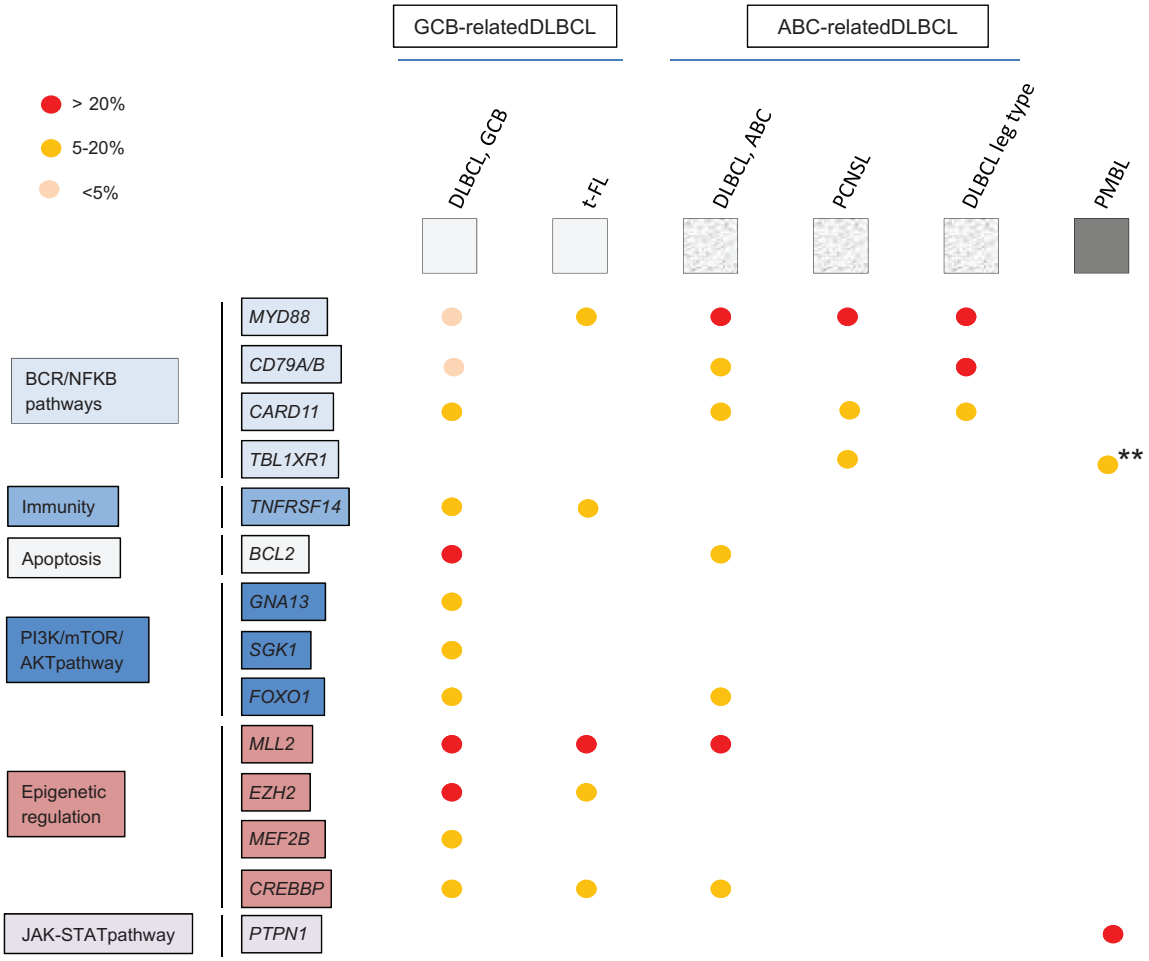


Figure 1. Selected panel of recurrently mutated genes identified by NGS in diffuse large B-cell lymphoma and related histopathological/molecular subtypes. DLBCL, diffuse large B-cell lymphoma; PCNSL, primary central nervous system lymphoma; PMBL, primary mediastinal B-cell lymphoma; t-FL, transformed follicular lymphoma. *TBLXR1 is recurrently involved in a TBLXR1–TP63 fusion transcript in PMBL. Circles are indicative of the mutation frequency.

Downloaded by [85.171.81.87] at 00:43 22 October 2015

domain, which interacts with the TIR domain of the TLR (Figure 2). The most prevalent mutation (an L265P substitution) targets the evolutionarily conserved B-B loop of the TIR domain. *In vitro* experiments indicate that this mutation induces rapid B cell division in the absence of exogenous TLR ligands, but is inhibited by inactivation of TLR9, suggesting the persistence of a partial dependence on upstream TLR activation. They also indicate that proliferation and NF-κB activation induced by *MYD88* L265P require additional genetic events, including the repression of TNFAIP3 (A20), an NF-κB pathway inhibitor that is frequently inactivated in ABC-DLBCL, and *BCL2* overexpression [29,30].

MYD88 is also recurrently targeted by somatic mutations in other DLBCL subtypes that predominantly display an ABC-related GEP. This includes primary cutaneous large B-cell lymphoma, leg type (DLBCL-LT), identified in the World Health Organization (WHO) classification as a DLCBL subtype characterized by the high expression of BCL2, MUM1 and FOXP1, indicative of an ABC phenotype [31]. DLBCL-LT displays frequent genetic alterations in the NF-κB pathway that resemble those observed in nodal ABC-DLBCL. Targeted sequencing of a relatively small cohort highlighted the elevated rate of mutation in the *CD79B*, *MYD88* and *CARD11* genes. Somatic *MYD88* L265P mutations have been reported in up to 70% of cases of DLBCL-LT [32–34]. Primary central nervous system DLBCL (PCNSL), which shares a similarly predominant ABC phenotype with DLBCL-LT, also displays a high rate of *MYD88* mutations. Studies searching for genomic alterations via NGS are scarce. WES has recently been performed in four human immunodeficiency virus (HIV)-negative patients with PCNSL in combination with comparative genomic hybridization/single nucleotide polymorphism (CGH/SNP) arrays. This approach identified novel somatic mutations in the *MYD88* and *TBLXR1* genes in 38% and 14% of cases, respectively [26]. A recent report from the American Society of Hematology (ASH) 2013 meeting confirmed these preliminary results in an extended cohort, with *MYD88*-activating mutations found in 69% of cases [35]. In keeping with these observations, *MYD88* mutations are found at a high frequency in other immune-privileged extranodal sites. An *MYD88* mutation at position L265P is observed in 68% (27/37) of primary testicular lymphomas (PTLs) [36].

Recent NGS data obtained from follicular lymphoma (FL) and subsequently matched transformed-FL (t-FL) yield insight into the oncogenic somatic events that may contribute to FL transformation to aggressive DLCBL. Among these

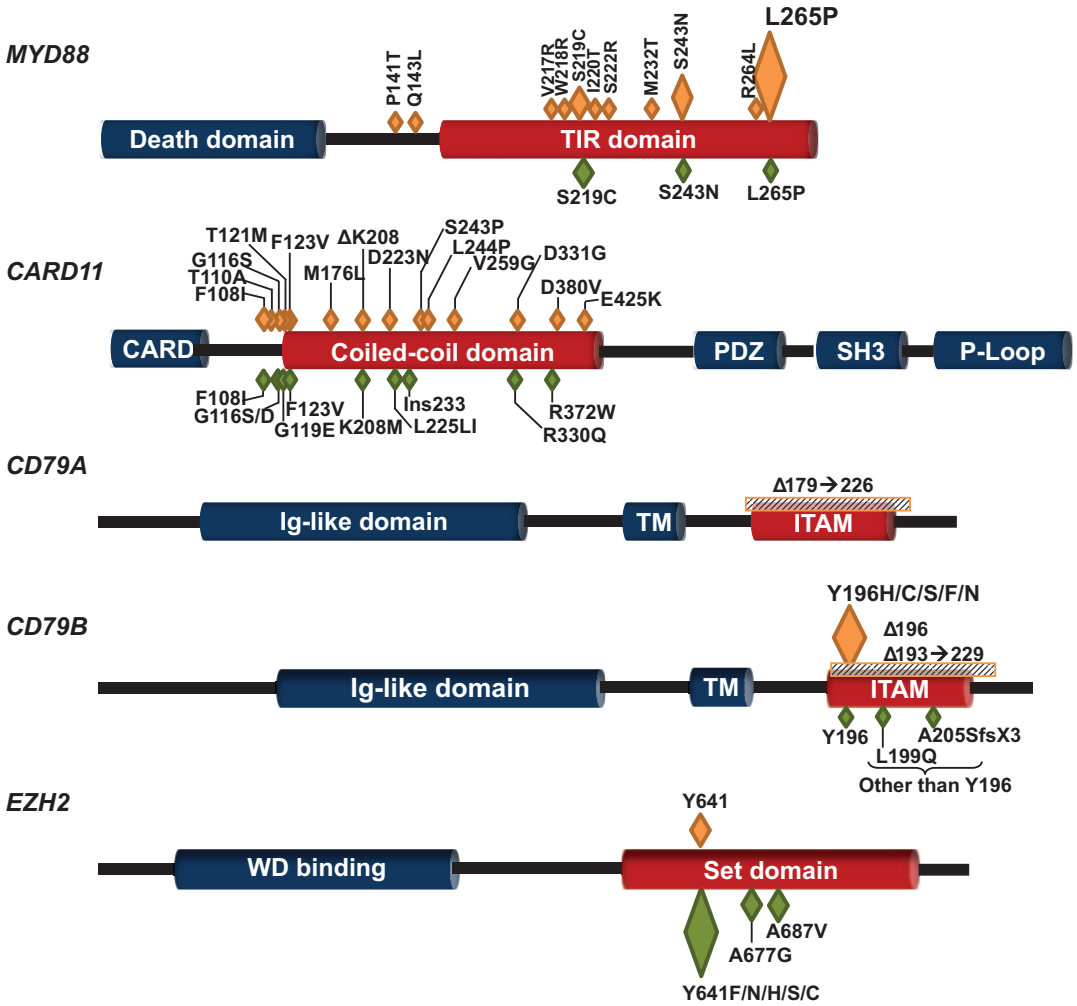


Figure 2. Structures of *CD79*, *MYD88*, *CARD11* and *EZH2* genes recurrently targeted by somatic mutations in DLBCL. Schematic representation of the domains and main mutations are indicated. Each diamond represents the position of the mutation with size proportional to the frequency of the mutation in ABC (orange) or GCB (green) subtypes. The shaded areas represent the deletions. According to references [15,16,18,28,43,49].

somatic events, acquisition of MYD88 SNV may be critical, since these mutations are observed exclusively in t-FL but not in the corresponding initial FL tumor, suggesting the possible role of MYD88 activation during the transformation process. By contrast, longitudinal analyses identified mutations in chromatin regulator genes (*CREBBP*, *EZH2* and *MLL2*) as early drivers [37].

Patients with DLCBL with *MYD88* mutations are significantly older than those without these mutations, consistent with the higher frequency of these mutations in cases of ABC-DLBCL and the increase in proportion of cases of ABC-DLBCL with age [27,28,38–40]. The *MYD88* L265P substitution does not seem to be a significant prognostic indicator by itself for DLBCL, but protein expression, regardless of the *MYD88* genetic status, may be significantly associated with tumor recurrence and shortened disease-free survival (DFS) [28,40]. While the *MYD88* L265P mutation is the hallmark of immunoglobulin M (IgM) lymphoplasmacytic lymphoma, IgM secretion is not a typical feature of *MYD88*-mutated DLBCL [6,41,42]. MYD88 protein, as detected by immunohistochemistry, appears to be overexpressed in approximately 40% of cases with DLBCL, mainly of the ABC subtype, and localized in the cytoplasm of lymphoma cells [28].

CD79A/CD79B and *CARD11* mutations in DLBCL

Mutations of the BCR proximal adaptors *CD79B/CD79A* have been identified in approximately 16–18% of patients, mostly within the ABC-DLBCL subtype [27,43]. A *CD79B* mutation is also found in 19% (7/37) of cases of PTL and coexists with *MYD88* variants in most cases [36] and in 20% of DLBCL-LT [34]. Conversely, targeted sequencing indicates that HIV-DLBCL displays a distinct SNV profile with an absence of mutations in *CD79A/CD79B* [44]. Mutations target the functional intracellular immunoreceptor tyrosine-based activation motif (ITAM); they favor surface BCR expression and reduce the activation of negative regulators of BCR signaling [43]. In *CD79B*, the majority of substitutions altered the first tyrosine (Y196) of the ITAM motif, reducing the negative regulation provided by the kinase LYN. In *CD79A*, one recurrent alteration deleted a large part of the ITAM domain, and altered a negative feedback loop mechanism for BCR activation driven by the phosphorylation of serine 97 by the kinase SYK [27,45] (Figure 2).

Downloaded by [85.171.81.87] at 00:43 22 October 2015

Downloaded by [85.171.81.87] at 00:43 22 October 2015

In normal B-cells, antigen receptor-induced NF-κB activation requires *CARD11*, a cytoplasmic scaffolding protein. The pathogenic driver status of *CARD11* alterations is supported by the discovery of gain-of-function germline mutations that drive constitutive NF-κB activation in congenital B-cell lymphocytosis [46]. In contrast with the initial report, *CARD11* mutants are equally distributed between the DLBCL subtypes, occurring in approximately 10% of nodal DLBCLs [10,19,27,30,47]. *CARD11* mutations are also observed in 16% of cases of PCNSL, 10% of DLBCL-LT and 10% of gastric DLBCL [34,48,49]. Several oncogenic *CARD11* mutations that enhance activity have been identified in DLBCL, and are mostly located in the coiled-coil (C-C) domain. Oncogenic mutations induce CARD11 hyperactivity by disrupting auto-inhibition by the CARD11 inhibitory domain (ID), leading to receptor-independent activation of NF-κB [50]. Mutations targeting other regulatory domains interacting with the ID domain, including the CARD and the LATCH domains, favor BCL10 interaction and NF-κB activation [51]. Recent experiments suggest that *CARD11* mutations may be implicated in early oncogenic events that drive B-cell transformation by favoring the switch from self-antigen-induced B-cell death into T-cell independent proliferation and plasmablast differentiation [52].

Histone methyltransferase enhancer of Zeste homolog 2 (*EZH2*) mutations

NGS data have reported an overlap of recurrent somatic mutations between FL and DLBCL, including recurrent somatic mutations in the *EZH2* histone methyltransferase, which encodes a subunit of polycomb-repressive complex 2 (PRC2) and is responsible for the trimethylation of lysine 27 of histone H3 (H3K27me3) [16]. The *EZH2* gene is mutated in 11% of FLs and up to 23% of GCB-DLBCLs [16]. The most frequent mutation is the heterozygous change of tyrosine residue 641 (Y641) in exon 16, which encodes the catalytic site of the SET domain (Figure 2), resulting in a mutant protein that acts synergistically with the wild-type enzyme to increase histone H3 trimethylation [53,54]. Alternative variants, A687V or A677G, occur in approximately 1–2% of patients with lymphoma and have been shown to be gain-of-function mutations [55,56]. *EZH2* is highly expressed in GC B-cells, and the conditional expression of mutant *EZH2* in mice induces GC hyperplasia and accelerated lymphomagenesis in cooperation with *BCL2* [57]. The Y641 variant affects the interaction of *EZH2* with a ubiquitin ligase (β-TrCP), which explains the increased *EZH2* stability and H3K27me3 hyperactivity [58]. No prognostic impacts of *EZH2* mutations in GCB-DLBCLs have been observed in distinct series of DLCBLs. In contrast, a recent report indicated that high-level expression of EZH2 (EZH2 ≥ 70%), detected by immunohistochemistry, was associated with superior overall survival (OS), but this study warrants further confirmation [59].

