UNIVERSITAT DE
BARCELONA

Development and validation of a sequencing panel for translocations, copy number alterations and mutations with clinical relevance in mature B-cell neoplasms

Andrea Gómez Llonín

Development and validation of a sequencing panel for translocations, copy number alterations and mutations with clinical relevance in mature B-cell neoplasms
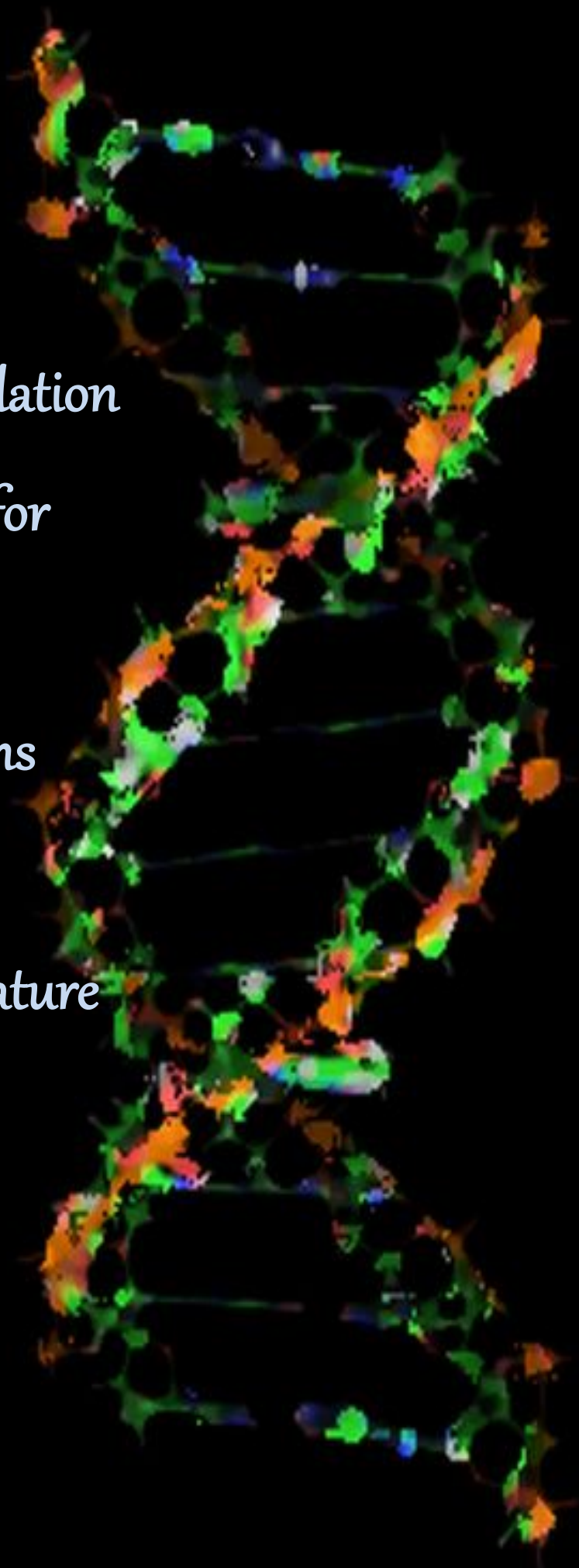
Andrea Gómez Llonín

Desarrollo y validación de un panel de secuenciación para translocaciones, alteraciones en número de copias y mutaciones con relevancia clínica en neoplasias de células B maduras

Memoria presentada por Andrea Gómez Llonín para optar al grado de Doctora por la Universitat de Barcelona

Este trabajo se ha desarrollado bajo la dirección de la Dra. Blanca Espinet Solà y la Dra. Anna Maria Puiggros Metje, en el Grup de Recerca Translacional en Neoplàsies Hematològiques de l´Institut Hospital del Mar d´Investigacions Mèdiques (IMIM), y la tutorización del Dr. Bru Cormand Rifà del Departamento de Genética, Microbiología y Estadística de la Universitat de Barcelona

Andrea Gómez Llonín
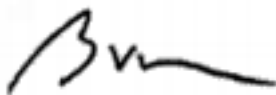
Directora de Tesis
Dra. Blanca Espinet Solà

Co-directora de Tesis
Dra. Anna Maria Puiggros Metje

Tutor de Tesis
Dr. Bru Cormand Rifà

Programa de Doctorado en Genética
Departamento de Genética, Microbiología y Estadística
Universitat de Barcelona
Barcelona, Diciembre de 2020

Agradecimientos

Acknowledgments

"Pintaba un día, el negro había invadido la tela por completo, sin formas, sin contrastes, sin transparencias.

En ese extremo vi de alguna manera la negación del negro. Las diferencias de textura reflejaban la luz con más o menos claridad, una luz pictórica, cuyo poder emocional particular animaba mi deseo de pintar.

Mi instrumento ya no era el negro, sino esa luz secreta procedente del negro"

Pierre Soulages

He estado a punto de no escribir este apartado. Sin duda alguna, me habría arrepentido enormemente, ya que para mí supone la parte más importante de la tesis, y por ello, lleva hasta su propia portada. Si hubiera podido ponerle una banda sonora, también la llevaría…

A pesar de no contener ningún valor científico, y por ello, ser prescindible, este apartado tiene un valor mayor, pues hace un reconocimiento a todas aquellas personas que, de una manera u otra, han sido imprescindibles en este camino, y suponen el esqueleto de este trabajo, sin el cual nada de esto hubiera sido posible. Pequeñas piezas de una gran estructura, en la que crecí y la que me ha permitido estar de pie a lo largo de todos estos años. Por ello, MIS AGRADECIMIENTOS deben ser un tributo a todas aquellas personas que contribuyen, y contribuyeron, a construir mis cimientos.

Especialmente durante los últimos años, he sido consciente de la huella que dejaron en mí algunas personas que forman parte de mi pasado. Creo que estas personas tienen una magia especial, pues, a pesar del tiempo, sigo sintiendo dentro de mí toda la confianza y el cariño que un día me dieron. Es mi deber hacérselo llegar. Por ello, quiero hacerles un especial reconocimiento a mis profesores, del insti, y alguno de la Universidad. Creo sinceramente que, en parte, es gracias a ellos que yo haya recorrido mi camino hasta aquí. La confianza (Loncho), la motivación y el interés por la ciencia (Juanjo), y creer en el caos de uno mismo (Jose Luis Vizmanos) son algunos de los regalos que me hicieron, y que, en los días feos, han salido a flote como un salvavidas.

Por otro lado, quiero dar las gracias a las personas que han seguido a mi lado. En especial a mi familia y amigos.

A mis padres, dos personas maravillosas y luchadoras. No he conocido a nadie como ellos. Quiero decirles que, a pesar de que creen que su mayor herencia son mis estudios, se equivocan. Sin duda lo son la constancia y el sacrificio que me han inculcado.

A mi madre, por ser la persona a la que acudo para lloriquear, y siempre me lo permite.

A mi padre, que nunca me permite lloriquear, y que, por el contrario, me regala frases casi filosóficas que me hacen mucha gracia y hacen que le quiera aún más.

A mi hermana Anne. Creo que si a alguien he de dar gracias es a ella. Esta tesis es casi tan suya como mía. Quien más me ha sufrido y me ha apoyado. Siempre estaré en deuda.

Gracias, por ser incondicionales. Por creer y confiar, siempre.

Me siento afortunada de teneros.

Por otro lado, quiero dar las gracias a todo el equipo.

A Blanca por dejarme ser miembro de este grupo. Por su confianza.

A Anna. Gracias por toda tu ayuda. No habría sido possible sin tí.

A las técnicos (Carme y Maria), por tener siempre un hueco y ayudarme con una sonrisa.

A Marta y Sergi, por esos ratos tan buenos, aunque escasos.

A Rosa, por los piropos, que siempre le alegran a uno el corazoncito.

A Silvia, por esos ratos en los que pensamos en fondue de chocolate para aliviar las penas de la tesis.

A Gonzalo. Por ser un mentor y un amigo.

También quiero dar las gracias al laboratorio de Suiza, por haberme acogido con los brazos abiertos, y haberme hecho sentir como en casa.