Activating mutations and therapeutic opportunities in diffuse large B-cell lymphoma

Because all the aforementioned mutations induce gain of function, they hold promising and already tested targeted therapeutic opportunities, and can be used to guide individualized DLBCL treatments.

Targeting *MYD88* and TLR signaling

Therapies targeting *MYD88* components, including IRAK kinase inhibitors either as single agents or in combination with drugs blocking BCR signaling, are promising and already being employed in certain strategies [60]. Many pharmacological strategies are in preclinical and clinical development [Figure 3(b)]. *MYD88* is a key linker protein in the TLR 7, 8 and 9 signaling pathway. IMO-8400, an oligonucleotide specifically designed to inhibit TLRs 7, 8 and 9, is currently being tested in phase I/II clinical trials for patients with relapsed or refractory Waldenstrom macroglobulinemia (number NCT02092909 at ClinicalTrials.gov). Other promising strategies include targeting the IRAK4 and TAK1 mediator proteins of *MYD88* signaling and homodimerization for the targeted therapy of *MYD88*-mutated DLBCLs [61,62]. However, growing evidence indicates that crosstalk between the BCR and TLR signaling components may contribute to lymphomagenesis, indicating that a dual blocking strategy would be part of an efficient therapeutic strategy [63].

Targeting BCR signaling-related mutations

The BCR signaling pathway is crucial for the development of normal B-cells and the pathogenesis of B-cell malignancies, and targeting this pathway may present a new treatment paradigm for many B-cell malignancies [64]. The BCR complex integrates a membrane immunoglobulin (IgM) that is non-covalently linked to CD79A and CD79B. Following IgM ligation, the cytoplasmic ITAM domains of CD79A and CD79B are phosphorylated by the Src family member LYN, leading to SYK activation and a subsequent PI3K–Bruton's tyrosine kinase (BTK) signaling cascade. After the Y551 and Y223 residues have been phosphorylated, BTK is able to form a complex with BLNK, promoting the production of DAG and IP3, which leads to the activation of protein kinase Cβ (PKCβ), the CARD11/BCL10/ MALT1 (CBM) complex and NF-κB signaling [Figure 3(a)].

In vivo experiments indicate that *CD79A/CD79B* mutant cells are highly sensitive to sotrastaurin (AEB071), a protein kinase C inhibitor, whereas *CARD11* mutations render ABC-DLBCL cell lines insensitive to the molecule [65]. Sotrastaurin is currently being tested as a single agent and in combination with the mammalian target of rapamycin (mTOR) inhibitor everolimus in patients with the CD79-mutant or the ABC-subtype of DLBCL (number NCT01854606 at ClinicalTrials.gov).

Ibrutinib, a BTK inhibitor, showed a 29% overall response rate (ORR) in a phase I clinical trial in DLBCLs. A subsequent phase II clinical trial demonstrated preferential activity for the ABC subtype compared to the GCB subtype of DLBCL, which showed response rates of 40% and 5%, respectively ($p = 0.01$) [66]. However, ibrutinib may also be active in GCB-DLBCLs and more active in *MYD88*-mutated ABC-DLBCLs, but not in those with *CARD11* mutations [67]. Interestingly, ibrutinib interacted favorably with a wide range of agents, including inhibitors of the PI3K–AKT pathway, other B-cell receptor pathway inhibitors, Bcl-2 family inhibitors and

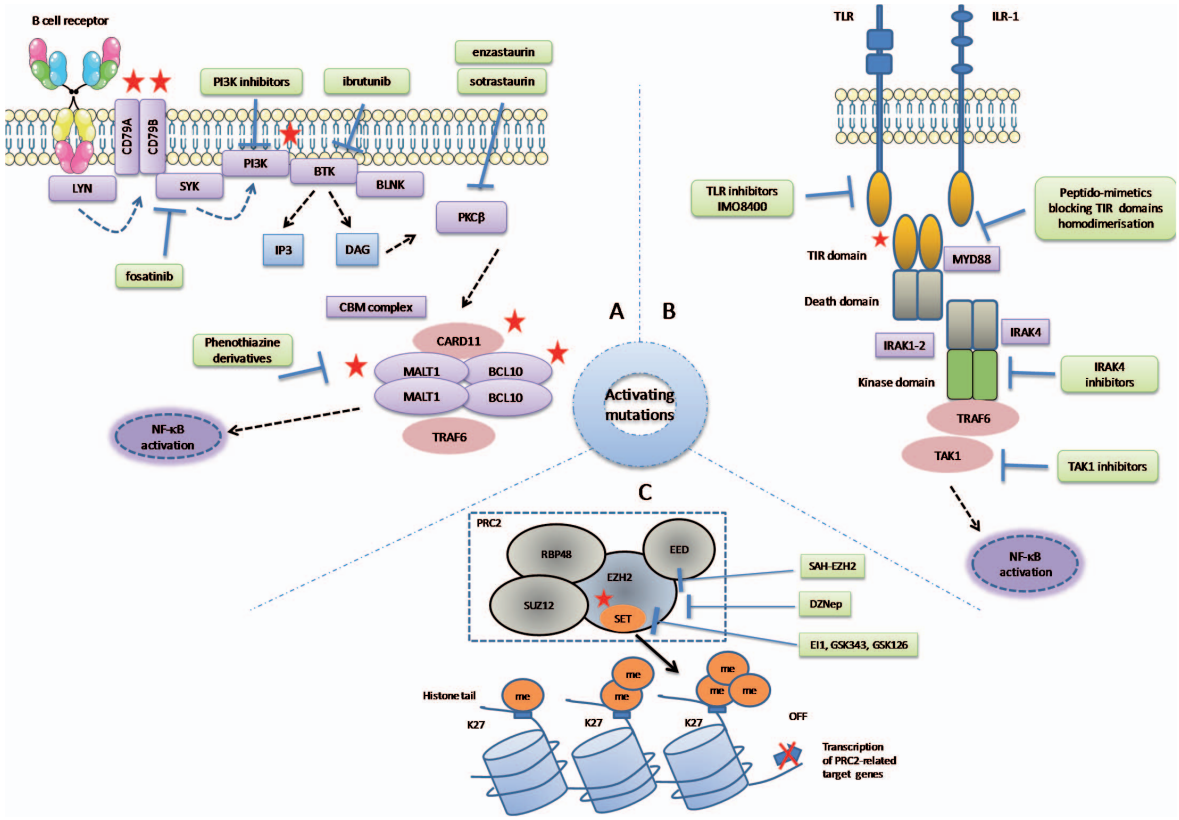


Figure 3. Therapeutic opportunities in patients with DLBCL with activating CD79A/B, EZH2, CARD11 or MYD88 variants. Recurrently mutated targeted genes in DLBCL are indicated by a star. Potential or effective therapeutic strategies are indicated in green squares. (A) CD79A/B, CARD11 and B cell receptor pathway. Following IgM ligation, the cytoplasmic ITAM domains of CD79A and CD79B are phosphorylated by the Src family member LYN, leading to SYK activation and subsequent PI3K–BTK signaling cascade. After phosphorylation of Tyr 551 and Tyr223 residues, BTK constitutes a complex with BLNK, promoting the production of DAG and IP3, activated PKCβ to activate the CARD11/BCL10/ MALT1 (CBM) complex and finally NF-κB signaling. Drugs targeting the BCR cascade including SYK, BTK, PI3K, PKCβ or CBM complex are indicated. (B) MYD88 and TLR signaling. MYD88 is a key linker protein in the signaling pathway of TLRs 7, 8, 9 and IL1-R. After ligand stimulation, TLR/IL1-R interacts with MYD88 by their Toll/interleukin-1 receptor (TIR) domain. MYD88 subsequently binds IRAK1 by its Death domain (DD), which in turn phosphorylates and activates IRAK4. In association with TRAF6 and TAK1, the IRAK1/IRAK4/MYD88 complex activates the NF-κB pathway. Strategies targeting TLR, IRAK4 and TAK1 mediator proteins of MYD88 signaling or MYD88 homodimerization are in clinical and preclinical development. (C) *EZH2* mutation and PRC2 complex-related pathway. The polycomb repressive complex (PRC2), which includes EED, SUZ12, RbAP46/48 and EZH2, catalyzes H3K27 di- and tri-methylation. The SET domain of EZH2 represents the catalytic domain of the PRC2 complex. After tri-methylation, the chromatin is condensed leading to PRC2 target gene silencing. EZH2 inhibition, leading to gene transcription, can be obtained by *S*-adenosyl homocysteine hydrolase inhibitor 3-deazaneplanocin A (DZNep), which induces PRC2 component depletion, by several agents that block the SET catalytic domain interacting with the *S*-adenosylmethionine (SAM) pocket (EI1, GSK343, GSK126), or by disruption of the EED–EZH2 complex (SAH–EZH2).

Downloaded by [85.171.81.87] at 00:43 22 October 2015

several standard cytotoxic agents used against DLBCLs [68]. Fostamatinib, a SYK inhibitor, also showed activity against relapsed DLBCL in a phase I/II study, and additional phase II trials are now under way [69]. Newly developed MALT1 small-molecule inhibitors suppress the growth of ABC-DLBCLs *in vitro* and *in vivo*. Recent progress in targeting the MALT1 proteolytic activity indicates that MALT1 inhibitors are a promising strategy for the treatment of B-cell lymphomas [70–72].

Targeting the PI3K pathway in *CD79A/CD79B* mutants

The PI3K pathway is active in many primary DLBCL tumor samples, independent of classification. Class IA/IB PI3Ks convert phosphatidylinositol-4,5-diphosphates to phosphatidylinositol-3,4,5-triphosphates, leading to activation of the effector kinases PDK1 (putative 3-phosphoinositide-dependent kinase 1) and AKT [Figure 3(c)]. For cancer therapy, clinical trials with different strategies have been initiated using many compounds that target the PI3K signaling pathway, including dual PI3K and mTOR inhibitors, pan-PI3K inhibitors or isoform-specific PI3K inhibitors [73–75]. The genetic mechanisms involved in the PI3K activation pathway likely influence the sensitivity to PI3K inhibition. PI3K and PDK1 have been shown to play a critical role in cell growth and survival in a subset of ABC-DLBCL cell lines characterized by *CD79B* mutations, suggesting that PI3K inhibition may be a promising strategy for treating aggressive lymphomas with *CD79B* mutations [76]. Of note, several mutations in PI3K pathway components, including *FOXO1*, *PIK3CD*, *PIK3R1* and *mTOR*, may also increase sensitivity to PI3K or mTOR inhibition [10]. Furthermore, *PTEN* (phosphatase and tensin homolog) deletions and mutations that lead to reduced or absent PTEN expression are observed in nearly 40% of GCB-DLBCLs, resulting in activation of the PI3K pathway and sensitivity to the corresponding inhibitors [77].

Downloaded by [85.171.81.87] at 00:43 22 October 2015

Targeting *EZH2* mutations

EZH2, which catalyzes H3K27 trimethylation, is a promising and attractive target to fight lymphoma [78]. *EZH2* represses proliferation checkpoint genes and helps to transiently suppress GC B-cell differentiation. These physiological effects are reinforced by acquired somatic mutations in DLBCLs, suggesting that GCB-DLBCLs are mostly affected by *EZH2* mutations, unlike the more differentiated ABC subtype [57]. Direct inhibition of EZH2 methyltransferase activity by agents such as GSK126, a potent and highly selective small-molecule inhibitor of EZH2 methyltransferase activity that competes for binding with *S*-adenosyl-methionine, has inhibited proliferation in both *EZH2* mutant DLBCL cell lines and *EZH2* mutant DLBCL xenografts in mice. *EZH2* targeting is associated with a global decrease in the H3K27me3 level and reactivates silenced PRC2 target genes [79,80]. Phase II clinical trials to evaluate different EZH2 inhibitors are currently ongoing in patients with relapsed or refractory DLBCL and t-FL (numbers NCT01897571, NCT02082977 at ClinicalTrials.gov), and considerable efforts to design specific EZH2/EZH1 inhibitor molecules are currently ongoing [81–84]. In contrast, a more global action based on DNA demethylation may be of interest in treating DLBCL, as chemoresistance is associated with aberrant DNA methylation programming. Prolonged exposure to low-dose DNA methyltransferase (DNMT) inhibitors reprogrammed chemoresistant cells to become doxorubicin-sensitive. Consistent with this fact, 5-azacytidine priming prior to R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone) has been shown to be feasible and results in global demethylation, restoration of the transforming growth factor-β (TGF-β) pathway and improved chemotherapy sensitivity in patients with newly diagnosed DLBCL [85].

Other recurrent mutations in DLBCL with potential pharmacological relevance

Between the several mutated genes and the inherent genetic heterogeneity of DLBCLs, a few SNVs are of especial interest, as the corresponding proteins are the object of pharmacological development. This includes PIM1, a member of a class of serine/threonine kinases with distinct molecular and biochemical features that regulate various oncogenic pathways, including B-cell development [86]. Inhibition of PIM1 suppressed cell proliferation and migration, induced apoptotic cell death and synergized with other chemotherapeutic agents. Ongoing trials in prostate cancers and B-cell lymphomas are under way [87].

RNA-sequencing (RNA-Seq) data obtained from a large series of more than 300 DLBCLs have shown that *BCL2* is the most highly mutated gene in GCB-DLBCL. Mutations are preferentially located in the flexible loop domain, sparing the BH domains, and result from somatic hypermutation [88]. BCL2 inhibitors such as ABT-199 may play a role in BCL2-overexpressing GCB tumors. The impact of somatic *BCL2* mutations in the BH domain is still uncertain, but may confer resistance to the BH3 mimetic ABT-199 [89,90]. Recurrent somatic coding-sequence mutations of the *PTPN1* gene have been reported in 22% of primary mediastinal B-cell lymphomas (PMBLs) [25]. Mutations consisted mainly of nonsense, missense and frameshift changes and deletions within a hotspot in the first exon. Mutations led to reduced phosphatase activity of the PTP1B protein, increasing the phosphorylation activity of the JAK–STAT pathway. These mutations likely synergize with other known driver mutations reported in PMBLs, including mutations in *SOCS1* and *STAT6*, and may identify patients who will benefit from JAK inhibitors [2,91,92]. Other genes are rarely mutated in DLBCLs but may be beneficial to targeted therapies, including *BRAF*, *PDGFRA*, *MTOR*, *KIT*, *IDH1*, *GNA13*, *SGK1*, the *IAP* gene family or *NOTCH* [10].