Al Nadie, que me han acompañado a lo largo de todos estos años, hasta el punto de sentir que son *como de la familia*. Gracias a sus desvaríos estoy un poquito más cuerda.

Por último, me gustaría concluir con una cita de E. E. Cummings. Es una frase que tengo muy presente, ya que parece haber siempre una fuerza que trata de homogeneizarlo todo. Creo que en ciencia, a veces, se nos olvida nadar con nuestro propio estilo.

"Ser únicamente quien eres, en un mundo que hace todo lo possible, continuamente, por

convertirte en todos los demás, implica luchar la batalla más difícil que pueda library

cualquier ser humano; y no dejar nunca de hacerlo..."

Gracias.

Abstract

The application of NGS techniques to the study of mature B-cell neoplasms at the research level is providing a large amount of new information, in certain cases with clinical implications, which includes the identification of molecular biomarkers with diagnostic, prognostic and predictive value. The incorporation of some of these molecular biomarkers into clinical practice has shown to be crucial for the correct diagnosis and management of these patients. Therefore, the integration of molecular information together with the genetic alterations detectable by NGS is of main relevance in order to transfer the new genomic knowledge generated to daily clinical practice. However, withing the framework of mature B-cell neoplasms, the application of these techniques has been mainly focused on the description of somatic mutation profiles, which hinders their applicability to the daily clinical practice of these entities. In this sense, we have developed a targeted NGS panel that allows the study of genetic alterations with diagnostic, prognostic and therapeutic value in mature B-cell neoplasms, currently determined with different techniques, using a single methodology. In addition, the panel also includes other relevant genetic alterations, which may have an impact on the diagnostic routine in the near future and that, otherwise, would not be studied in the current clinical practice. Furthermore, to analyze the results obtained with the NGS panel, we have also designed a bioinformatic algorithm, which performs an integrative analysis using different bioinformatics tools. This approach allowed an increase in the specificity and sensitivity of the panel. In this sense, we have successfully validated most of the genomic alterations detected by the designed panel using different conventional genetic techniques (CBA, FISH, genomic microarrays, PCR and Sanger sequencing) in addition to other massive sequencing panels for SNV validation. Furthermore, this has allowed us to compare the sensitivity and specificity of the panel with respect to genetic techniques performed in routine laboratories in addition to other NGS panels. To our knowledge, this is the first panel developed that allows integrated analysis of SNV and small indels, CNA and chromosomal rearrangements in MBCN, and although some additional studies are required to improve its performance, the designed panel could potentially be implemented in clinical practice.

Resumen

La aplicación de las técnicas de NGS al estudio de las neoplasias de células B maduras (NCBM) a nivel de investigación está aportando una gran cantidad de información nueva, en algunos casos con implicaciones clínicas, que incluye la identificación de biomarcadores moleculares con valor diagnóstico, pronóstico y predictivo. La incorporación de algunos de estos biomarcadores moleculares en la práctica clínica ha demostrado ser crucial para el correcto diagnóstico y manejo de estos pacientes. Por tanto, la integración de la información molecular junto con las alteraciones genéticas detectables por NGS es de principal relevancia para trasladar el nuevo conocimiento genómico generado a la práctica clínica diaria. Sin embargo, en el marco de las neoplasias de células B maduras, la aplicación de estas técnicas se ha centrado principalmente en la descripción de los perfiles de mutaciones somáticas, lo que dificulta su aplicabilidad en la rutina diagnóstica de estas entidades. En este sentido, hemos desarrollado un panel de NGS dirigido que permite el estudio de alteraciones genéticas con valor diagnóstico, pronóstico y terapéutico en neoplasias de células B maduras, actualmente determinadas con diferentes técnicas, utilizando una única metodología. Además, el panel también incluye otras alteraciones genéticas relevantes, que pueden tener un impacto en la rutina diagnóstica en un futuro próximo y que, de lo contrario, no serían estudiadas en la práctica clínica. Además, para analizar los resultados obtenidos con el panel desarrollado, también hemos diseñado un algoritmo bioinformático, que realiza un análisis integrado utilizando diferentes herramientas bioinformáticas. Este enfoque permite aumentar la especificidad y sensibilidad del panel. En este sentido, hemos validado con éxito la mayoría de las alteraciones genómicas detectadas por el panel diseñado utilizando diferentes técnicas genéticas convencionales (CBA, FISH, microarrays genómicos, PCR y secuenciación de Sanger) además de otros paneles de secuenciación masiva para la validación de SNV, lo que nos ha permitido comparar la sensibilidad y especificidad del panel con respecto a la de estas técnicas. Hasta donde sabemos, este es el primer panel desarrollado que permite el análisis integrado de SNV y pequeños indels, CNA y reordenamientos cromosómicos en NCBM, y aunque se requieren algunos estudios adicionales para mejorar su rendimiento, el panel diseñado podría potencialmente ser implementado en la práctica clínica.

Table of contents

List of abbreviations

A

ABC: Activated B-cell

ABR: Alternative Breakpoint Cluster

ACMGG: American College of Medical Genetics and Genomics

aCN-LOH: acquired Copy Neutral Loss Of Heterozygosity

AMP: Association for Molecular Pathology

AnxA1: *Annexin A1*

ARID1A: AT-Rich Interaction Domain 1A

ARID1B: AT-Rich Interaction Domain 1B

ASCO: American Society of Clinical Oncology

ATM: Ataxia Telangiectasia Mutated

AT-rich regions: Adenina-Timina rich regions

B

B2M: Beta-2-Microglobulin

BAC: *Bacterial* Artificial Chromosome

BCL10: B-cell lymphoma/leukemia 10

BCL11A: BAF Chromatin Remodeling Complex Subunit BCL11A

BCL2: B-cell lymphoma 6

BCL2L11: Bcl-2-like protein 11

BCL6: B-cell lymphoma 6

BCL7A: BAF Chromatin Remodeling Complex Subunit BCL7A

BCR: B-cell receptor

BED: Browser Extensible Data

BIRC3: Baculoviral IAP Repeat Containing 3

BL: Burkitt Lymphoma

BM: Bone Marrow

BNIP1: BCL2 Interacting Protein 1

bp: Base pair

BRAF: B-Raf Proto-Oncogene, Serine/Threonine Kinase

BTK: Bruton Tyrosine Kinase

BWA: Burrows-Wheeler Alignment tool

C

C2H2 domain: Cys2His2 zinc-finger domain

CAP: College of American Pathologists

CARD11: Caspase Recruitment Domain Family Member 11

CAV: Caveolin-1

CBA: Chromosome Banding Analysis

C-banding: Centromere banding

CBS: Circular Binary Segmentation

CCND1: Cyclin D1

CCND2: Cyclin D2

CCND3: Cyclin D3

CCR4: C-C Motif Chemokine Receptor 4

CD5: Cluster of differentiation 5

CD10: Cluster of differentiation 10

CD11c: Cluster of differentiation 11c

CD19: Cluster of differentiation 19

CD20: Cluster of differentiation 20

CD21: Cluster of differentiation 21

CD22: Cluster of differentiation 22

CD23: Cluster of differentiation 23

CD24: Cluster of differentiation 24

CD25: Cluster of differentiation 25

CD35: Cluster of differentiation 35

CD38: Cluster of differentiation 38

CD43: Cluster of differentiation 43

CD58: Cluster of differentiation 58

CD77: Cluster of differentiation 77

CD79a: Cluster of differentiation 79a

CD79b: Cluster of differentiation 79b

CD103: Cluster of differentiation 103

CD123: Cluster of differentiation 123

CD138: Cluster of differentiation 138

CD200: Cluster of differentiation 200

CDK: Cyclin-dependent kinase

CDK6: Cell division protein kinase 6

CDKN2A: Cyclin-dependent kinase inhibitor 2A

CDKN2B: Cyclin-dependent kinase inhibitor 2B

cDNA: complementary DNA

CE: Capillary Electrophoresis

CGH: Comparative Genomic Hybridization

cHCL: Classical hairy cell leukemia

CHD2: Chromodomain-helicase-DNA-binding protein 2

CK: Complex karyotype

CDK4: Cyclin-dependent kinase 4

CDK6: Cyclin-dependent kinase 6

CDKN1B: Cyclin Dependent Kinase Inhibitor 1B

CLL: Chronic lymphocytic leukemia

CNA: Copy Number Alterations

CPC: Common Progenitor Cell

CREBBP: CREB Binding Protein gene

CRT: Cyclic Reversible Termination

CSR: Class-switch recombination

Cul1: Cullin-1

D

ddNTPs: dideoxynucleotide triphosphates

DE lymphoma: Double-expressor lymphoma

del(1p36): Deletion of 1p36

del(6q): Deletion of 6q

del(7q): Deletion of 7q

del(11q): Deletion of 11q22-q23

del(13q): Deletion of 13q14

del(17p): Deletion of 17p13

DH: Double hit

DLBCL: Diffuse large B-cell lymphoma

DLEU1: Deleted genes in leukemia-1

DLEU2: Deleted genes in leukemia-2

DNA: Deoxyribonucleic acid

dNTPs: deoxynucleotide triphosphates

dsDNA: double strand DNA

DTX1: Deltex E3 Ubiquitin Ligase 1

E

EBF1: EBF Transcription Factor 1

EBV: Epstein-Barr virus

emPCR: emulsion PCR

EP300: E1A Binding Protein P300 gene

EZH2: Enhancer of zeste homolog 2

F

FBXW7: F-box and WD40 repeat domain containing-7

FDCS: Follicular dendritic cell sarcoma

FFPE: Formalin-Fixed Paraffin-Embedded

FGF18: Fibroblast Growth Factor 18

FISH: Fluorescence In Situ Hybridization

FL: Follicular Lymphoma

FLN-C: Filamin-C

FOXO1: Forkhead box O1

FOXP1: Forkhead box protein P1

G

G-banding: Giemsa banding

GC: Germinal center

GCB: Germinal center B-cell

GC-rich regions: Guanina-Citosina rich regions

gDNA: genomic DNA

GNA13: G Protein Subunit Alpha 13

GQ: Genotype Quality

GRCh37: Genome Reference Consortium Human Build 37

H

HC2: Hemocyanin subunit type 2

HCL: Hairy cell leukemia

HHV-8: Human herpesvirus 8

HIV: human immunodeficiency virus

I

IARC: International Agency for Research on Cancer

ICGC: International Cancer Genome Consortium

icr: Intermediate breakpoint cluster region

ID3: Inhibitor Of DNA Binding 3

IG: Immunoglobulin Gene

IGH: Immunoglobulin Heavy Gene

IGHV: Immunoglobulin heavy chain vaiable region

IgA: Immunoglobulin A

IgD: Immunoglobulin D

IgG: Immunoglobulin G

IGK: Immunoglobulin kappa locus

IGL: Immunoglobulin light gene

IgM: Immunoglobulin M

Indels: insertions and deletions

ING3: Inhibitor of growth family member 3

IMIM: Institut Hospital del Mar d'Investigacions Mèdiques

IRAK4: Interleukin 1 Receptor Associated Kinase 4

IRF4/ MUM1: Interferon Regulatory Factor 4

ISCN: International System for Human Cytogenetics Nomenclature

ITPKB: Inositol-Trisphosphate 3-Kinase B

J

JAK: Janus Kinase

K

Kb: Kilobase

KLF2: Krüppel-like Factor 2

KLHL6: Kelch-like protein 6

KMT2C/MLL3: Lysine N-methyltransferase 2C

KTM2D/MLL2: Histone-lysine N-methyltransferase 2D

L

LB: Luria-Bertani

LOH: loss of heterozygosity

LPL: Lymphoplasmacytic lymphoma

M

MACS: Model-based Analysis for ChIPseq

MALTL: Mucosa-associated lymphoid tissue lymphoma

MALT1: Mucosa-associated lymphoid tissue lymphoma translocation protein 1

MAPK: Mitogen activated protein kinase pathway

MAP2K1: Mitogen-Activated Protein Kinase Kinase 1

Mb: Megabase

MBCN: Mature B-Cell Neoplasms

MBR: Major Breakpoint Region

MCL: Mantle Cell Lymphoma

Mcr: Minor breakpoint cluster region

MDM2: Mouse double minute 2 homolog

MDR: Minimal deleted region

MEF2B: Myocyte enhancer binding factor 2B

MEK1: Mitogen-activated protein kinase 1

m7-FLIPI: m7–FL International Prognostic Index

MGR: Minimal gain region

M-IGHV: Mutated immunoglobulin heavy chain vaiable region

MIR15A: MicroRNA 15a

MIR16-1: MicroRNA 16-1

mRNA: messenger RNA

MSX2: Msh Homeobox 2 protein coding gene

MTC: Major translocation cluster

mTORC1: Mammalian target of rapamycin complex 1

MYD88: Myeloid differentiation primary response 88

MZ: Marginal Zone

MZL: Marginal Zone Lymphoma

N

NFκB: Nuclear factor-kappa B

NGS: Next-Generation Sequencing

NHL: Non-Hodgkin Lymphoma

NOTCH1: Notch receptor 1

NOTCH2: Notch receptor 2

O

OS: Overall Survival

P

P2RY8: P2Y Receptor Family Member 8

PAX5: Paired box 5

PAXS: Paired box genes

PB: Peripheral Blood

PBMC: Peripheral Blood Mononuclear Cells

PCR: Polymerase Chain Reaction

PEST domain: peptide sequence rich in proline (P), glutamic acid (E),

serine (S), and threonine (T)

PHA: Phytohemaglutinin

PI3K: phosphatidylinositol 3-kinase

POT1: Protection of telomere 1

PRDM1/BLIMP1: PR domain zinc finger protein 1

PTEN: Phosphatase and tensin homolog

PV1: Panel Version 1
PV2: Panel Version 2

Q

Q-banding: Quinacrine banding

qRT-PCR: quantitative Real-Time PCR

R

RAG: Recombination-Activating Gene

RB1: Retinoblastoma

R-banding: Reverse banding

RNA: Ribonucleic acid

RT-PCR: Reverse Transcription PCR

S

S1PR2: Sphingosine-1-Phosphate Receptor 2

SBL: Sequencing By Ligation

SBS: Sequencing By Synthesis

SDRPL: Splenic Diffuse Red Pulp small B-cell Lymphoma

SF3B1: Splicing factor 3B subunit 1

*SHH: S*onic hedgehog

SHM: Somatic hypermutation

SLL: Small Lymphocytic Lymphoma

SMYD1: SET and MYND Domain Containing 1

SMZL: Splenic Marginal Zone Lymphoma

SNP: Single Nucleotide Polymorphismsm

SNV: Single Nucleotide Variants

SOCS1: Suppressor Of Cytokine Signaling 1

SOX11: SRY-Box Transcription Factor 11

ssDNA: single-stranded DNA

STAT: Signal Transducer and Activator of Transcription

SV: Structural Variations

SYK: Spleen tyrosine kinase

T

TBX21: T-box 21

TCF3: Transcription factor 3

TCGA: The Cancer Genome Atlas

TdT: Terminal deoxynucleotidyl Transferase

tFL: Transformed Follicular Lymphoma

TH: Triple hit

TLR: Toll-like receptor

TLX3: T-cell leukemia, homeobox protein 3

TNFAIP3: TNF Alpha Induced Protein 3

TP53: Tumor protein P53

TPA: 12-O-Tetradecanoylphorbol-13-acetate

TRAF2: TNF receptor-associated factor 2

TRAF3: TNF receptor-associated factor 3

TRAP: Triiodothyronine Receptor Auxiliary Protein

TTFT: Time To First Treatment

U

U2AF1: U2 small nuclear RNA auxiliary factor 1

UCSC: University of California Santa Cruz

UM-IGHV: Non-mutated immunoglobulin heavy chain vaiable región

3'-UTR: three prime untranslated region

V

VAF: Variant Allele Frequency

VCF: Variant Call File

vHCL: Variant Hairy Cell Leukemia

W

WES: Whole-Exome Sequencing

WGS: Whole-Genome Sequencing

WHO: World Health Organization

WT: Wild Type

Y

YAC: Yeast artificial chromosome

List of tables and figures

List of tables

List of figures

Introduction

1. Introduction to mature B-cell neoplasms

Hematological neoplasms represent the fifth most frequent malignant tumors in the western world. Within these, mature B-cell neoplasms (MBCN) account for 90% of B-cell lymphomas, and constitute approximately 4% of all new cancer cases each year. MBCN represent a broad and heterogeneous group of diseases characterized by clonal proliferation of mature B-cells at different stages in the bone marrow, peripheral blood or in other tissues. The World Health Organization (WHO) includes more than 30 entities in this category (Table 1), being the follicular lymphoma (FL) and the diffuse large B-cell lymphoma (DLBCL) the most frequent types of MBCN, which together make up >60% of all lymphomas other than Hodgkin lymphoma and plasma cell myeloma. However, the frequency of each entity varies according to geographical location (Figure 1). Most types of mature B cell neoplasms have a male predominance and a median age at diagnosis of between 60 and 70 years (Swerdlow et al., 2016).