Routine targeted sequencing in diffuse large B-cell lymphoma

Recent advances in high-throughput DNA sequencing have provided a vast and comprehensive catalog of cancer genes [93]. NGS offers a great opportunity to identify drug targets and guide the therapeutic strategy, supporting the concept of precision medicine for DLBCL. To reach this goal, a limited and relevant panel of genes must be designed and compact sequencers must be used, currently widely available in most academic laboratories, including the most popular Ion Personal Genome Machine (PGM) (Life Technologies) and MiSeq (Illumina). Several clinical trials to test targeted therapies based on genomic profiling are currently ongoing [94]. NGS technology platforms dedicated to targeted resequencing provide sensitive evaluation of the tumor genotype in fewer than 3 working days, as well as additional information in comparison to the standard Sanger method [95]. This includes minimal residual assessment and longitudinal sampling of the tumor or blood-derived circulating tumor DNA, information concerning clonal architecture according to variant allele frequency (VAF) and gene copy number variation (CNV) abnormalities [93,95–98]. An example of DLBCL genome analysis obtained by routine NGS with the PGM is shown in Figure 4.

Interestingly, NGS can be performed from both formalin-fixed, paraffin embedded (FFPE) and fresh-frozen specimens. Preliminary studies have suggested that DNA damage caused by fixation was supposed to be involved in 1% of observed SNVs. However, more recently, direct comparison between paired matched fresh-frozen and FFPE tumor specimens showed a similar error rate between the sample sources, indicating that FFPE tissues can be used in routine NGS processing [99,100]. Furthermore, the minimum required DNA amount (10 ng) for targeted NGS by some platforms enables the analysis of fine-needle aspiration specimens, extending the field of NGS usage to the vast majority of real-life patients with DLBCL [101]. The growing and rapid development of novel agents (some examples are indicated in Figure 3) targeting oncogenic drivers and their usage in a clinical trial setting or in daily practice currently requires a minimal molecular characterization of the DLBCL tumor. This includes, at the very least, determination of the cell of origin subtype, MYC rearrangement status and the somatic mutation profile of a gene panel with theranostic value. The complexity of NGS data interpretation implies integrating molecular biologists in our routine hospital medical

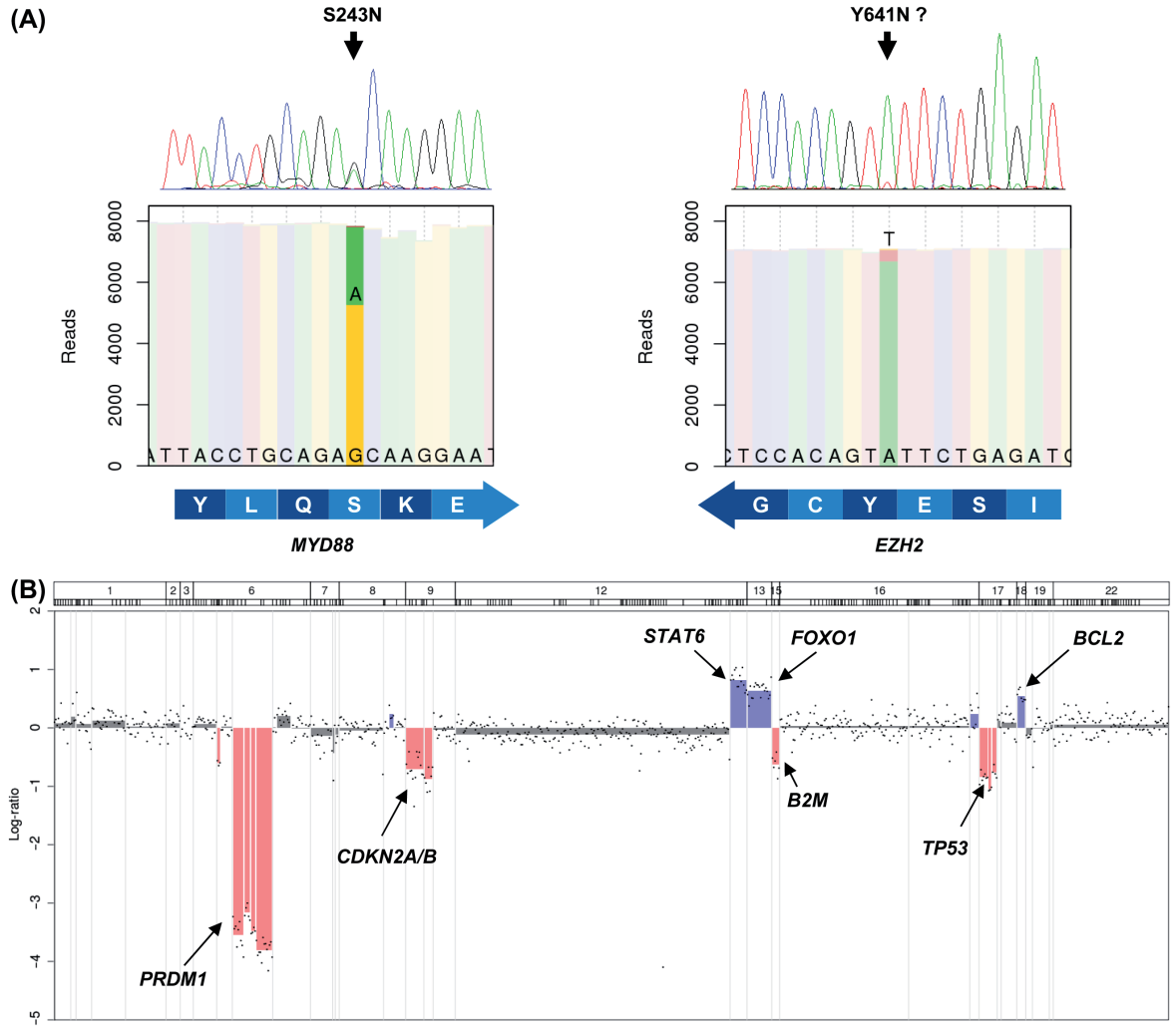


Figure 4. Example of somatic profile obtained by targeted sequencing using Personal Genome Machine (PGM) in routine diagnosis. (A) Example of a case of GCB-DLBCL harboring a heterozygous *MYD88* mutation and an *EZH2* mutation at low allelic frequency. The Sanger sequencing profiles in the upper panels are presented aligned to the histograms of the nucleotides found in the PGM sequencing reads in the middle panels, and the translated amino acids of the protein of interest in the lower panels. The depths obtained with PGM enable confirmation of the subclonal *EZH2* mutation. (B) Gene copy number variations of an ABC-DLBCL sample showing *PRDM1*, *CDKN2A/B*, *B2M* and *TP53* deletions (in pink) and gains of copy of *STAT6*, *FOXO1* and *BCL2* (in blue). Gene copy number abnormalities are detected by comparing the sequencing depths between sample and normal control with a dedicated algorithm (personal laboratory data, INSERM U918).

Downloaded by [85.171.81.87] at 00:43 22 October 2015

committees, as currently done for pathologists or nuclear physicians.

Conclusion

The characterization of DLBCL somatic mutations is still in progress, but has thus far revealed tremendous heterogeneity and a limited number of cancer genes that are shared with other cancers [93]. Incorporating this comprehensive characterization of somatic gene mutations to assist in the daily practice of managing DLBCL still presents a challenge but will be widely used in the near future. The first challenge is to establish a comprehensive mutation profile of genes that are targetable or "actionable" in a clinical setting. A consensus lymphopanel still must be defined, and should incorporate somatic mutations with proven diagnostic, prognostic and/or theranostic relevance. According to these criteria, a lymphopanel of approximately 30 genes should be sufficiently informative to tailor genotype-specific treatments. Of note, despite the likely physiopathological relevance, the prognostic value of these mutations has not been properly evaluated in prospective clinical trials, and the potential of these mutations to be resistant to conventional immunochemotherapy in B-cell lymphoma treatment is unknown. Finally, clinical trials with therapies directed against well-defined targets assessed by targeted NGS but also in combination with other relevant biomarkers, such as GEP routine methods, will accelerate the development of drugs and therapeutic strategies in DLBCL.

Potential conflict of interest: Disclosure forms provided by the authors are available with the full text of this article at www.informahealthcare.com/lal.

References

[1] Lenz G, Staudt LM. Aggressive lymphomas. N Engl J Med 2010;362:1417–1429.
[2] Steidl C, Gascoyne RD. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. Blood 2011;118:2659–2669.
[3] Mullighan CG. Genome sequencing of lymphoid malignancies. Blood 2013;122:3899–3907.
[4] Kridel R, Meissner B, Rogic S, et al. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 2012;119:1963–1971.
[5] Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. N Engl J Med 2011;364:2305–2315.
[6] Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. N Engl J Med 2012;367:826–833.
[7] Rossi D, Trifonov V, Fangazio M, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med 2012;209:1537–1551.
[8] Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nat Rev Clin Oncol 2014;11:12–23.
[9] Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495–501.
[10] Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci USA 2013;110: 1398–1403.
[11] Gu X, Shivarov V, Strout MP. The role of activation-induced cytidine deaminase in lymphomagenesis. Curr Opin Hematol 2012; 19:292–298.
[12] Liu M, Schatz DG. Balancing AID and DNA repair during somatic hypermutation. Trends Immunol 2009;30:173–181.
[13] Khodabakhshi AH, Morin RD, Fejes AP, et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. Oncotarget 2012; 3:1308–1319.
[14] Morin RD, Mungall K, Pleasance E, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 2013;122:1256–1265.
[15] Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298–303.
[16] Morin RD, Johnson NA, Severson TM, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 2010;42:181–185.
[17] Scott DW, Mungall KL, Ben-Neriah S, et al. TBL1XR1/TP63: a novel recurrent gene fusion in B-cell non-Hodgkin lymphoma. Blood 2012;119:4949–4952.
[18] Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature 2011;470:115–119.
[19] Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 2011;43:830–837.
[20] Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci USA 2012;109:3879–3884.
[21] Steidl C, Shah SP, Woolcock BW, et al. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature 2011;471:377–381.
[22] Pasqualucci L, Dominguez-Sola D, Chiarenza A, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature 2011;471:189–195.
[23] Ying CY, Dominguez-Sola D, Fabi M, et al. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol 2013;14:1084–1092.
[24] Trinh DL, Scott DW, Morin RD, et al. Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. Blood 2013;121:3666–3674.
[25] Gunawardana J, Chan FC, Telenius A, et al. Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. Nat Genet 2014;46:329–335.
[26] Gonzalez-Aguilar A, Idbaih A, Boisselier B, et al. Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. Clin Cancer Res 2012;18:5203–5211.
[27] Bohers E, Mareschal S, Bouzelfen A, et al. Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell lymphoma. Genes Chromosomes Cancer 2014;53:144–153.
[28] Choi JW, Kim Y, Lee JH, et al. MYD88 expression and L265P mutation in diffuse large B-cell lymphoma. Hum Pathol 2013;44: 1375–1381.
[29] Wang JQ, Jeelall YS, Beutler B, et al. Consequences of the recurrent MYD88(L265P) somatic mutation for B cell tolerance. J Exp Med 2014;211:413–426.
[30] Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature 2009;459:717–721.
[31] Willemze R, Jaffe ES, Burg G, et al. WHO-EORTC classification for cutaneous lymphomas. Blood 2005;105:3768–3785.
[32] Pham-Ledard A, Cappellen D, Martinez F, et al. MYD88 somatic mutation is a genetic feature of primary cutaneous diffuse large B-cell lymphoma, leg type. J Invest Dermatol 2012;132:2118–2120.
[33] Pham-Ledard A, Prochazkova-Carlotti M, Andrique L, et al. Multiple genetic alterations in primary cutaneous large B-cell lymphoma, leg type support a common lymphomagenesis with activated B-cell-like diffuse large B-cell lymphoma. Mod Pathol 2014;27:402–411.
[34] Koens L, Zoutman WH, Ngarmlertsirichai P, et al. Nuclear factor-kappaB pathway-activating gene aberrancies in primary cutaneous large B-cell lymphoma, leg type. J Invest Dermatol 2014;134:290–292.
[35] Braggio E, O'Neill B, Van Wier S, et al. The genomic landscape of primary central nervous system lymphomas. Blood 2013;122(Suppl.1): Abstract 504.
[36] Kraan W, van Keimpema M, Horlings HM, et al. High prevalence of oncogenic MYD88 and CD79B mutations in primary testicular diffuse large B-cell lymphoma. Leukemia 2014;28:719–720.
[37] Okosun J, Bodor C, Wang J, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat Genet 2014;46:176–181.
[38] Mareschal S, Lanic H, Ruminy P, et al. The proportion of activated B-cell like subtype among de novo diffuse large B-cell lymphoma increases with age. Haematologica 2011;96:1888–1890.
[39] Kraan W, Horlings HM, van Keimpema M, et al. High prevalence of oncogenic MYD88 and CD79B mutations in diffuse large B-cell lymphomas presenting at immune-privileged sites. Blood Cancer J 2013;3:e139.
[40] Kim Y, Ju H, Kim DH, et al. CD79B and MYD88 mutations in diffuse large B-cell lymphoma. Hum Pathol 2014;45:556–564.
[41] Gachard N, Parrens M, Soubeyran I, et al. IGHV gene features and MYD88 L265P mutation separate the three marginal zone lymphoma entities and Waldenstrom macroglobulinemia/lymphoplasmacytic lymphomas. Leukemia 2013;27:183–189.
[42] Cox MC, Di Napoli A, Scarpino S, et al. Clinicopathologic characterization of diffuse-large-B-cell lymphoma with an associated serum monoclonal IgM component. PLoS One 2014;9:e93903.
[43] Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature 2010;463:88–92.
[44] Capello D, Gloghini A, Martini M, et al. Mutations of CD79A, CD79B and EZH2 genes in immunodeficiency-related non-Hodgkin lymphomas. Br J Haematol 2011;152:777–780.
[45] Heizmann B, Reth M, Infantino S. Syk is a dual-specificity kinase that self-regulates the signal output from the B-cell antigen receptor. Proc Natl Acad Sci USA 2010;107:18563–18568.
[46] Snow AL, Xiao W, Stinson JR, et al. Congenital B cell lymphocytosis explained by novel germline CARD11 mutations. J Exp Med 2012;209:2247–2261.
[47] Lenz G, Davis RE, Ngo VN, et al. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. Science 2008;319:1676–1679.
[48] Montesinos-Rongen M, Schmitz R, Brunn A, et al. Mutations of CARD11 but not TNFAIP3 may activate the NF-kappaB pathway in primary CNS lymphoma. Acta Neuropathol 2010;120:529–535.
[49] Dong G, Chanudet E, Zeng N, et al. A20, ABIN-1/2, and CARD11 mutations and their prognostic value in gastrointestinal diffuse large B-cell lymphoma. Clin Cancer Res 2011;17:1440–1451.
[50] Lamason RL, McCully RR, Lew SM, et al. Oncogenic CARD11 mutations induce hyperactive signaling by disrupting autoinhibition by the PKC-responsive inhibitory domain. Biochemistry 2010;49: 8240–8250.
[51] Chan W, Schaffer TB, Pomerantz JL. A quantitative signaling screen identifies CARD11 mutations in the CARD and LATCH domains that induce Bcl10 ubiquitination and human lymphoma cell survival. Mol Cell Biol 2013;33:429–443.
[52] Jeelall YS, Wang JQ, Law HD, et al. Human lymphoma mutations reveal CARD11 as the switch between self-antigen-induced B cell death or proliferation and autoantibody production. J Exp Med 2012;209:1907–1917.
[53] Yap DB, Chu J, Berg T, et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered

Downloaded by [85.171.81.87] at 00:43 22 October 2015

PRC2 catalytic activity, to increase H3K27 trimethylation. Blood 2011;117:2451–2459.
[54] Sneeringer CJ, Scott MP, Kuntz KW, et al. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci USA 2010;107:20980–20985.
[55] Majer CR, Jin L, Scott MP, et al. A687V EZH2 is a gain-of-function mutation found in lymphoma patients. FEBS Lett 2012;586:3448–3451.
[56] McCabe MT, Graves AP, Ganji G, et al. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). Proc Natl Acad Sci USA 2012;109:2989–2994.
[57] Beguelin W, Popovic R, Teater M, et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell 2013;23:677–692.
[58] Sahasrabuddhe AA, Chen X, Chung F, et al. Oncogenic Y641 mutations in EZH2 prevent Jak2/beta-TrCP-mediated degradation. Oncogene 2014 Jan 27. [Epub ahead of print]
[59] Lee HJ, Shin DH, Kim KB, et al. Polycomb protein EZH2 expression in diffuse large B-cell lymphoma is associated with better prognosis in patients treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone. Leuk Lymphoma 2014;55:2056–2063.
[60] Loiarro M, Ruggiero V, Sette C. Targeting the Toll-like receptor/interleukin 1 receptor pathway in human diseases: rational design of MyD88 inhibitors. Clin Lymphoma Myeloma Leuk 2013;13:222–226.
[61] Loiarro M, Volpe E, Ruggiero V, et al. Mutational analysis identifies residues crucial for homodimerization of myeloid differentiation factor 88 (MyD88) and for its function in immune cells. J Biol Chem 2013;288:30210–30222.
[62] Ansell SM, Hodge LS, Secreto FJ, et al. Activation of TAK1 by MYD88 L265P drives malignant B-cell Growth in non-Hodgkin lymphoma. Blood Cancer J 2014;4:e183.
[63] Dufner A, Schamel WW. B cell antigen receptor-induced activation of an IRAK4-dependent signaling pathway revealed by a MALT1-IRAK4 double knockout mouse model. Cell Commun Signal 2011;9:6.
[64] Niemann CU, Wiestner A. B-cell receptor signaling as a driver of lymphoma development and evolution. Semin Cancer Biol 2013;23:410–421.
[65] Naylor TL, Tang H, Ratsch BA, et al. Protein kinase C inhibitor sotrastaurin selectively inhibits the growth of CD79 mutant diffuse large B-cell lymphomas. Cancer Res 2011;71:2643–2653.
[66] Advani RH, Buggy JJ, Sharman JP, et al. Bruton tyrosine kinase inhibitor ibrutinib (PCI-32765) has significant activity in patients with relapsed/refractory B-cell malignancies. J Clin Oncol 2013;31:88–94.
[67] Zheng X, Ding N, Song Y, et al. Different sensitivity of germinal center B cell-like diffuse large B cell lymphoma cells towards ibrutinib treatment. Cancer Cell Int 2014;14:32.
[68] Mathews Griner LA, Guha R, Shinn P, et al. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. Proc Natl Acad Sci USA 2014;111:2349–2354.
[69] Friedberg JW, Sharman J, Sweetenham J, et al. Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia. Blood 2010;115:2578–2585.
[70] Fontan L, Melnick A. Molecular pathways: targeting MALT1 paracaspase activity in lymphoma. Clin Cancer Res 2013;19:6662–6668.
[71] Hailfinger S, Lenz G, Ngo V, et al. Essential role of MALT1 protease activity in activated B cell-like diffuse large B-cell lymphoma. Proc Natl Acad Sci USA 2009;106:19946–19951.
[72] Nagel D, Spranger S, Vincendeau M, et al. Pharmacologic inhibition of MALT1 protease by phenothiazines as a therapeutic approach for the treatment of aggressive ABC-DLBCL. Cancer Cell 2012;22:825–837.
[73] Jabbour E, Ottmann OG, Deininger M, et al. Targeting the phosphoinositide 3-kinase pathway in hematologic malignancies. Haematologica 2014;99:7–18.
[74] Choi MY, Kipps TJ. Inhibitors of B-cell receptor signaling for patients with B-cell malignancies. Cancer J 2012;18:404–410.
[75] Lannutti BJ, Meadows SA, Herman SE, et al. CAL-101, a p110delta selective phosphatidylinositol-3-kinase inhibitor for the treatment of B-cell malignancies, inhibits PI3K signaling and cellular viability. Blood 2011;117:591–594.
[76] Kloo B, Nagel D, Pfeifer M, et al. Critical role of PI3K signaling for NF-kappaB-dependent survival in a subset of activated B-cell-like diffuse large B-cell lymphoma cells. Proc Natl Acad Sci USA 2011;108:272–277.
[77] Abubaker J, Bavi PP, Al-Harbi S, et al. PIK3CA mutations are mutually exclusive with PTEN loss in diffuse large B-cell lymphoma. Leukemia 2007;21:2368–2370.
[78] Melnick A. Epigenetic therapy leaps ahead with specific targeting of EZH2. Cancer Cell 2012;22:569–570.
[79] McCabe MT, Ott HM, Ganji G, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature 2012;492:108–112.
[80] Knutson SK, Kawano S, Minoshima Y, et al. Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. Mol Cancer Ther 2014;13:842–854.
[81] Garapaty-Rao S, Nasveschuk C, Gagnon A, et al. Identification of EZH2 and EZH1 small molecule inhibitors with selective impact on diffuse large B cell lymphoma cell growth. Chem Biol 2013;20:1329–1339.
[82] Diaz E, Machutta CA, Chen S, et al. Development and validation of reagents and assays for EZH2 peptide and nucleosome high-throughput screens. J Biomol Screen 2012;17:1279–1292.
[83] Woo J, Kim HY, Byun BJ, et al. Biological evaluation of tanshindiols as EZH2 histone methyltransferase inhibitors. Bioorg Med Chem Lett 2014;24:2486–2492.
[84] Kim W, Bird GH, Neff T, et al. Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. Nat Chem Biol 2013;9:643–650.
[85] Clozel T, Yang S, Elstrom RL, et al. Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. Cancer Discov 2013;3:1002–1019.
[86] Baron BW, Anastasi J, Hyjek EM, et al. PIM1 gene cooperates with human BCL6 gene to promote the development of lymphomas. Proc Natl Acad Sci USA 2012;109:5735–5739.
[87] Arunesh GM, Shanthi E, Krishna MH, et al. Small molecule inhibitors of PIM1 kinase: July 2009 to February 2013 patent update. Expert Opin Ther Pat 2014;24:5–17.
[88] Schuetz JM, Johnson NA, Morin RD, et al. BCL2 mutations in diffuse large B-cell lymphoma. Leukemia 2012;26:1383–1390.
[89] Scarfo L, Ghia P. Reprogramming cell death: BCL2 family inhibition in hematological malignancies. Immunol Lett 2013;155:36–39.
[90] Fresquet V, Rieger M, Carolis C, et al. Acquired mutations in BCL2 family proteins conferring resistance to the BH3 mimetic ABT-199 in lymphoma. Blood 2014;123:4111–4119.
[91] Mullally A, Ebert BL. Janus reveals another face: the biologic rationale for targeting Janus kinase 2 in lymphoma. J Clin Oncol 2012;30:4168–4170.
[92] Younes A, Romaguera J, Fanale M, et al. Phase I study of a novel oral Janus kinase 2 inhibitor, SB1518, in patients with relapsed lymphoma: evidence of clinical and biologic activity in multiple lymphoma subtypes. J Clin Oncol 2012;30:4161–4167.
[93] Watson IR, Takahashi K, Futreal PA, et al. Emerging patterns of somatic mutations in cancer. Nat Rev Genet 2013;14:703–718.
[94] Thomas F, Desmedt C, Aftimos P, et al. Impact of tumor sequencing on the use of anticancer drugs. Curr Opin Oncol 2014;26:347–356.
[95] Hullein J, Jethwa A, Stolz T, et al. Next-generation sequencing of cancer consensus genes in lymphoma. Leuk Lymphoma 2013;54:1831–1835.
[96] Bodor C, Grossmann V, Popov N, et al. EZH2 mutations are frequent and represent an early event in follicular lymphoma. Blood 2013;122:3165–3168.
[97] Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. Bioinformatics 2012;28:1307–1313.
[98] Saieg MA, Geddie WR, Boerner SL, et al. EZH2 and CD79B mutational status over time in B-cell non-Hodgkin lymphomas detected by high-throughput sequencing using minimal samples. Cancer Cytopathol 2013;121:377–386.
[99] Spencer DH, Sehn JK, Abel HJ, et al. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn 2013;15:623–633.
[100] Kerick M, Isau M, Timmermann B, et al. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC Med Genomics 2011;4:68.
[101] Hadd AG, Houghton J, Choudhary A, et al. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. J Mol Diagn 2013;15:234–247.

Downloaded by [85.171.81.87] at 00:43 22 October 2015

Bibliographie

1. Abbas, A. K., Lichtman, A. H. & Pillai, S. *Cellular and molecular immunology* 6e éd. isbn : 978-1-4160-3122-2 978-0-8089-2358-9 (Saunders Elsevier, Philadelphia, 2007).
2. Nagasawa, T. The chemokine CXCL12 and regulation of HSC and B lymphocyte development in the bone marrow niche. *Adv. Exp. Med. Biol.* **602,** 69–75 (2007).
3. Bryder, D. & Sigvardsson, M. Shaping up a lineage–lessons from B lymphopoesis. *Curr. Opin. Immunol.* **22,** 148–153 (avr. 2010).
4. Blom, B. & Spits, H. Development of human lymphoid cells. *Annu. Rev. Immunol.* **24,** 287–320 (2006).
5. Clark, M. R., Mandal, M., Ochiai, K. & Singh, H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nat. Rev. Immunol.* **14,** 69–80 (fév. 2014).
6. Chung, J. B., Silverman, M. & Monroe, J. G. Transitional B cells : step by step towards immune competence. *Trends Immunol.* **24,** 343–349 (juin 2003).
7. Steiniger, B., Timphus, E. M. & Barth, P. J. The splenic marginal zone in humans and rodents : an enigmatic compartment and its inhabitants. *Histochem. Cell Biol.* **126,** 641–648 (déc. 2006).
8. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15,** 137–148 (mar. 2015).
9. National Cancer Institute sponsored study of classifications of non-Hodgkin's lymphomas : summary and description of a working formulation for clinical usage. The Non-Hodgkin's Lymphoma Pathologic Classification Project. *Cancer* **49,** 2112–2135 (mai 1982).
10. Gérard-Marchant, R *et al.* Classification of non-Hodgkin's lymphomas. *Lancet* **2,** 405–408 (août 1974).
11. Stansfeld, A. G. *et al.* Updated Kiel classification for lymphomas. *Lancet* **1,** 292–293 (fév. 1988).
12. Chan, J. K. *et al.* A revised European-American classification of lymphoid neoplasms proposed by the International Lymphoma Study Group. A summary version. *Am. J. Clin. Pathol.* **103,** 543–560 (mai 1995).
13. Harris, N. L. *et al.* The World Health Organization classification of neoplastic diseases of the haematopoietic and lymphoid tissues : Report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997. *Histopathology* **36,** 69–86 (jan. 2000).
14. Swerdlow, S. H., International Agency for Research on Cancer & World Health Organization. *WHO classification of tumours of haematopoietic and lymphoid tissues* isbn : 978-92-832-2431-0 92-832-2431-0 (International Agency for Research on Cancer, Lyon, France, 2008).
15. Monnereau, A *et al. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012 : étude à partir des registres des cancers du réseau Francim* isbn : 978-2-11-138316-6 978-2-11-138347-0 (Institut de veille sanitaire, Saint-Maurice, 2013).
16. Treon, S. P. *et al.* MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N. Engl. J. Med.* **367,** 826–833 (août 2012).
17. Camara, D. A. *et al.* Immunoblastic morphology in diffuse large B-cell lymphoma is associated with a non-germinal center immunophenotypic profile. *Leuk. Lymphoma* **48,** 892–896 (mai 2007).
18. Xu-Monette, Z. Y. *et al.* Clinical and biological significance of de novo CD5+ diffuse large B-cell lymphoma in Western countries. *Oncotarget* **6,** 5615–5633 (mar. 2015).
19. Nakamura, T. *et al.* Recurrent mutations of CD79B and MYD88 are the hallmark of primary central nervous system lymphomas. *Neuropathol. Appl. Neurobiol.* doi :10.1111/nan.12259 (juin 2015).
20. Gebauer, N. *et al.* Prevalence of targetable oncogenic mutations and genomic alterations in Epstein-Barr virus-associated diffuse large B-cell lymphoma of the elderly. *Leuk. Lymphoma* **56,** 1100–1106 (avr. 2015).
21. Bonnet, C. *et al.* Prise en charge du lymphome B diffus à grandes cellules en 2012. *Rev Med Suisse* **8,** 1582–1584, 1586–1590 (août 2012).
22. Lenz, G. *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 13520–13525 (sept. 2008).
23. Dunleavy, K. *et al.* Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. *N. Engl. J. Med.* **368,** 1408–1416 (avr. 2013).
24. Tilly, H. *et al.* Diffuse large B-cell lymphoma (DLBCL) : ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **26 Suppl 5,** v116–v125 (sept. 2015).
25. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* **89,** 3909–3918 (juin 1997).
26. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N. Engl. J. Med.* **329,** 987–994 (sept. 1993).
27. Ziepert, M. *et al.* Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J. Clin. Oncol.* **28,** 2373–2380 (mai 2010).
28. Sehn, L. H. *et al.* The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood* **109,** 1857–1861 (mar. 2007).
29. Nguyen, V. X., Nguyen, C. C. & Nguyen, B. D. 18F-FDG PET/CT imaging of the pancreas : spectrum of diseases. *JOP* **12,** 557–566 (nov. 2011).
30. Cheson, B. D. *et al.* Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma : the Lugano classification. *J. Clin. Oncol.* **32,** 3059–3068 (sept. 2014).
31. Maurer, M. J. *et al.* Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J. Clin. Oncol.* **32,** 1066–1073 (avr. 2014).

32. Tarella, C. *et al.* Rate of primary refractory disease in B and T-cell non-Hodgkin's lymphoma : correlation with long-term survival. *PLoS ONE* **9,** e106745 (2014).

33. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470 (oct. 1995).

34. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14,** 457–460 (déc. 1996).

35. Golub, T. R. *et al.* Molecular classification of cancer : class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537 (oct. 1999).

36. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511 (fév. 2000).

37. Alizadeh, A. *et al.* The lymphochip : a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harb. Symp. Quant. Biol.* **64,** 71–78 (1999).

38. Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346,** 1937–1947 (juin 2002).

39. Hakim, R. B. F., Subanar & Winarko, E. in *Database Theory and Application, Bio-Science and Bio-Technology* (éds. Zhang, Y. *et al.*) *Communications in Computer and Information Science* 118, 207–225 (Springer Berlin Heidelberg, 2010). isbn : 978-3-642-17621-0 978-3-642-17622-7.

40. Wright, G. *et al.* A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9991–9996 (août 2003).

41. Lenz, G *et al.* Stromal gene signatures in large-B-cell lymphomas. *N. Engl. J. Med.* **359,** 2313–2323 (nov. 2008).

42. Shaffer, A. L., Young, R. M. & Staudt, L. M. Pathogenesis of human B cell lymphomas. *Annu. Rev. Immunol.* **30,** 565–610 (2012).