Table 1. WHO classification of mature B-cell neoplasms. The neoplasms addressed in this thesis are identified in yellow (Swerdlow et al., 2016).

WHO classification of Mature B-cell neoplasms:

- Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL)
  - Monoclonal B-cell lymphocytosis (MBL)
- B-cell prolymphocytic leukemia (B-PLL)
- Splenic marginal zone lymphoma (SMZL)
- Hairy cell leukaemia (HCL)
- Splenic B-cell lymphoma/leukemia, unclassifiable
  - Splenic diffuse red pulp small B-cell lymphoma (SDRPL)
  - Hairy cell leukemia variant (HCL-v)
- Lymphoplasmacytic lymphoma (LPL)
- lgM Monoclonal gammopathy of undetermined significance (MGUS)
- Heavy chain diseases (HCDs)
  - Mu heavy chain disease
  - Gamma heavy chain disease
  - Alpha heavy chain disease
- Plasma cell neoplasms
  - Non-lgM monoclonal gammopathy of undetermined significance (non-lgM MGUS)
  - Plasma cell myeloma (PCM)
  - Plasma cell myeloma variants
    - Smouldering (asymptomatic) plasma cell myeloma (SPCM)
    - Non-secretory myeloma
    - Plasma cell leukaemia

- o Plasmacytoma
  - ▪ Solitary plasmacytoma of bone
  - ▪ Extraosseous plasmacytoma
- Monoclonal immunoglobulin deposition diseases
  - o Primary amyloidosis
  - o Light chain and heavy chain deposition diseases
- Plasma cell neoplasms with associated paraneoplastic syndrome
  - o POEMS syndrome
  - o TEMPI syndrome
- Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT lymphoma)
- Nodal marginal zone lymphoma (NMZL)
  - o Paediatric nodal marginal zone lymphoma
- Follicular lymphoma (FL)
  - o Testicular follicular lymphoma
  - o In situ follicular neoplasia (ISFN)
  - o Duodenal-type follicular lymphoma
- Paediatric-type follicular lymphoma (PTFL)
- Large B-cell lymphoma with IRF4 rearrangement
- Primary cutaneous follicle centre lymphoma (PCFCL)
- Mantle cell lymphoma (MCL)
  - o Leukemic non-nodal mantle cell lymphoma
  - o In situ mantle cell neoplasia (ISMCN)
- Diffuse large B-cell lymphoma, NOS (DLBCL)
- High grade B-cell lymphoma (HGBL), with rearrangements of *MYC* and *BCL2* and/or *BCL6*
- T-cell/histiocyte-rich large B-cell lymphoma (THRLBCL)
- Primary diffuse large B-cell lymphoma of the CNS
- Primary cutaneous diffuse large B-cell lymphoma, leg type
- EBV-positive diffuse large B-cell lymphoma, NOS (EBV+ DLBCL, NOS)
- EBV-positive mucocutaneous ulcer (EBVMCU)
- Diffuse large B-cell lymphoma associated with chronic inflammation
  - o Fibrin-associated diffuse large B-cell lymphoma
- Lymphomatoid granulomatosis (LYG)
- Primary mediastinal (thymic) large B-cell lymphoma (PMBL)
- Intravascular large B-cell lymphoma
- ALK-positive large B-cell lymphoma (LBCL)
- Plasmablastic lymphoma (PBL)
- Primary effusion lymphoma (PEL)
- HHV8-associated lymphoproliferative disorders
  - o Multicentric Castleman disease (MCD)
  - o HHV8-positive diffuse large B-cell lymphoma, NOS
  - o HHV8-positive germinotropic lymphoproliferative disorder (GLPD)
- Burkitt lymphoma (BL)
- Burkitt-like lymphoma with 11q aberration
- High-grade B-cell lymphoma (HGBL)
  - o High-grade B-cell lymphoma (HGBL), with *MYC* and *BCL2* and/or *BCL6* rearrangements
  - o High-grade B-cell lymphoma (HGBL), NOS
- B-cell lymphoma, unclassifiable, with features intermediate between DLBCL and classic Hodgkin lymphoma
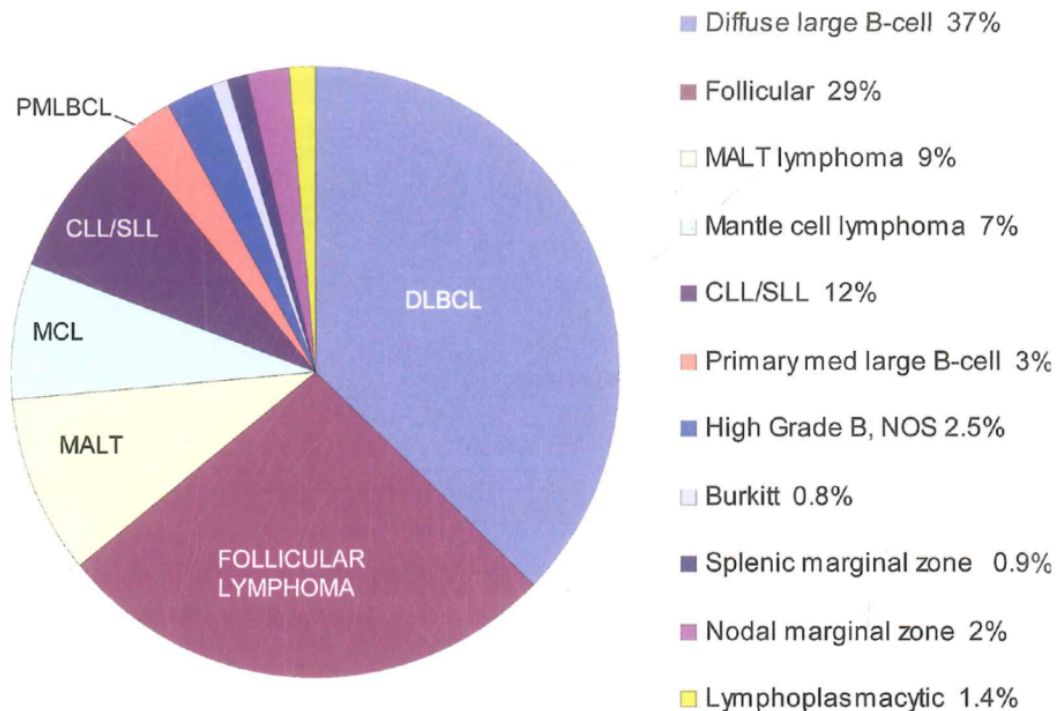
Figure 1. WHO described frequency of different mature B-cell neoplasms (Swedlow et al., 2008)**.**

Mature B-cell neoplasms seem to mimic diverse stages of normal B-cell differentiation. This resemblance to normal cellular stages is the basis for its nomenclature and classification (Rickert et al., 2013) (Figure 2). In this regard, MBCN originate from B-cells that have left their niche in the bone marrow. At this point, these cells present surface immunoglobulins (IG) and a functional B cell receptor (BCR). Then, a clonal expansion will occur in any moment of the cell maturation process. In fact, the same genetic mechanisms that enable the development and maturation of normal B-cells are frequently involved in their malignant transformation (Küppers et al., 2001; Küppers et al., 2005). Some lymphomas originate from naive B-cells that have not yet passed through the germinal center (GC), such as mantle cell lymphoma (MCL) and a subset of chronic lymphocytic leukemia (CLL). However, the vast majority of lymphomas derived from GC B-cells, as it is the case for CLL, FL, Burkitt lymphoma (BL), and DLBCL.