43. Lossos, I. S. *et al.* Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 10209–10213 (août 2000).

44. Mittrücker, H. W. *et al.* Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* **275,** 540–543 (jan. 1997).

45. Davis, R. E., Brown, K. D., Siebenlist, U. & Staudt, L. M. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* **194,** 1861–1874 (déc. 2001).

46. Dunleavy, K. *et al.* Differential efficacy of bortezomib plus chemotherapy within molecular subtypes of diffuse large B-cell lymphoma. *Blood* **113,** 6069–6076 (juin 2009).

47. Milhollen, M. A. *et al.* MLN4924, a NEDD8-activating enzyme inhibitor, is active in diffuse large B-cell lymphoma models : rationale for treatment of NF-{kappa}B-dependent lymphoma. *Blood* **116,** 1515–1523 (sept. 2010).

48. *A Study of the Bruton's Tyrosine Kinase Inhibitor, PCI-32765 (Ibrutinib), in Combination With Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Patients With Newly Diagnosed Non-Germinal Center B-Cell Subtype of Diffuse Large B-Cell Lymphoma (PHOENIX)* mai 2013. <https://clinicaltrials.gov/ct2/show/NCT01855750> (visité le 11/08/2015).

49. Hans, C. P. *et al.* Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103,** 275–282 (jan. 2004).

50. Ott, G. *et al.* Immunoblastic morphology but not the immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL. *Blood* **116,** 4916–4925 (déc. 2010).

51. Gutiérrez-García, G. *et al.* Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood* **117,** 4836–4843 (mai 2011).

52. Benesova, K. *et al.* The Hans algorithm failed to predict outcome in patients with diffuse large B-cell lymphoma treated with rituximab. *Neoplasma* **60,** 68–73 (2013).

53. Read, J. A. *et al.* Evaluating cell-of-origin subtype methods for predicting diffuse large B-cell lymphoma survival : a meta-analysis of gene expression profiling and immunohistochemistry algorithms. *Clin Lymphoma Myeloma Leuk* **14,** 460–467.e2 (déc. 2014).

54. Coutinho, R. *et al.* Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for diffuse large B-cell lymphoma : implications for therapeutic strategies. *Clin. Cancer Res.* **19,** 6686–6695 (déc. 2013).

55. Choi, W. W. L. *et al.* A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin. Cancer Res.* **15,** 5494–5502 (sept. 2009).

56. Meyer, P. N. *et al.* Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J. Clin. Oncol.* **29,** 200–207 (jan. 2011).

57. Visco, C. *et al.* Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma : a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia* **26,** 2103–2113 (sept. 2012).

58. De Jong, D. *et al.* Immunohistochemical prognostic markers in diffuse large B-cell lymphoma : validation of tissue microarray as a prerequisite for broad clinical applications–a study from the Lunenburg Lymphoma Biomarker Consortium. *J. Clin. Oncol.* **25,** 805–812 (mar. 2007).

59. Muris, J. J. F. *et al.* Immunohistochemical profiling based on Bcl-2, CD10 and MUM1 expression improves risk stratification in patients with primary nodal diffuse large B cell lymphoma. *J. Pathol.* **208,** 714–723 (avr. 2006).

60. Natkunam, Y. *et al.* LMO2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy with and without rituximab. *J. Clin. Oncol.* **26,** 447–454 (jan. 2008).

61. Nyman, H., Jerkeman, M., Karjalainen-Lindsberg, M.-L., Banham, A. H. & Leppä, S. Prognostic impact of activated B-cell focused classification in diffuse large B-cell lymphoma patients treated with R-CHOP. *Mod. Pathol.* **22,** 1094–1101 (août 2009).

62. Hermine, O. *et al.* Prognostic significance of bcl-2 protein expression in aggressive non-Hodgkin's lymphoma. Groupe d'Etude des Lymphomes de l'Adulte (GELA). *Blood* **87,** 265–272 (jan. 1996).

63. Nyman, H. *et al.* Prognostic impact of immunohistochemically defined germinal center phenotype in diffuse large B-cell lymphoma patients treated with immunochemotherapy. *Blood* **109,** 4930–4935 (juin 2007).

64. Zu, Y. *et al.* Validation of tissue microarray immunohistochemistry staining and interpretation in diffuse large B-cell lymphoma. *Leuk. Lymphoma* **46,** 693–701 (mai 2005).

65. Culpin, R. E. *et al.* Prognostic significance of immunohistochemistry-based markers and algorithms in immunochemotherapy-treated diffuse large B cell lymphoma patients. *Histopathology* **63,** 788–801 (déc. 2013).

66. Lu, T.-X. *et al.* Immunohistochemical algorithm alone is not enough for predicting the outcome of patients with diffuse large B-cell lymphoma treated with R-CHOP. *Int J Clin Exp Pathol* **8,** 275–286 (2015).

67. Linderoth, J. *et al.* Tissue microarray is inappropriate for analysis of BCL6 expression in diffuse large B-cell lymphoma. *Eur. J. Haematol.* **79,** 146–149 (août 2007).

68. Rizzardi, A. E. *et al.* Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* **7,** 42 (2012).

69. Von Ahlfen, S., Missel, A., Bendrat, K. & Schlumpberger, M. Determinants of RNA quality from FFPE samples. *PLoS ONE* **2,** e1261 (2007).

70. Williams, P. M. *et al.* A novel method of amplification of FFPET-derived RNA enables accurate disease classification with microarrays. *J Mol Diagn* **12,** 680–686 (sept. 2010).

71. Kurn, N. *et al.* Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. *Clin. Chem.* **51,** 1973–1981 (oct. 2005).

72. Barrans, S. L. *et al.* Whole genome expression profiling based on paraffin embedded tissue can be used to classify diffuse large B-cell lymphoma and predict clinical outcome. *Br. J. Haematol.* **159,** 441–453 (nov. 2012).

73. April, C. *et al.* Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. *PLoS ONE* **4,** e8162 (2009).

74. Jais, J.-P. *et al.* The expression of 16 genes related to the cell of origin and immune response predicts survival in elderly patients with diffuse large B-cell lymphoma treated with CHOP and rituximab. *Leukemia* **22,** 1917–1924 (oct. 2008).

75. Rimsza, L. M. *et al.* Accurate classification of diffuse large B-cell lymphoma into germinal center and activated B-cell subtypes using a nuclease protection assay on formalin-fixed, paraffin-embedded tissues. *Clin. Cancer Res.* **17,** 3727–3732 (juin 2011).

76. Martel, R. R. *et al.* Multiplexed screening assay for mRNA combining nuclease protection with luminescent array detection. *Assay Drug Dev Technol* **1,** 61–71 (nov. 2002).

77. Collie, A. M. B. *et al.* Molecular subtype classification of formalin-fixed, paraffin-embedded diffuse large B-cell lymphoma samples on the ICEPlex® system. *Br. J. Haematol.* **167,** 281–285 (oct. 2014).

78. Hall, J. S. *et al.* QuantiGene Plex Represents a Promising Diagnostic Tool for Cell-of-Origin Subtyping of Diffuse Large B-Cell Lymphoma. *J Mol Diagn* **17,** 402–411 (juil. 2015).

79. Harris, E. *et al.* Detection of Trypanosoma brucei spp. in human blood by a nonradioactive branched DNA-based technique. *J. Clin. Microbiol.* **34,** 2401–2407 (oct. 1996).

80. Masqué-Soler, N., Szczepanowski, M., Kohler, C. W., Spang, R. & Klapper, W. Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens. *Blood* **122,** 1985–1986 (sept. 2013).

81. Scott, D. W. *et al.* Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **123,** 1214–1217 (fév. 2014).

82. Veldman-Jones, M. H. *et al.* Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System. *Clin. Cancer Res.* **21,** 2367–2378 (mai 2015).

83. Technologies, N. *NanoString Technologies Enters into Biomarker Companion Diagnostic Collaboration with Celgene Corporation to Support Development of REVLIMID as Treatment for Patients with Diffuse Large B-Cell Lymphoma* mar. 2014. <http://www.nanostring.com/company/corp_press_release?id=116> (visité le 12/08/2015).

84. *Efficacy and Safety Study of Lenalidomide Plus R-CHOP Chemotherapy Versus Placebo Plus R-CHOP Chemotherapy in Untreated ABC Type Diffuse Large B-cell Lymphoma (ROBUST)* avr. 2014. <https://clinicaltrials.gov/ct2/show/NCT02285062> (visité le 12/08/2015).

85. Scott, D. W. Cell-of-Origin in Diffuse Large B-Cell Lymphoma : Are the Assays Ready for the Clinic ? *Am Soc Clin Oncol Educ Book* **35,** e458–466 (2015).

86. Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105,** 1851–1861 (mar. 2005).

87. Hummel, M. *et al.* A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N. Engl. J. Med.* **354,** 2419–2430 (juin 2006).

88. Klapper, W. *et al.* Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. *Blood* **112,** 1374–1381 (août 2008).

89. Salaverria, I. *et al.* Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. *Blood* **118,** 139–147 (juil. 2011).

90. Rosolowski, M. *et al.* Massive transcriptional perturbation in subgroups of diffuse large B-cell lymphomas. *PLoS ONE* **8,** e76287 (2013).

91. Dybkær, K. *et al.* Diffuse large B-cell lymphoma classification system that associates normal B-cell subset phenotypes with prognosis. *J. Clin. Oncol.* **33,** 1379–1388 (avr. 2015).

92. Lossos, I. S. *et al.* Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* **350,** 1828–1837 (avr. 2004).

93. Shipp, M. A. *et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* **8,** 68–74 (jan. 2002).

94. Lossos, I. S. *et al.* Expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. *Blood* **98,** 945–951 (août 2001).

95. Winter, J. N. *et al.* Prognostic significance of Bcl-6 protein expression in DLBCL treated with CHOP or R-CHOP : a prospective correlative study. *Blood* **107,** 4207–4213 (juin 2006).

96. Sjö, L. D., Poulsen, C. B., Hansen, M., Møller, M. B. & Ralfkiaer, E. Profiling of diffuse large B-cell lymphoma by immunohistochemistry : identification of prognostic subgroups. *Eur. J. Haematol.* **79,** 501–507 (déc. 2007).

97. Dunleavy, K., Davis, R. E., Landgren, O., Staudt, L. M. & Wilson, W. H. BCL-6 and rituximab in diffuse large B-cell lymphoma : where are we ? *Blood* **109,** 843–844 ; discussion 844–845 (jan. 2007).

98. Alizadeh, A. A. *et al.* Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. *Blood* **118,** 1350–1358 (avr. 2011).

99. Barrans, S. L., Fenton, J. A. L., Banham, A., Owen, R. G. & Jack, A. S. Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood* **104,** 2933–2935 (nov. 2004).

100. Banham, A. H. *et al.* Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma. *Clin. Cancer Res.* **11,** 1065–1072 (fév. 2005).

101. Lane, D. P. On the expression of the p53 protein in human cancer. *Mol. Biol. Rep.* **19,** 23–29 (jan. 1994).

102. Piris, M. A. *et al.* p53 expression in non-Hodgkin's lymphomas : a marker of p53 inactivation ? *Leuk. Lymphoma* **17,** 35–42 (mar. 1995).

103. Xie, Y. *et al.* p53 expression is a strong marker of inferior survival in de novo diffuse large B-cell lymphoma and may have enhanced negative effect with MYC coexpression : a single institutional clinicopathologic study. *Am. J. Clin. Pathol.* **141,** 593–604 (avr. 2014).

104. Scholzen, T. & Gerdes, J. The Ki-67 protein : from the known and the unknown. *J. Cell. Physiol.* **182,** 311–322 (mar. 2000).

105. He, X. *et al.* Ki-67 is a valuable prognostic predictor of lymphoma but its utility varies in lymphoma subtypes : evidence from a systematic meta-analysis. *BMC Cancer* **14,** 153 (2014).

106. Cojoc, M. *et al.* Emerging targets in cancer management : role of the CXCL12/CXCR4 axis. *Onco Targets Ther* **6,** 1347–1361 (2013).

107. Chen, J. *et al.* Dysregulated CXCR4 expression promotes lymphoma cell survival and independently predicts disease progression in germinal center B-cell-like diffuse large B-cell lymphoma. *Oncotarget* **6,** 5597–5614 (mar. 2015).

108. Johnson, N. A. *et al.* Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J. Clin. Oncol.* **30,** 3452–3459 (oct. 2012).

109. Slack, G. W., Steidl, C., Sehn, L. H. & Gascoyne, R. D. CD30 expression in de novo diffuse large B-cell lymphoma : a population-based study from British Columbia. *Br. J. Haematol.* **167,** 608–617 (déc. 2014).

110. *Brentuximab Vedotin in Treating Patients With Relapsed or Refractory CD30+ Lymphoma* août 2012. <https://clinicaltrials.gov/ct2/show/NCT01703949> (visité le 14/08/2015).

111. *Brentuximab Vedotin + Rituximab as Frontline Therapy for Pts w/ CD30+ and/or EBV+ Lymphomas* fév. 2013. <https://clinicaltrials.gov/ct2/show/NCT01805037> (visité le 14/08/2015).

112. Alencar, A. J. *et al.* MicroRNAs are independent predictors of outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Clin. Cancer Res.* **17,** 4125–4135 (juin 2011).

113. Wang, W.-l. *et al.* MicroRNA-23a expression in paraffin-embedded specimen correlates with overall survival of diffuse large B-cell lymphoma. *Med. Oncol.* **31,** 919 (avr. 2014).

114. Hedström, G. *et al.* Low expression of microRNA-129-5p predicts poor clinical outcome in diffuse large B cell lymphoma (DLBCL). *Int. J. Hematol.* **97,** 465–471 (avr. 2013).

115. Iqbal, J. *et al.* Global microRNA expression profiling uncovers molecular markers for classification and prognosis in aggressive B-cell lymphoma. *Blood* **125,** 1137–1145 (fév. 2015).

116. Berglund, M., Hedström, G., Amini, R.-M., Enblad, G. & Thunberg, U. High expression of microRNA-200c predicts poor clinical outcome in diffuse large B-cell lymphoma. *Oncol. Rep.* **29,** 720–724 (fév. 2013).

117. Song, G. *et al.* Deregulated expression of miR-224 and its target gene : CD59 predicts outcome of diffuse large B-cell lymphoma patients treated with R-CHOP. *Curr Cancer Drug Targets* **14,** 659–670 (2014).

118. Bai, H. *et al.* MicroRNA-21 regulates the sensitivity of diffuse large B-cell lymphoma cells to the CHOP chemotherapy regimen. *Int. J. Hematol.* **97,** 223–231 (fév. 2013).

119. Jørgensen, L. K. *et al.* MicroRNAs as novel biomarkers in diffuse large B-cell lymphoma–a systematic review. *Dan Med J* **62** (mai 2015).

120. Song, G. *et al.* Serum microRNA expression profiling predict response to R-CHOP treatment in diffuse large B cell lymphoma patients. *Ann. Hematol.* **93,** 1735–1743 (oct. 2014).

121. Salles, G. *et al.* Prognostic significance of immunohistochemical biomarkers in diffuse large B-cell lymphoma : a study from the Lunenburg Lymphoma Biomarker Consortium. *Blood* **117,** 7070–7078 (juin 2011).