GCs are the histological structures dedicated to the generation and the selection of B-cells that produce high-affinity antibodies (Klein & Dalla-Favera, 2008; Victora & Nussenzweig, 2012). To this end, GC B-cells proliferate and undergo two events that allow the diversification of the immunoglobulin genes, which are somatic hypermutation (SHM) and class-switch recombination (CSR). The cellular pathways regulating GC formation and maintenance are often involved in the malignant transformation process that leads to lymphomagenesis. In fact, a common feature of many B-cell neoplasms is that they present chromosomal translocations involving IG genes and SHMs, both of which are dependent on IG remodelling mechanisms including V(D)J recombination, SHM and CSR. Besides the common elements, each MBCN is characterized by distinct genetic alterations that are often major determinants of their phenotypes (Figure 3) (Morgan et al., 2014); Basso & Dalla-Favera, 2015). Finally, there are tumors that originate from post GC B-cells and that have become memory B-cells, such as lymphoplasmacytic lymphoma (LPL), marginal zone lymphoma (MZL) and plasma cell neoplasms (Figure 2) (Swerdlow et al., 2016).
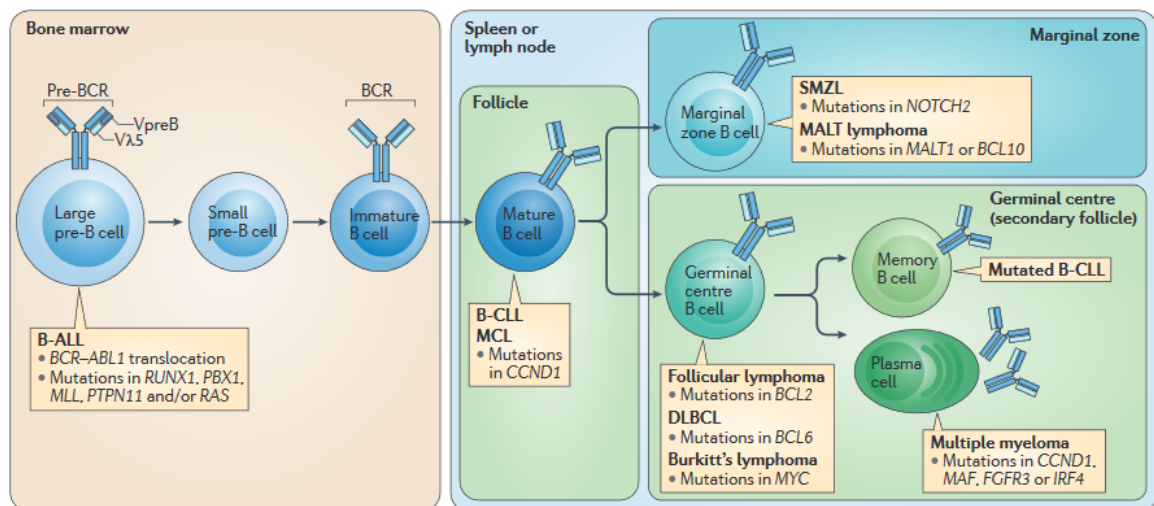
Figure 2. Normal B-cell differentiation and its relationship to major B-cell neoplasms. Hematologic neoplasms that occur at each stage and their characteristic genetic alterations are detailed in the boxes (Rickert et al., 2013).
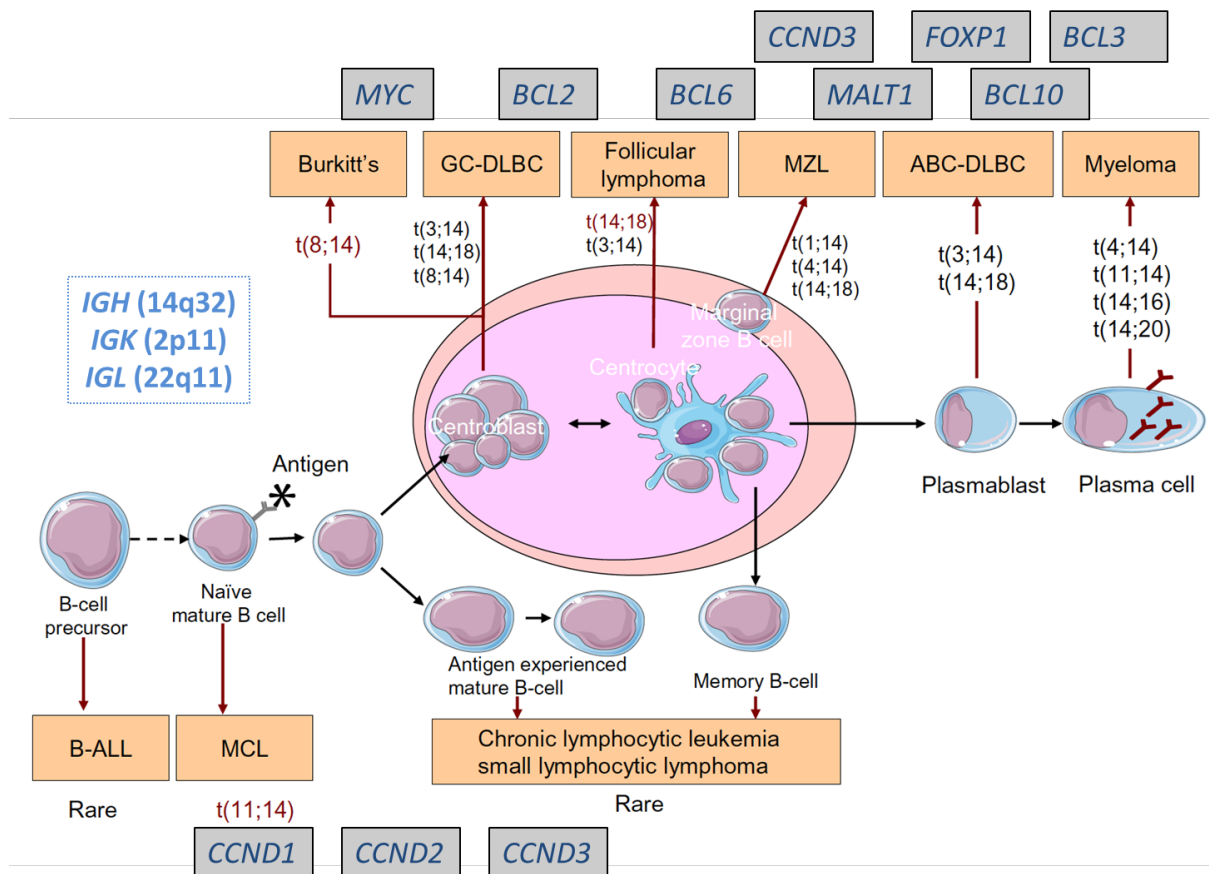
Figure 3. The most common genetic lesions associated with the major subtypes of B-cell lymphomas. (Adapted from Morgan et al., 2014).

However, some neoplasms can show lineage heterogeneity or plasticity, in exceptional cases, or do not clear correspond to a normal B-cell differentiation stage, such as hairy cell leukemia (HCL). Consequently, the normal counterpart of neoplastic cells cannot be used as the sole basis for its classification. Thus, the 2008 WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues schema employed a multiple-parameter approach based on clinical, morphologic and biologic features, keeping in mind that a precise separation between entities is not always possible (Swerdlow et al., 2008).

In the last decade, after the complete sequencing of the human genome, a technological revolution has taken place with the development of the next generation of sequencing (NGS) techniques (Meyerson et al., 2010). The application of these techniques in the B-cell lymphoid neoplasms have led to major advances

not only in better understanding the pathology, but also to discoveries of diagnostic importance. Therefore, since the WHO classification was published in 2008, disease definitions have continued to evolve and expand, with new entities and variants being recognized in the 2016 WHO revision (Swerdlow et al., 2016). Furthermore, in this update the WHO highlights the importance of the genetic evaluation for the classification and prognostic stratification of several B-cell neoplasms.