122. Radmacher, M. D., McShane, L. M. & Simon, R. A paradigm for class prediction using gene expression profiles. *J. Comput. Biol.* **9,** 505–511 (2002).

123. Mareschal, S. *et al.* The proportion of activated B-cell like subtype among de novo diffuse large B-cell lymphoma increases with age. *Haematologica* **96,** 1888–1890 (déc. 2011).

124. Lanic, H. *et al.* Interim positron emission tomography scan associated with international prognostic index and germinal center B cell-like signature as prognostic index in diffuse large B-cell lymphoma. *Leuk. Lymphoma* **53,** 34–42 (jan. 2012).

125. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **58,** 236–244 (mar. 1963).

126. Thunberg, U., Enblad, G. & Berglund, M. Classification of diffuse large B-cell lymphoma by immunohistochemistry demonstrates that elderly patients are more common in the non-GC subgroup and younger patients in the GC subgroup. *Haematologica* **97,** e3 ; author reply e4 (fév. 2012).

127. Ruminy, P. *et al.* The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. *Leukemia* **25,** 681–688 (avr. 2011).

128. Guney, S. *et al.* Several mechanisms lead to the inactivation of the CDKN2A (P16), P14ARF, or CDKN2B (P15) genes in the GCB and ABC molecular DLBCL subtypes. *Genes Chromosomes Cancer* **51,** 858–867 (sept. 2012).

129. Bertrand, P. *et al.* The costimulatory molecule CD70 is regulated by distinct molecular mechanisms and is associated with overall survival in diffuse large B-cell lymphoma. *Genes Chromosomes Cancer* **52,** 764–774 (août 2013).

130. Bohers, E. *et al.* Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell lymphoma. *Genes Chromosomes Cancer* **53,** 144–153 (fév. 2014).

131. Dubois, S. *et al.* Immunohistochemical and genomic profiles of diffuse large B-cell lymphomas : Implications for targeted EZH2 inhibitor therapy ? *Oncotarget* **6,** 16712–16724 (juin 2015).

132. Staudt, L. M. *NCBI GEO Accession GSE11318* sept. 2008. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11318> (visité le 12/10/2015).

133. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (éds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer New York, 2005). isbn : 978-0-387-25146-2 978-0-387-29362-2.

134. Scott, S. D. Rituximab : a new therapeutic monoclonal antibody for non-Hodgkin's lymphoma. *Cancer Pract* **6,** 195–197 (juin 1998).

135. Swerdlow, S. H. *et al.* Expression of cyclin D1 protein in centrocytic/mantle cell lymphomas with and without rearrangement of the BCL1/cyclin D1 gene. *Hum. Pathol.* **26,** 999–1004 (sept. 1995).

136. Hans, C. P. *et al.* Expression of PKC-beta or cyclin D2 predicts for inferior survival in diffuse large B-cell lymphoma. *Mod. Pathol.* **18,** 1377–1384 (oct. 2005).

137. Mareschal, S. *et al.* Accurate Classification of Germinal Center B-Cell-Like/Activated B-Cell-Like Diffuse Large B-Cell Lymphoma Using a Simple and Rapid Reverse Transcriptase-Multiplex Ligation-Dependent Probe Amplification Assay : A CALYM Study. *J Mol Diagn* **17,** 273–283 (mai 2015).

138. Quindroit, A. Des chercheurs rouennais franchissent un grand pas contre le cancer. *Paris Normandie* **21752,** 4 (avr. 2015).

139. Ruminy, P. *et al. Accurate Classification Of GCB/ABC and MYC/BCL2 Diffuse Large B-Cell Lymphoma With a 14 Genes Expression Signature and a Simple and Robust RT-MLPA Assay* in *55th ASH Annual Meeting and Exposition* Session 622. Non-Hodgkin Lymphoma : Biology, excluding Therapy : Genetic Signature in Non-Hodgkin Lymphomas (#84) (New Orleans, USA, déc. 2013).

140. Mareschal, S. *et al. Développement d'un outil simple et fiable de classification des lymphomes B diffus à grandes cellules GCB et ABC* in *Congrès de la Société Française d'Hématologie 2014* SCO8 - Syndromes lymphoprolifératifs - Biologie #1 (04-07) (Paris, France, mar. 2014).

141. Mareschal, S. *et al. Développement d'un outil simple et fiable de classification des lymphomes B diffus à grandes cellules GCB et ABC* in *7e journées scientifiques du Cancéropole Nord Ouest* Session Jeunes Chercheurs #4 (Deauville, France, mai 2014).

142. Rosenberg, R. N. *The Molecular and Genetic Basis of Neurologic and Psychiatric Disease* p 22. isbn : 978-0-7817-6956-3 (Lippincott Williams & Wilkins, 2008).

143. Pfeiffer, P., Goedecke, W. & Obe, G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* **15,** 289–302 (juil. 2000).

144. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4,** 712–720 (sept. 2003).

145. Harousseau, J.-L. Les leucémies induites par les chimiothérapies anticancéreuses. *Bulletin du Cancer* **86,** 929–38 (déc. 1999).

146. Clancy, S. Genetic Recombination. *Nature Education* **1,** 40 (2008).

147. Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **11,** 196–207 (mar. 2010).

148. Salmon, I. *et al.* Relationship between DNA ploidy level, nuclear size, and survival in large cell lymphoma. *Am. J. Clin. Pathol.* **103,** 568–573 (mai 1995).

149. Lackowska, B. *et al.* Diagnostic, predictive and prognostic verification of DNA flow cytometric measurements performed at diagnosis for non-Hodgkin's lymphoma adult patients. *Pol J Pathol* **63,** 18–24 (mar. 2012).

150. Cigudosa, J. C. *et al.* Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. *Genes Chromosomes Cancer* **25,** 123–133 (juin 1999).

151. Mitelman, F., Johansson, B. & Mertens, F. *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* 2015. <http://cgap.nci.nih.gov/Chromosomes/Mitelman> (visité le 02/09/2015).

152. Wessendorf, S. *et al.* Further delineation of chromosomal consensus regions in primary mediastinal B-cell lymphomas : an analysis of 37 tumor samples using high-resolution genomic profiling (array-CGH). *Leukemia* **21,** 2463–2469 (déc. 2007).

153. Scandurra, M. *et al.* Genomic lesions associated with a different clinical outcome in diffuse large B-Cell lymphoma treated with R-CHOP-21. *Br. J. Haematol.* **151,** 221–231 (nov. 2010).

154. Monti, S. *et al.* Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma. *Cancer Cell* **22,** 359–372 (sept. 2012).

155. Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366,** 704–707 (déc. 1993).

156. Sherr, C. J. Tumor surveillance via the ARF-p53 pathway. *Genes Dev.* **12,** 2984–2991 (oct. 1998).

157. Jardin, F. *et al.* Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment : a GELA study. *Blood* **116,** 1092–1104 (août 2010).

158. Herman, J. G., Jen, J., Merlo, A. & Baylin, S. B. Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. *Cancer Res.* **56,** 722–727 (fév. 1996).

159. Lee, B.-s. *et al.* The BCL11A transcription factor directly activates RAG gene expression and V(D)J recombination. *Mol. Cell. Biol.* **33,** 1768–1781 (mai 2013).

160. Küppers, R. *et al.* Lack of somatic hypermutation of IG V(H) genes in lymphoid malignancies with t(2 ;14)(p13 ;q32) translocation involving the BCL11A gene. *Leukemia* **16,** 937–939 (mai 2002).

161. Yin, C. C. *et al.* Chronic lymphocytic leukemia With t(2 ;14)(p16 ;q32) involves the BCL11A and IgH genes and is associated with atypical morphologic features and unmutated IgVH genes. *Am. J. Clin. Pathol.* **131,** 663–670 (mai 2009).

162. Liu, H. *et al.* Functional studies of BCL11A : characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Mol. Cancer* **5,** 18 (2006).

163. Fukuhara, N. *et al.* Characterization of target genes at the 2p15-16 amplicon in diffuse large B-cell lymphoma. *Cancer Sci.* **97,** 499–504 (juin 2006).

164. Taskinen, M. *et al.* Deregulation of COMMD1 is associated with poor prognosis in diffuse large B-cell lymphoma. *PLoS ONE* **9,** e91031 (2014).

165. Maine, G. N., Mao, X., Komarck, C. M. & Burstein, E. COMMD1 promotes the ubiquitination of NF-kappaB subunits through a cullin-containing ubiquitin ligase. *EMBO J.* **26,** 436–447 (jan. 2007).

166. Kwiecinska, A. *et al.* Amplification of 2p as a genomic marker for transformation in lymphoma. *Genes Chromosomes Cancer* **53,** 750–768 (sept. 2014).

167. Meier, C. *et al.* Recurrent numerical aberrations of JAK2 and deregulation of the JAK2-STAT cascade in lymphomas. *Mod. Pathol.* **22,** 476–487 (mar. 2009).

168. Hao, Y. *et al.* Selective JAK2 inhibition specifically decreases Hodgkin lymphoma and mediastinal large B-cell lymphoma growth in vitro and in vivo. *Clin. Cancer Res.* **20,** 2674–2683 (mai 2014).

169. *Pilot Study of Ruxolitinib in Relapsed or Refractory Hodgkin Lymphoma and Primary Mediastinal Large B-cell Lymphoma - Full Text View - ClinicalTrials.gov* oct. 2013. <https://clinicaltrials.gov/ct2/show/NCT01965119?term=jak2&rank=19> (visité le 04/09/2015).

170. *A Phase II Study of Oral JAK1/JAK2 Inhibitor INC424 in Adult Patients With Relapsed/Refractory Classical Hodgkin's Lymphoma* juin 2013. <https://clinicaltrials.gov/ct2/show/NCT01877005?term=jak2&rank=11> (visité le 04/09/2015).

171. Wu, D. *et al.* No evidence for the JAK2 (V617F) or JAK2 exon 12 mutations in primary mediastinal large B-cell lymphoma. *Diagn. Mol. Pathol.* **18,** 144–149 (sept. 2009).

172. Shi, M. *et al.* Expression of programmed cell death 1 ligand 2 (PD-L2) is a distinguishing feature of primary mediastinal (thymic) large B-cell lymphoma and associated with PDCD1LG2 copy gain. *Am. J. Surg. Pathol.* **38,** 1715–1723 (déc. 2014).

173. Taborelli, M. *et al.* Chromosome band 6q deletion pattern in malignant lymphomas. *Cancer Genet. Cytogenet.* **165,** 106–113 (mar. 2006).

174. Monni, O. *et al.* BCL2 overexpression associated with chromosomal amplification in diffuse large B-cell lymphoma. *Blood* **90,** 1168–1174 (août 1997).

175. Kichina, J. V. *et al.* Targeted disruption of the mouse ing1 locus results in reduced body size, hypersensitivity to radiation and elevated incidence of lymphomas. *Oncogene* **25,** 857–866 (fév. 2006).

176. Coles, A. H. *et al.* p37Ing1b regulates B-cell proliferation and cooperates with p53 to suppress diffuse large B-cell lymphomagenesis. *Cancer Res.* **68,** 8705–8714 (nov. 2008).

177. Hassan, H. M., Dave, B. J. & Singh, R. K. TP73, an underappreciated player in non-Hodgkin lymphoma pathogenesis and management. *Curr. Mol. Med.* **14,** 432–439 (mai 2014).

178. Tsai, A. G. & Lieber, M. R. Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics* **11 Suppl 1,** S1 (2010).

179. Goldman, J. M. & Melo, J. V. Chronic myeloid leukemia–advances in biology and new approaches to treatment. *N. Engl. J. Med.* **349,** 1451–1464 (oct. 2003).

180. Marculescu, R. *et al.* Recombinase, chromosomal translocations and lymphoid neoplasia : targeting mistakes and repair failures. *DNA Repair (Amst.)* **5,** 1246–1258 (sept. 2006).

181. Khodabakhshi, A. H. *et al.* Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget* **3,** 1308–1319 (nov. 2012).

182. Willis, T. G. & Dyer, M. J. The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood* **96,** 808–822 (août 2000).

183. Summers, K. E. *et al.* Frequency of the Bcl-2/IgH rearrangement in normal individuals : implications for the monitoring of disease in patients with follicular lymphoma. *J. Clin. Oncol.* **19,** 420–424 (jan. 2001).

184. Küppers, R. & Dalla-Favera, R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20,** 5580–5594 (sept. 2001).

185. Montoto, S. *et al.* Risk and clinical implications of transformation of follicular lymphoma to diffuse large B-cell lymphoma. *J. Clin. Oncol.* **25,** 2426–2433 (juin 2007).

186. Bastard, C. Cytogénétique des lymphomes malins et du myélome multiple. *Pathol. Biol.* **51,** 375–381 (août 2003).

187. Huang, J. Z. *et al.* The t(14 ;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. *Blood* **99,** 2285–2290 (avr. 2002).

188. Akasaka, T. *et al.* Refinement of the BCL2/immunoglobulin heavy chain fusion gene in t(14 ;18)(q32 ;q21) by polymerase chain reaction amplification for long targets. *Genes Chromosomes Cancer* **21,** 17–29 (jan. 1998).

189. Iqbal, J. *et al.* BCL2 predicts survival in germinal center B-cell-like diffuse large B-cell lymphoma treated with CHOP-like therapy and rituximab. *Clin. Cancer Res.* **17,** 7785–7795 (déc. 2011).

190. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7,** 233–245 (avr. 2007).

191. Neri, A., Barriga, F., Knowles, D. M., Magrath, I. T. & Dalla-Favera, R. Different regions of the immunoglobulin heavy-chain locus are involved in chromosomal translocations in distinct pathogenetic forms of Burkitt lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 2748–2752 (avr. 1988).

192. Liu, D. *et al.* t(8 ;14 ;18) : a 3-way chromosome translocation in two patients with Burkitt's lymphoma/leukemia. *Mol. Cancer* **6,** 35 (2007).

193. Sewastianik, T., Prochorec-Sobieszek, M., Chapuy, B. & Juszczyński, P. MYC deregulation in lymphoid tumors : molecular mechanisms, clinical consequences and therapeutic implications. *Biochim. Biophys. Acta* **1846,** 457–467 (déc. 2014).

194. Barrans, S. *et al.* Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J. Clin. Oncol.* **28,** 3360–3365 (juil. 2010).

195. Aukema, S. M. *et al.* Double-hit B-cell lymphomas. *Blood* **117,** 2319–2331 (fév. 2011).

196. Busslinger, M., Klix, N., Pfeffer, P., Graninger, P. G. & Kozmik, Z. Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 6129–6134 (juin 1996).

197. Morton, L. M. *et al.* Molecular characteristics of diffuse large B-cell lymphoma in human immunodeficiency virus-infected and -uninfected patients in the pre-highly active antiretroviral therapy and pre-rituximab era. *Leuk. Lymphoma* **55,** 551–557 (mar. 2014).

198. Huret, J. BCL6 (B-Cell Lymphoma 6). *Atlas Genet Cytogenet Oncol Haematol* **17,** 371–379 (2013).

199. Offit, K. *et al.* Rearrangement of the bcl-6 gene as a prognostic marker in diffuse large-cell lymphoma. *N. Engl. J. Med.* **331,** 74–80 (juil. 1994).

200. Niitsu, N. *et al.* Prognostic impact of chromosomal alteration of 3q27 on nodal B-cell lymphoma : correlation with histology, immunophenotype, karyotype, and clinical outcome in 329 consecutive patients. *Leuk. Res.* **31,** 1191–1197 (sept. 2007).

201. Ohno, H. Pathogenetic and clinical implications of non-immunoglobulin ; BCL6 translocations in B-cell non-Hodgkin's lymphoma. *J Clin Exp Hematop* **46,** 43–53 (nov. 2006).

202. Wang, W., Hu, S., Lu, X., Young, K. H. & Medeiros, L. J. Triple-hit B-cell Lymphoma With MYC, BCL2, and BCL6 Translocations/Rearrangements : Clinicopathologic Features of 11 Cases. *Am. J. Surg. Pathol.* **39,** 1132–1139 (août 2015).

203. Laurent, C. *et al.* Anaplastic lymphoma kinase-positive diffuse large B-cell lymphoma : a rare clinicopathologic entity with poor prognosis. *J. Clin. Oncol.* **27,** 4211–4216 (sept. 2009).

204. Cuneo, A, Rigolin, G. & Ciccone, M. ALK-positive diffuse large B-cell lymphoma. *Atlas Genet Cytogenet Oncol Haematol* **15,** 352–353 (2011).

205. Lee, S. E., Kang, S. Y., Takeuchi, K. & Ko, Y. H. Identification of RANBP2-ALK fusion in ALK positive diffuse large B-cell lymphoma. *Hematol Oncol* **32,** 221–224 (déc. 2014).

206. Mariño-Enríquez, A. & Dal Cin, P. ALK as a paradigm of oncogenic promiscuity : different mechanisms of activation and different fusion partners drive tumors of different lineages. *Cancer Genet* **206,** 357–373 (nov. 2013).

207. Steidl, C. *et al.* MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. *Nature* **471,** 377–381 (mar. 2011).

208. Scott, D. W. *et al.* TBL1XR1/TP63 : a novel recurrent gene fusion in B-cell non-Hodgkin lymphoma. *Blood* **119,** 4949–4952 (mai 2012).

209. Kallioniemi, A. *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258,** 818–821 (oct. 1992).

210. Theisen, A. Microarray-based Comparative Genomic Hybridization (aCGH). **1** (2008).

211. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31,** 265–273 (déc. 2003).

212. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (oct. 2004).

213. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (mar. 2007).

214. Willenbrock, H. & Fridlyand, J. A comparison study : applying segmentation to array CGH data for downstream analyses. *Bioinformatics* **21,** 4084–4091 (nov. 2005).

215. Aguirre, A. J. *et al.* High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 9067–9072 (juin 2004).

216. Hanemaaijer, N. M. *et al.* Practical guidelines for interpreting copy number gains detected by high-resolution array in routine diagnostics. *Eur. J. Hum. Genet.* **20,** 161–165 (fév. 2012).

217. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).

218. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12,** 115–121 (fév. 2015).

219. R Core Team. *R : A Language and Environment for Statistical Computing* <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2015).

220. *GNU Public License 3.0* juin 2007. <http://www.gnu.org/licenses/gpl-3.0.fr.html> (visité le 24/08/2015).

221. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9,** e1003118 (2013).

222. Genolini, C. *Petit traité de programmation orientée objet sous R : Programmation, Construction de packages, Bonnes pratiques* isbn : 978-2-8041-6251-1 (De Boeck, Bruxelles, sept. 2010).

223. *maressyl/R.Rgb* <https://github.com/maressyl/R.Rgb> (visité le 24/08/2015).

224. *Rgb 1.4.1* <http://bioinformatics.ovsa.fr/72/Rgb_141> (visité le 24/08/2015).

225. Mareschal, S., Ruminy, P., Jardin, F. & Tilly, H. *Rgb : a native genome browser for R* in *21st Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and 12th European Conference on Computational Biology (ECCB)* G - Genome Organization and Annotation (#13) (Berlin, Allemagne, août 2013).

226. Mareschal, S., Dubois, S., Lecroq, T. & Jardin, F. Rgb : a scriptable genome browser for R. *Bioinformatics* **30,** 2204–2205 (août 2014).

227. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (avr. 2002).

228. Leprêtre, F *et al.* Waved aCGH : to smooth or not to smooth. *Nucleic Acids Res.* **38,** e94 (avr. 2010).

229. Neuvial, P. *et al.* Spatial normalization of array-CGH data. *BMC Bioinformatics* **7,** 264 (2006).

230. Jardin, F. *et al.* Detection of somatic quantitative genetic alterations by multiplex polymerase chain reaction for the prediction of outcome in diffuse large B-cell lymphomas. *Haematologica* **93,** 543–550 (avr. 2008).

231. Picard, F., Robin, S., Lavielle, M., Vaisse, C. & Daudin, J.-J. A statistical approach for array CGH data analysis. *BMC Bioinformatics* **6,** 27 (2005).

232. Engler, D. A., Mohapatra, G., Louis, D. N. & Betensky, R. A. A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics* **7,** 399–421 (juil. 2006).

233. Van de Wiel, M. A. *et al.* CGHcall : calling aberrations for array CGH tumor profiles. *Bioinformatics* **23,** 892–894 (avr. 2007).

234. Hupé, P., Stransky, N., Thiery, J.-P., Radvanyi, F. & Barillot, E. Analysis of array CGH data : from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20,** 3413–3422 (déc. 2004).

235. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36,** 949–951 (sept. 2004).

236. Duclos, A. *et al.* Pitfalls in the use of DGV for CNV interpretation. *Am. J. Med. Genet. A* **155A,** 2593–2596 (oct. 2011).

237. Bastida-Lertxundi, N. *et al.* Errors in the interpretation of copy number variations due to the use of public databases as a reference. *Cancer Genet* **207,** 164–167 (avr. 2014).

238. Uddin, M. *et al.* A high-resolution copy-number variation resource for clinical and population genetics. *Genet. Med.* doi :10.1038/gim.2014.178 (déc. 2014).

239. Fitoussi, O. *et al.* Survival impact of rituximab combined with ACVBP and upfront consolidation autotransplantation in high-risk diffuse large B-cell lymphoma for GELA. *Haematologica* **96,** 1136–1143 (août 2011).

240. Peyrade, F. *et al.* Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma : a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* **12,** 460–468 (mai 2011).

241. Récher, C. *et al.* Intensified chemotherapy with ACVBP plus rituximab versus standard CHOP plus rituximab for the treatment of diffuse large B-cell lymphoma (LNH03-2B) : an open-label randomised phase 3 trial. *Lancet* **378,** 1858–1867 (nov. 2011).

242. Delarue, R. *et al.* Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study) : a randomised phase 3 trial. *Lancet Oncol.* **14,** 525–533 (mai 2013).

243. Ketterer, N. *et al.* Phase III study of ACVBP versus ACVBP plus rituximab for patients with localized low-risk diffuse large B-cell lymphoma (LNH03-1B). *Ann. Oncol.* **24,** 1032–1037 (avr. 2013).

244. Molina, T. J. *et al.* Young patients with non-germinal center B-cell-like diffuse large B-cell lymphoma benefit from intensified chemotherapy with ACVBP plus rituximab compared with CHOP plus rituximab : analysis of data from the Groupe d'Etudes des Lymphomes de l'Adulte/lymphoma study association phase III trial LNH 03-2B. *J. Clin. Oncol.* **32,** 3996–4003 (déc. 2014).

245. Wang, X. *et al.* High expression of oncoprotein DEK predicts poor prognosis of small cell lung cancer. *Int J Clin Exp Pathol* **7,** 5016–5023 (2014).

246. Yi, H.-C. *et al.* Overexpression of DEK gene is correlated with poor prognosis in hepatocellular carcinoma. *Mol Med Rep* **11,** 1318–1323 (fév. 2015).

247. Rossi, E. *et al.* Forced expression of RDH10 gene retards growth of HepG2 cells. *Cancer Biol. Ther.* **6,** 238–244 (fév. 2007).

248. Trinh, D. L. *et al.* Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood* **121,** 3666–3674 (mai 2013).

249. Leenders, F. *et al.* PKN3 is required for malignant prostate cell growth downstream of activated PI 3-kinase. *EMBO J.* **23,** 3303–3313 (août 2004).

250. Inoue, M., Saeki, M., Egusa, H., Niwa, H. & Kamisaki, Y. PIH1D1, a subunit of R2TP complex, inhibits doxorubicin-induced apoptosis. *Biochem. Biophys. Res. Commun.* **403,** 340–344 (déc. 2010).

251. Jardin, F. *et al. Integrated Analysis of High-Resolution Gene Expression and Copy Number Profiling Identified Biallelic Deletion of CDKN2A/2B Tumor Suppressor Locus As the Most Frequent and Unique Genomic Abnormality in Diffuse Large B-Cell Lymphoma (DLBCL) with Strong Prognostic Value in Both GCB and ABC Subtypes and Not Overcome by a Dose-Intensive Immunochemotherapy Regimen Plus Rituximab. Results of a Prospective GELA Clinical Trial Program* in *54th ASH Annual Meeting and Exposition* Session 622. Non-Hodgkin Lymphoma - Biology, excluding Therapy : Genetics of Diffuse Large B-cell Lymphoma (#415) (Atlanta, USA, déc. 2012).

252. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15,** 585–598 (sept. 2014).

253. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (août 2013).

254. Jiang, Y., Soong, T. D., Wang, L., Melnick, A. M. & Elemento, O. Genome-wide detection of genes targeted by non-Ig somatic hypermutation in lymphoma. *PLoS ONE* **7,** e40332 (2012).

255. Morin, R. D. *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122,** 1256–1265 (août 2013).

256. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3,** 246–259 (jan. 2013).

257. Imamura, J., Miyoshi, I. & Koeffler, H. P. p53 in hematologic malignancies. *Blood* **84,** 2412–2421 (oct. 1994).

258. Migliazza, A. *et al.* Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 12520–12524 (déc. 1995).

259. Gordon, M. S., Kanegai, C. M., Doerr, J. R. & Wall, R. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proc. Natl. Acad. Sci. U.S.A.* **100,** 4126–4131 (avr. 2003).

260. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476,** 298–303 (août 2011).

261. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43,** 830–837 (sept. 2011).

262. Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 3879–3884 (mar. 2012).

263. Zhang, J. *et al.* Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 1398–1403 (jan. 2013).

264. Turajlic, S., McGranahan, N. & Swanton, C. Inferring mutational timing and reconstructing tumour evolutionary histories. *Biochim. Biophys. Acta* **1855,** 264–275 (avr. 2015).

265. De Miranda, N. F. C. C. *et al.* Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. *Blood* **124,** 2544–2553 (oct. 2014).

266. Novak, A. J. *et al.* Whole-exome analysis reveals novel somatic genomic alterations associated with outcome in immunochemotherapy-treated diffuse large B-cell lymphoma. *Blood Cancer J* **5,** e346 (2015).

267. Fabbri, G. *et al.* Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J. Exp. Med.* **210,** 2273–2288 (oct. 2013).

268. Pasqualucci, L. *et al.* Genetics of follicular lymphoma transformation. *Cell Rep* **6,** 130–140 (jan. 2014).

269. Jiang, Y. *et al.* Deep sequencing reveals clonal evolution patterns and mutation events associated with relapse in B-cell lymphomas. *Genome Biol.* **15,** 432 (2014).

270. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (jan. 2014).

271. Dubois, S. *et al. Recurrent somatic mutations in Diffuse Large B Cell Lymphoma assessed by high-throughput targeted sequencing highlight molecular subtypes' genetic divergence : a LYSA study.* in *13th International Conference on Malignant Lymphoma* Session 1 - Lymphoma biology and genomics I (#049) (Lugano, Suisse, juin 2015).

272. Dubois, S. *et al.* Next Generation Sequencing in Diffuse Large B Cell Lymphoma Highlights Molecular Divergence and Therapeutic Opportunities : a LYSA Study. En révision (07/10/2015).

273. Dubois, S. & Jardin, F. The role of next-generation sequencing in understanding the genomic basis of non-Hodgkin lymphoma and advancing targeted therapies. *Expert Rev Hematol.* Revue commandée, sous presse.

274. Kim, Y. *et al.* CD79B and MYD88 mutations in diffuse large B-cell lymphoma. *Hum. Pathol.* **45,** 556–564 (mar. 2014).

275. Fernández-Rodríguez, C. *et al.* MYD88 (L265P) mutation is an independent prognostic factor for outcome in patients with diffuse large B-cell lymphoma. *Leukemia* **28,** 2104–2106 (oct. 2014).

276. Xu-Monette, Z. Y. *et al.* Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP : report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood* **120,** 3986–3996 (nov. 2012).

277. Schuetz, J. M. *et al.* BCL2 mutations in diffuse large B-cell lymphoma. *Leukemia* **26,** 1383–1390 (juin 2012).

278. Béguelin, W. *et al.* EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23,** 677–692 (mai 2013).

279. Arunesh, G. M., Shanthi, E., Krishna, M. H., Sooriya Kumar, J. & Viswanadhan, V. N. Small molecule inhibitors of PIM1 kinase : July 2009 to February 2013 patent update. *Expert Opin Ther Pat* **24,** 5–17 (jan. 2014).

280. Paik, J. H. *et al.* Clinicopathologic implication of A20/TNFAIP3 deletion in diffuse large B-cell lymphoma : an analysis according to immunohistochemical subgroups and rituximab treatment. *Leuk. Lymphoma* **54,** 1934–1941 (sept. 2013).

281. Cheung, K.-J. J. *et al.* Acquired TNFRSF14 mutations in follicular lymphoma are associated with worse prognosis. *Cancer Res.* **70,** 9166–9174 (nov. 2010).

282. Arcaini, L. *et al.* The NOTCH pathway is recurrently mutated in diffuse large B-cell lymphoma associated with hepatitis C virus infection. *Haematologica* **100,** 246–252 (fév. 2015).

283. Reichel, J. *et al.* Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood* **125,** 1061–1072 (fév. 2015).

284. Ritz, O. *et al.* Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. *Blood* **114,** 1236–1242 (août 2009).

285. Billard, C. BH3 mimetics : status of the field and new developments. *Mol. Cancer Ther.* **12,** 1691–1700 (sept. 2013).

286. Baron, B. W. *et al.* PIM1 gene cooperates with human BCL6 gene to promote the development of lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 5735–5739 (avr. 2012).

287. Sneeringer, C. J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 20980–20985 (déc. 2010).

288. Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat. Med.* doi :10.1038/nm.3940 (sept. 2015).

289. Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat. Med.* doi :10.1038/nm.3943 (sept. 2015).

290. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471,** 189–195 (mar. 2011).

291. Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463,** 88–92 (jan. 2010).

292. Wilson, W. H. *et al.* Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat. Med.* **21,** 922–926 (août 2015).

293. Ritz, O. *et al.* STAT6-mediated BCL6 repression in primary mediastinal B-cell lymphoma (PMBL). *Oncotarget* **4,** 1093–1102 (juil. 2013).

294. Gunawardana, J. *et al.* Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. *Nat. Genet.* **46,** 329–335 (avr. 2014).

295. Melzner, I. *et al.* Biallelic mutation of SOCS-1 impairs JAK2 degradation and sustains phospho-JAK2 action in the MedB-1 mediastinal lymphoma line. *Blood* **105,** 2535–2542 (mar. 2005).