Regarding the etiology of MBCN, diverse factors that could promote the development of several types of mature B-cell lymphomas have been described. One known major risk factor seems to be an abnormality of the inmune system. Immunodeficient patients have a markedly increased incidence of B-cell neoplasms, especially for DLBCL and BL, even though there is no evidence of immune system abnormality in most MBCN patients (Salavoura et al., 2008). Lesions in genes involved in the control of lymphocyte apoptosis have also been associated with a higher risk of both autoimmune diseases and B-cell lymphomas (Straus et al., 2001). Furthermore, genome-wide association studies have identified a substantial number of single nucleotide polymorphismsm (SNP), the majority in immunoregulatory genes, which are associated with increased risk of lymphoma (Forrest et al., 2006; Wang et al., 2007). Another factor that has been shown to contribute to the development of different MBCN is the presence of infectious agents. Epstein-Barr virus (EBV), human herpesvirus 8 (HHV-8) and Hepatitis C virus are some of the agents that appear to influence lymphoma development through the activation of the B-cell immune response. On the other hand, bacteria, or at least the immune response to their antigens, have also been associated with the pathogenesis of B-lymphomas, in particular to mucosa-associated lymphoid tissue (MALT) lymphoma (Troppan et al., 2015; Swerdlow et al., 2016). Finally, epidemiological studies have associated the exposure to some environmental factors, such as herbicides and pesticides, to the development of FL and DLBCL (Hartge et al., 2005; Colt et al., 2006).

2. Genetic abnormalities in mature B-cell neoplasms

MBCN constitute a heterogeneous group of disorders that can often have overlapping clinical and/or morphological features and, consequently, can present diagnostic challenges. Moreover, patients with relatively homogeneous pathologic findings can have highly variable clinical outcomes, what makes its clinical management even more difficult. In the last decades the approach to the diagnosis of lymphoid neoplasms has changed drastically, from solely morphologic assessment to the integration of clinical, immunophenotypic, morphologic, cytogenetic, and molecular genetic findings. While histomorphology and immunophenotipe are sufficient for the diagnosis of many lymphomas, genomic studies carried out in recent years have improved the understanding of B-cell lymphomagenesis and have revealed novel molecular biomarkers with diagnostic, prognostic and predictive value in different mature B-cell neoplasms. The integration of these molecular biomarkers has proven to be crucial in the diagnosis and management of B-cell neoplasms, what is reflected in the evolving classification of B-cell malignancies, as well as in the consensus guidelines for patient management (Swerdlow et al., 2016).

The most frequent and relevant genetic alterations in the most common entities of mature B- cell neoplasms are detailed below, being these entities the ones included in our study.

2.1. Chronic lymphocytic leukemia/small lymphocytic lymphoma

Chronic lymphocytic leukemia/small lymphocytic lymphoma (SLL) is the most common leukemia in adults in the Western world, with an estimated incidence of 4 to 6 cases per 100,000 individuals per year, and with a median age of onset of 65 years (Swerdlow et al., 2016). It is a neoplasm of small, monomorphic B-cells, usually coexpressing CD5, CD200 and CD23, with a characteristic weak expression of B-cell markers CD20, CD22, CD79b and surface immunoglobulin light chains. However, some patients may have an atypical CLL immunophenotype (Mora et al., 2019). CLL is characterized by a highly heterogeneous clinical course with some patients

showing an asymptomatic disease, while others require therapeutic interventions due to early progression and short survival (Zenz et al., 2010; Zenz et al., 2012). Several clinical and biological prognostic factors have been identified that can be used in clinical practice to break through this heterogeneity.

Regarding genetic aberrations, a variety of characteristic, well-stablished cytogenetic abnormalities are seen in CLL, including deletions of 13q14, 11q22-q23, 17p13, 6q21 and trisomy 12. Importantly, none of these alterations is specific for CLL, so the main reason for their detection is their role as prognostic markers of the disease (Döhner et al. 2000) (Figure 4A).

Deletion of 13q14 [del(13q)] represents the most frequent alteration, which occurs in 50-60% of all cases. The 13q14 deletions are heterogeneous and include different chromosome bands, some of which also include the retinoblastoma gene (*RB1*). However, all cases present a minimal deleted region (MDR) that contains the deleted genes in leukemia-1 and 2 (*DLEU1* and *DLEU2*), as well as the *MIR15A* and *MIR16-1* microRNAs (Calin et al., 2002). Although del(13q) as sole alteration has been reported as good prognosis marker, differences in the clinical course of the disease have been observed associated to the size of the deleted region (type II deletions including *RB1* gene showed shorter survival), and to the percentage of nuclei with this alteration (when the evidence is >70%). Besides, the prognostic relevance of biallelic vs monoallelic 13q deletion remains controversial (Puiggros et al., 2013).

Trisomy 12 is present in approximately 20% of cases and is associated with an atypical morphologic and immunophenotypic features (Autore et al., 2018). CLL patients carrying this alteration constitute a heterogeneous group with an intermediate prognosis. It has also been linked to aberrations in the *NOTCH1* gene (Balatti et al., 2012) and to Richter transformation, both events associated with a more aggressive disease and inferior clinical outcome.

Deletion of 11q22-q23 [del(11q)] occurs in 10-20% of CLL and is often associated with bulky lymphadenopathy, rapid progression, and reduced overall survival (OS) (Döhner et al., 1997). It mainly involves deletion of the *ATM* (ataxia telangiectasia

mutated) gene, that plays a key role in the protection of genome integrity by activating DNA-repair pathways, or by inducing apoptosis when the damage cannot be repaired (Edelmann et al., 2012). In addition, 20-30% of del(11q) CLL cases show concurrent *ATM* aberrations. Similarly, only a few proportion of *ATM* mutated CLL (10-20%) presents a concomitant 11q22-q23 deletion (Skowronska et al., 2012). Conversely, there are a small number of cases carrying the del(11q) with *ATM* gene intact, but have a loss of the neighboring *BIRC3* gene instead (Rossi et al., 2012). *BIRC3* can be disruptive in CLL by single nucleotide variants (SNV), deletions or a combination of both. Loss of *BIRC3* leads to constitutive activation of the NFκB pathway through downregulation of the p53 protein via MDM2, and is a poor prognosis marker independent of other factors.

Deletion of 17p13 [del(17p)] is found in 5-8% of cases at diagnosis and leads to the loss of tumor suppressor gene *TP53,* and is associated with chemotherapy resistance and poor survival (Döhner et al., 1995; Gonzalez et al., 2011). Lesions of *TP53* are found in ~5-7% of patients with CLL, and have also been associated with very poor prognosis (Monti et al., 2020). Among cases with del(17p), more than 80% display aberrations in the remaining *TP53* allele. *TP53* lesions without del(17p) are much more rare, but exhibit a similar effect on chemotherapy response and OS, and are also associated with higher genetic complexity in CLL. In addition, the frequency of *TP53* lesions and/or deletions has been seen to increase in the disease progression and reaches approximately 40% in refractory CLL (Buccheri et al., 2018; Monti et al., 2020).

Deletions involving the long arm of chromosome 6 [del(6q)], especially 6q21, have been identified in 5-7% of CLL patients, and frequently display peculiar morphology and are associated with an intermediate prognosis (Cuneo et al., 2004). It is currently unknown which genes are affected, but a recent study revealed the presence of both SNV and deletions involving *ZNF292* (at 6q15) in 3% and 2% of cases, respectively. However, the outcome of patients carrying the deletions but not the SNV seems to be worse, suggesting different mechanisms of alteration of this gene in CLL patients (Puente et al., 2015).