296. *The Cancer Genome Atlas* <http://cancergenome.nih.gov/> (visité le 09/10/2015).

297. *International Cancer Genome Consortium* <https://icgc.org/> (visité le 09/10/2015).

298. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res.* **43,** D662–669 (jan. 2015).

299. *GRCh37.p13 - Assembly - NCBI* juin 2013. <http://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.25> (visité le 10/10/2015).

300. Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A. & Gilissen, C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum. Mutat.* **36,** 815–822 (août 2015).

301. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11,** 31–46 (jan. 2010).

302. Hodges, E. *et al.* Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* **39,** 1522–1527 (déc. 2007).

303. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27,** 182–189 (fév. 2009).

304. Asan, n. *et al.* Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol.* **12,** R95 (2011).

305. *History of Illumina Sequencing* <http://www.illumina.com/technology/next-generation-sequencing/solexa-technology.html> (visité le 11/10/2015).

306. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (nov. 2008).

307. Metzker, M. L. *Poster : Sequencing technologies - the next generation* nov. 2008. <http://www.nature.com/nrg/posters/sequencing/index.html> (visité le 09/10/2015).

308. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls : the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **11,** 11.10.1–11.10.33 (oct. 2013).

309. *POSIX 1003.1 - man page for getopts* 2013. <http://www.unix.com/man-page/posix/1p/getopts/> (visité le 11/10/2015).

310. Li, H. *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* arXiv e-print 1303.3997 (mar. 2013). <http://arxiv.org/abs/1303.3997> (visité le 14/11/2013).

311. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (mar. 1998).

312. Institute, B. *Picard Tools* <http://broadinstitute.github.io/picard/> (visité le 11/10/2015).

313. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (mai 2011).

314. Van der Auwera, G. A. *Base Quality Score Recalibration (BQSR)* "For humans, with dbSNP and now 1000 Genomes available, almost all of the mismatches - even in cancer - will be errors, and an accurate error model (essential for downstream analysis) can be ascertained.". Juil. 2015. <http://gatkforums.broadinstitute.org/discussion/44/base-quality-score-recalibration-bqsr> (visité le 11/10/2015).

315. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (août 2009).

316. Koboldt, D. C. *et al.* VarScan 2 : somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22,** 568–576 (mar. 2012).

317. Stead, L. F., Sutton, K. M., Taylor, G. R., Quirke, P. & Rabbitts, P. Accurately identifying low-allelic fraction variants in single samples with next-generation sequencing : applications in tumor subclone resolution. *Hum. Mutat.* **34,** 1432–1438 (oct. 2013).

318. Wang, Q. *et al.* Detecting somatic point mutations in cancer genome sequencing data : a comparison of mutation callers. *Genome Med* **5,** 91 (2013).

319. Roberts, N. D. *et al.* A comparative analysis of algorithms for somatic SNV detection in cancer. *Bioinformatics* **29,** 2223–2230 (sept. 2013).

320. Xu, H., DiCarlo, J., Satya, R. V., Peng, Q. & Wang, Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genomics* **15,** 244 (2014).

321. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (avr. 2010).

322. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4,** 1073–1081 (2009).

323. Dees, N. D. *et al.* MuSiC : identifying mutational significance in cancer genomes. *Genome Res.* **22,** 1589–1598 (août 2012).

324. Jardin, F. *et al. Whole Exome Sequencing of refractory aggressive B-cell lymphomas identified recurrent mutations of the exportin 1 gene (XPO1) in Primary Mediastinal B-cell Lymphoma subtype. A LYSA study.* in *13th International Conference on Malignant Lymphoma* Session 1 - Lymphoma biology and genomics I (#048) (Lugano, Suisse, juin 2015).

325. Jardin, F. *et al.* Recurrent mutations of the exportin 1 gene (XPO1) in primary mediastinal B-cell lymphoma : a LYSA study. En révision (12/10/2015).

326. Wendl, M. C. *et al.* PathScan : a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27,** 1595–1602 (juin 2011).

327. Amarasinghe, K. C. *et al.* Inferring copy number and genotype in tumour exome data. *BMC Genomics* **15,** 732 (2014).

328. Mareschal, S. *et al.* Whole exome sequencing of relapsed / refractory patients expands the repertoire of somatic mutations in DLBCL. *Genes Chromosomes Cancer.* En édition.

329. Mareschal, S. *et al. Exome sequencing of refractory Diffuse Large B-Cell Lymphomas highlights candidate genes for targeted resequencing* in *American Association for Cancer Research (AACR) annual meeting 2015* Molecular and Cellular Biology, Genomics 3 (#4810) (Philadelphia, PA, USA, avr. 2015).

330. Mareschal, S. *et al. Exome sequencing underscores the mutational heterogeneity of refractory / relapsed diffuse large B-cell lymphomas and identifies emerging patterns of somatic mutations* in *4emes journées scientifiques de l'Institut pour la Recherche et l'Innovation en Biomédecine (IRIB)* Communication affichée #59 (Rouen, France, juin 2015).

331. Mareschal, S. *et al. Exome sequencing underscores the mutational heterogeneity of refractory / relapsed diffuse large B-cell lymphomas and identifies emerging patterns of somatic mutations* in *8emes journées scientifiques du Cancéropole Nord-Ouest* Communication affichée #37 (Deauville, France, juin 2015).

332. Begum, N. A., Stanlie, A., Nakata, M., Akiyama, H. & Honjo, T. The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. *J. Biol. Chem.* **287,** 32415–32429 (sept. 2012).

333. Daniel, J. A. & Nussenzweig, A. The AID-induced DNA damage response in chromatin. *Mol. Cell* **50,** 309–321 (mai 2013).

334. Bahler, D. W. & Levy, R. Clonal evolution of a follicular lymphoma : evidence for antigen selection. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 6770–6774 (août 1992).

335. Limpens, J. *et al.* Lymphoma-associated translocation t(14 ;18) in blood B cells of normal individuals. *Blood* **85,** 2528–2536 (mai 1995).

336. Roulland, S. *et al.* t(14 ;18) Translocation : A predictive blood biomarker for follicular lymphoma. *J. Clin. Oncol.* **32,** 1347–1355 (mai 2014).

337. Tolar, P., Hanna, J., Krueger, P. D. & Pierce, S. K. The constant region of the membrane immunoglobulin mediates B cell-receptor clustering and signaling in response to membrane antigens. *Immunity* **30,** 44–55 (jan. 2009).

338. Leithäuser, F., Bäuerle, M., Huynh, M. Q. & Möller, P. Isotype-switched immunoglobulin genes with a high load of somatic hypermutation and lack of ongoing mutational activity are prevalent in mediastinal B-cell lymphoma. *Blood* **98,** 2762–2770 (nov. 2001).

339. Lenz, G. *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319,** 1676–1679 (mar. 2008).

340. Yang, Y. *et al.* Exploiting synthetic lethality for the therapy of ABC diffuse large B cell lymphoma. *Cancer Cell* **21,** 723–737 (juin 2012).

341. Isaza-Correa, J. M., Liang, Z., van den Berg, A., Diepstra, A. & Visser, L. Toll-like receptors in the pathogenesis of human B cell malignancies. *J Hematol Oncol* **7,** 57 (2014).

342. McCully, R. R. & Pomerantz, J. L. The protein kinase C-responsive inhibitory domain of CARD11 functions in NF-kappaB activation to regulate the association of multiple signaling cofactors that differentially depend on Bcl10 and MALT1 for association. *Mol. Cell. Biol.* **28,** 5668–5686 (sept. 2008).

343. Ngo, V. N. *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470,** 115–119 (fév. 2011).

344. Heyninck, K. & Beyaert, R. The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF-kappaB activation at the level of TRAF6. *FEBS Lett.* **442,** 147–150 (jan. 1999).

345. Steidl, C. *et al.* Genome-wide copy number analysis of Hodgkin Reed-Sternberg cells identifies recurrent imbalances with correlations to treatment outcome. *Blood* **116,** 418–427 (juil. 2010).

346. Pastore, A. *et al.* Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma : a retrospective analysis of a prospective clinical trial and validation in a population-based registry. *Lancet Oncol.* **16,** 1111–1122 (sept. 2015).

347. Tam, W. F., Lee, L. H., Davis, L. & Sen, R. Cytoplasmic sequestration of rel proteins by IkappaBalpha requires CRM1-dependent nuclear export. *Mol. Cell. Biol.* **20,** 2269–2284 (mar. 2000).

348. Thomas, M. *et al.* Analysis of transcriptional regulation of the human miR-17-92 cluster ; evidence for involvement of Pim-1. *Int J Mol Sci* **14,** 12273–12296 (2013).

349. Zippo, A., De Robertis, A., Serafini, R. & Oliviero, S. PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. *Nat. Cell Biol.* **9,** 932–944 (août 2007).

350. Kumar, A. *et al.* Crystal structures of proto-oncogene kinase Pim1 : a target of aberrant somatic hypermutations in diffuse large cell lymphoma. *J. Mol. Biol.* **348,** 183–193 (avr. 2005).

351. Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. p53 in survival, death and metabolic health : a lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* **16,** 393–405 (juil. 2015).

352. Basso, K. & Dalla-Favera, R. BCL6 : master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv. Immunol.* **105,** 193–210 (2010).

353. Pasqualucci, L. *et al.* Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* **101,** 2914–2923 (avr. 2003).

354. De Silva, N. S., Simonetti, G., Heise, N. & Klein, U. The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol. Rev.* **247,** 73–92 (mai 2012).

355. Lenz, G. *et al.* Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *J. Exp. Med.* **204,** 633–643 (mar. 2007).

356. Martins, G. & Calame, K. Regulation and functions of Blimp-1 in T and B lymphocytes. *Annu. Rev. Immunol.* **26,** 133–169 (2008).

357. Mandelbaum, J. *et al.* BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell* **18,** 568–579 (déc. 2010).

358. Challa-Malladi, M. *et al.* Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer Cell* **20,** 728–740 (déc. 2011).

359. Rimsza, L. M. *et al.* Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors : a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood* **103,** 4251–4258 (juin 2004).

360. Technologies, N. *nCounter Workflow* <http://www.nanostring.com/products/workflow> (visité le 23/10/2015).

361. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26,** 317–325 (mar. 2008).

362. Dennis, L., Demirkan, G. & Filanoski, B. *White paper - Exploring Immuno-oncology using nCounter RNA :Protein Profiling Technology* LBL-10128-02. Sept. 2015. <http://www.nanostring.com/products/rna-protein> (visité le 23/10/2015).

363. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (août 2000).

364. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (mar. 2009).

365. Wallden, B. *et al.* Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics* **8,** 54 (2015).

366. Technologies, N. *NanoString Launches Its First Commercial Diagnostic Product in the European Union and Israel - Press Release* fév. 2013. <http://www.nanostring.com/company/corp_press_release?id=74> (visité le 23/10/2015).

367. Technologies, N. *NanoString Technologies Receives FDA 510(k) Clearance for Prosigna™ Breast Cancer Prognostic Gene Signature Assay - Press Release* sept. 2013. <http://www.nanostring.com/company/corp_press_release?id=93> (visité le 23/10/2015).

368. Senkus, E. *et al.* Primary breast cancer : ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **26 Suppl 5,** v8–v30 (sept. 2015).

369. Scott, D. W. *et al.* Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. *J. Clin. Oncol.* **33,** 2848–2856 (sept. 2015).

370. Scott, D. W., Wright, G. W. & Williams, M. *Determining Cell-Of-Origin Subtypes In Diffuse Large B-Cell Lymphoma Using Gene Expression Profiling On Formalin-Fixed Paraffin-Embedded Tissue – An L.L.M.P.P. Project* in *55th ASH Annual Meeting and Exposition* Session 622. Non-Hodgkin Lymphoma : Biology, excluding Therapy : Genetic Signature in Non-Hodgkin Lymphomas (#73) (New Orleans, USA, déc. 2013).

371. Technologies, N. *nCounter Selection Guide* <http://www.nanostring.com/products/nCounterSelectionGuide> (visité le 24/10/2015).

372. Ruminy, P. *et al.* Multiplexed targeted sequencing of recurrent fusion genes in acute leukaemia. *Leukemia.* doi :10.1038/leu.2015.177 (juil. 2015).

373. Keane, C., Vari, F. & Hertzberg, M. Ratios of T-cell immune effectors and checkpoint molecules as prognostic biomarkers in diffuse large B-cell lymphoma : a population-based study. *Lancet Hematol.* **2,** e445–e455 (oct. 2015).

374. *BMS_PD-L1_onco : Assessment of the PD-L1 Protein as a Biomarker in Oncology and Hematology* juil. 2012. <https://clinicaltrials.gov/ct2/show/NCT01660776?term=PD-1+lymphoma&rank=4> (visité le 25/10/2015).

375. *Pembrolizumab After ASCT for Hodgkin Lymphoma and DLBCL* fév. 2015. <https://clinicaltrials.gov/ct2/show/NCT02362997?term=PD-1+lymphoma&rank=2> (visité le 25/10/2015).

376. *Study of Pembrolizumab (MK-3475) in Participants With Advanced Primary Mediastinal Large B-cell Lymphoma (MK-3475-170/KEYNOTE-170) - Full Text View - ClinicalTrials.gov* oct. 2015. <https://clinicaltrials.gov/ct2/show/NCT02576990?term=PD-1+lymphoma&rank=7> (visité le 25/10/2015).

377. Popova, T. *et al.* Genome Alteration Print (GAP) : a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol.* **10,** R128 (2009).

378. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 16910–16915 (sept. 2010).

379. Affymetrix, I. *Affymetrix Price Sheet - European List Price* Site de l'Université de Wurtzbourg. Fév. 2014. <http://www.sysmed.uni-wuerzburg.de/fileadmin/ext00315/_temp_/FEB2014_European_List_Price__EURLIST__EURLIST-EUR.pdf> (visité le 25/10/2015).

380. Boeva, V. *et al.* Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics* **30,** 3443–3450 (déc. 2014).

381. Coiffier, B. *et al.* CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346,** 235–242 (jan. 2002).

382. Wang, C. *et al.* PatternCNV : a versatile tool for detecting copy number changes from exome sequencing data. *Bioinformatics* **30,** 2678–2680 (sept. 2014).

383. Fontanilles, M., Marguet, F. & Bohers, E. *Somatic Mutations Detected in Plasma Cell-Free DNA By Targeted Sequencing : Assessment of Liquid Biopsy in Primary Central Nervous System Lymphoma* in *57th ASH Annual Meeting and Exposition* Session 622. Non-Hodgkin Lymphoma - Biology, excluding Therapy : Bench to Bedside (#332) (Orlando, USA, déc. 2015).

384. Bohers, E. *et al.* Somatic mutations of cell-free circulating DNA detected by next-generation sequencing reflect the genetic changes in both germinal center B-cell-like and activated B-cell-like diffuse large B-cell lymphomas at the time of diagnosis. *Haematologica* **100,** e280–284 (juil. 2015).

385. Abdelhamid, E. *et al.* Quantification of JAK2V617F mutation by next-generation sequencing technology. *Am. J. Hematol.* **88,** 536–537 (juin 2013).

386. Bohers, E. *et al.* Activating somatic mutations in diffuse large B-cell lymphomas : lessons from next generation sequencing and key elements in the precision medicine era. *Leuk. Lymphoma* **56,** 1213–1222 (mai 2015).