As a last reference to cytogenetic alterations, the presence of a complex karyotype (CK), generally defined as the finding of ≥3 abnormalities in two or more metaphases, has been associated with inferior outcomes and worse response to treatments, including novel drugs (Puiggros et al., 2017). Moreover, a recent study from Baliakas et al. confirmed that high CK, carrying ≥5 chromosomal aberrations, emerges as prognostically adverse, independently of other biomarkers (Figure 4B)
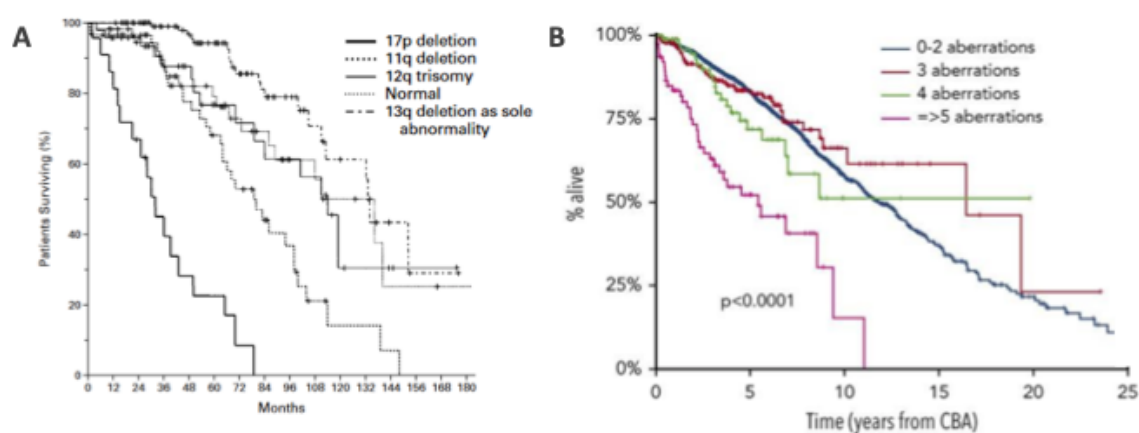
(Panagiotis Baliakas et al., 2019).

Figure 4. Kaplan-Meier plots for OS and (A) the five cytogenetic categories determined by FISH; (B) the number of cytogenetic aberrations by chromosome banding analysis. (adapted from Döhner et al., 2000 and Baliakas et al., 2019).

In addition, several studies based on whole-genome sequencing (WGS) and whole-exome sequencing (WES) have identified more than 40 recurrently mutated driver genes in CLL, being the most frequent mutations those involving *SF3B1, TP53*, and *NOTCH1* genes (5-15% of all CLL cases) (Landau et al., 2015; Puente et al., 2015) (Figure 5). Mutations in some of these drivers, like *NOTCH1, TP53, BIRC3, and SF3B1*, are associated with a poor outcome and/or aggressive disease and refractoriness to chemotherapy (Stilgenbauer et al., 2014). Conversely, mutations in *MYD88, CHD2*, and *KLHL6* are found mainly in patients with more indolent disease (Quesada et al., 2012). Besides, novel recurrent mutations in non-coding regions have been recently

described, which include mutations of the 3' UTR of *NOTCH1*, that are associated with more aggressive disease, and mutations in an enhancer region of chromosome 9p13, resulting in the reduced expression of *PAX5* (Puente et al., 2015). Interestingly, *TP53* and *NOTCH1* mutations are strong predictors of worse outcome, even when they are presented in small subclones (<1%), suggesting the clinical relevance of the subclonal status in CLL (Nadeu et al., 2016).
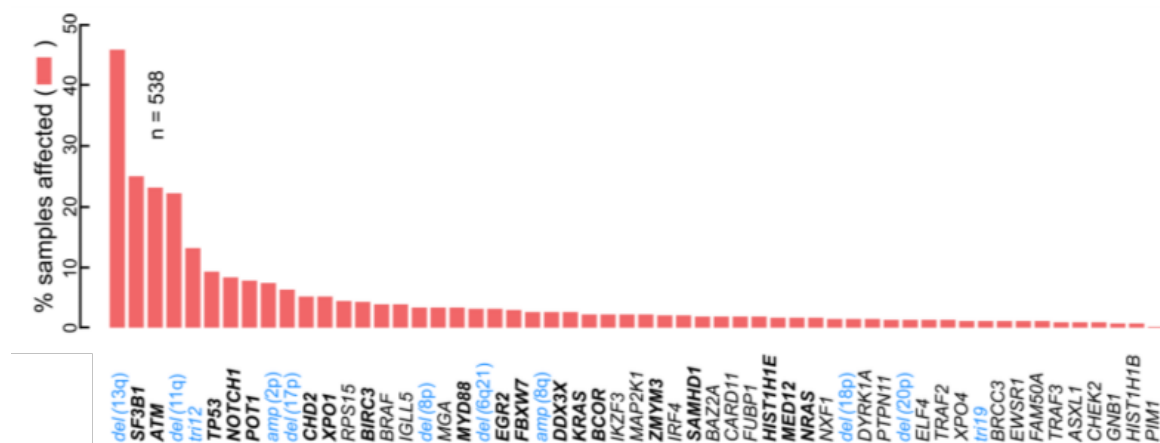
Figure 5. The landscape of putative driver gene mutations and recurrent somatic copy number variations in chronic lymphocytic leukemia. Colums represent the frequency of somatic mutations and somatic copy number alterations found by WES. Recurrent somatic CNA are labeled in blue, and putative CLL cancer genes previously identified in bold (Adapted from Landau et al., 2015).

Integration of mutational and cytogenetic data has evidenced the association between certain cytogenetic alterations and particular mutations, such as *SF3B1* mutations with del(11q), *NOTCH1* and *FBXW7* mutations with trisomy 12, and *MYD88* with del(13q), and has led to the stratification of CLL patients into four prognostic subgroups: (1) high-risk, harboring *TP53* and/or *BIRC3* abnormalities; (2) intermediate-risk, harboring *NOTCH1* and/or *SF3B1* mutations and/or del(11q); (3) low-risk, harboring +12 or a normal genetics; (4) and very low-risk, harboring

del(13q) only, whose 10-year survival did not show significant differences from a matched general population (Rossi et al., 2013) (Figure 6).
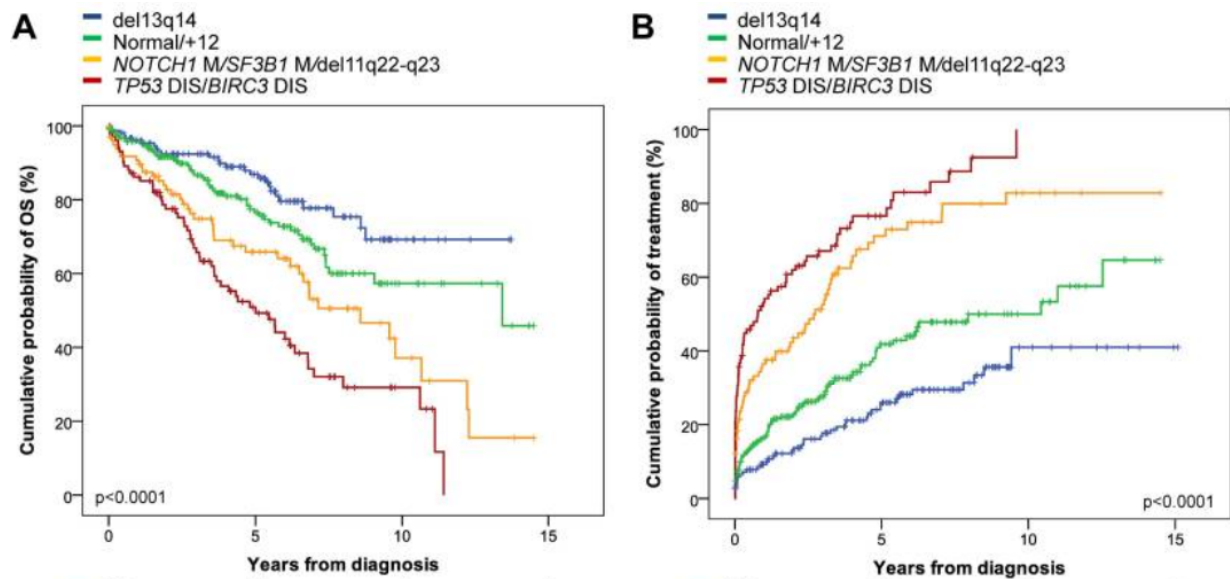
Figure 6. The four subgroups of chronic lymphocytic leukemia based on the integration of their cytogenetic and genetic alterations. Kaplan-Meier estimates of OS (A) and probability of progressive disease requiring treatment according to IWCLL-NCI guidelines as indicated by treatment-free interval (B). (Rossi et al., 2013)

Clinical significance of the main genetic biomarkers in chronic lymphocytic leukemia is shown in Table 2.

Table 2. Clinical significance of the main genetic biomarkers in chronic lymphocytic leukemia.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
13q14 deletion	Loss of *DLEU2/miR-15a/16-1 locus* which has a tumor-suppressor role and modulate de proliferation	50%-60% of all cases	Considered an early event in the disease: similar frequency in MBL and often detectable as a single lesion	Good (strong)	
Trisomy 12	Unknown	20% of all cases	Associate with an atypical morphologic and immunophenotypic features	Good/intermediate (strong) Increased risk of progression to aggressive B-cell lymphoma transformation	
11q22 (*ATM*) deletion	Loss of 1 copy of the *ATM* gene; DNA repair, NF-κβ signaling	10-20% of all cases	Associate with bulky lymphadenopathy, rapid progression	Poor (strong) Shorter OS	
17p13 (*TP53*) deletion	Loss of 1 copy of the *TP53* gene; DNA repair, cell cycle control, apoptosis, senescence, DNA repair	5-8% of cases at diagnosis Enriched frequency in progressive cases (40%)	Associate with *NOTCH1* and *SF3B1* mutation Adverse prognostic factors: UM-IGHV status, advanced Binet stages, higher levels of CD38 expression	Poor (strong) Increased risk of progression to aggressive B-cell lymphoma transformation (50-60% frequency in RS) Shorter TTT, PFS, OS	Chemo-refractoriness to fludarabine-containing regimens
TP53 mutation	DNA repair, cell cycle control, apoptosis, senescence, DNA repair	5-7% of all cases Enriched frequency in progressive cases (40%)	Associate with *NOTCH1* and *SF3B1* mutation Adverse prognostic factors: UM-IGHV status, advanced Binet stages, higher levels of CD38 expression	Poor (strong) High-risk of progression to aggressive B-cell lymphoma transformation (40% frequency in RS) Shorter TTT, PFS, OS	Chemo-refractoriness to fludarabine-containing regimens

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
NOTCH1 mutation	Ligand-activated transmembrane protein regulating downstream pathways involved in cell growth mRNA splicing and processing	5-10% newly diagnosed cases Enriched frequency in progressive cases (15-20%)	Correlate with adverse clinical parameters: UM-IGHV, trisomy 12, higher levels of CD38 and ZAP-70 expression	Poor (strong) High-risk of lymphomas, progression to aggressive DLBCL in RS (present in 90% of the cases) Sorter OS, PFS and TTT	Decreased benefit of addition of rituximab to fluradabine treatment Benefit from anti-CD52 treatment
SF3B1 mutation	mRNA splicing and processing	4-12% in early CLL Enriched frequency in progressive cases (17-24%)	Association with adverse prognostic factors: UM-IGHV, advanced Binet stages, higher levels of CD38 expression Association with del(11q) Less frequently present in association with del(13q); mutually exclusive trisomy 12	Poor (strong) Shorter TTT and OS Risk stratification of patients with other cytogenetic abnormalities (shorter OS in patients with del(11q) and del(13q)	Chemo-refractoriness to fludarabine-containing regimens
ATM mutation	PI3-/PI4-kinase family, cell cycle checkpoint kinase	12% of newly diagnosed cases	Associate with bulky lymphadenopathy, rapid progression	Poor (intermediate) Shorter TTT and PFS	Chemo-refractoriness to alkylating agent or purine analog treatment
BIRC3 mutation	DNA repair, NF-κβ signaling	4% of newly diagnosed cases Enriched mutations in fludarabine-resistant-patients (24%)	Correlation with UM-IGHV, trisomy 12, del(11q) Exclusive from *TP53* mutations	Poor Shorter OS	Chemo-refractoriness to fludarabine-containing regimens
MYD88 mutation	NF-κβ signaling, TLR signaling	1.5-4% of newly diagnosed cases	Association with favorable prognostic factors: younger age (≤50 years), low expression of CD38 and ZAP-70, higher frequency on M-IGHV	Good No difference with the age-matched healthy population	No treatment needed

Abbreviations: OS: overall survival; PFS: progression-free survival; RS: Richter Syndrome; TTT: time to first treatment; mRNA: messenger RNA; IGHV: immunoglobulin heavy chain variable region; UM-IGHV: unmutated immunoglobulin heavy chain variable region; M-IGHV: mutated immunoglobulin heavy chain variable region; MBL: monoclonal B-cell lymphocytosis.

(Adapted from Onaindia et al., 2017)

2.2. Splenic marginal zone lymphoma

SMZL is a rare low-grade B-cell lymphoma. The median age at diagnosis is 69 years, and the overall age-adjusted incidence is 0.13 per 100,000 habitants per year, with most of the patients being Caucasian and with an increasing trend to male predominance (Matutes et al., 2008). SMZL originates from B memory lymphocytes present in the marginal zone of secondary lymphoid follicles (Zinzani et al., 2012). The diagnosis of SMZL is based on a combination of features that include lymphocyte morphology, immunophenotype, cytogenetic abnormalities as well as bone marrow (BM) and spleen histology (when available) (Swerdlow et al., 2016). Pathological cells appear as a small to medium sized mature B-cells with round or oval nuclei and condensed chromatin, basophilic cytoplasm, and most of the cases present with typical irregular membrane projections called villi (Traverse-Glehen et al., 2011). There is no specific immunophenotypic pattern for SMZL, however these cases usually express CD19, CD20, CD22, CD79a, CD79b, FMC7 and IgM and are negative for CD5, CD10, CD43, bcl-6, cyclin D1 (CCND1) and CD103 (Behdad & Bailey, 2014). The clinical course is usually indolent, with median overall survival of 10 years and more than 60% of patients are alive at 5 years (Olszewski & Castillo, 2013; Swerdlow et al., 2016).

Cytogenetic abnormalities are present in approximately 80% of SMZL (Salido et al., 2010; Robinson & Cutucache., 2018). The most common chromosomal abnormality is the 7q deletion [del(7q)], occurring in 30% to 40% of patients. Its frequency is much higher in SMZL than in similar B-cell neoplasms and it has been found as a single anomaly in some patients. For this reason, it has been proposed as the primary diagnostic marker (Salido et al., 2010; Watkins et al., 2010). The main region resulting in the loss of heterozygosity (LOH) has been identified between 7q21 and 7q33, but the precise chromosomal location responsible, and the mechanisms underlying this alteration, remain unknown (Watkins et al., 2010; Robledo et al., 2011). Some potentially important genes described in this region include the protection of telomere 1 (POT1), sonic hedgehog (SHH), cullin-1 (Cul1), the enhancer of zeste homolog 2 (EZH2), Filamin-C (FLN-C), and caveolin-1 (CAV1) genes (Parry et al., 2013). In addition, several miRNAs located in the 7q region, have been proposed

to be significant contributors in SMZL pathogenesis (Watkins et al., 2012; Robinson & Cutucache., 2018). However, none of the molecular studies carried out have been able to demonstrate an essential role in the pathogenesis of the disease for any of these genes. Further, the direct effects of the del(7q) on SMZL also remain controversial (Robinson & Cutucache., 2018).

In addition to del(7q), gains of 3q (20-30% of cases) and 12q (15-20% of cases) are the most frequent cytogenetic aberrations. Trisomy 3 is a common abnormality in marginal zone B-cell lymphomas, especially in nodal and extranodal MALT lymphomas (Spina & Rossi, 2017). However, unlike other marginal cell lymphomas, in SMZL the gain specifically occurs at 3q and it has been suggested that a gene dosage effect for genes localized in this region could be involved in the development and/or disease progression of marginal zone B-cell lymphomas (Spina & Rossi, 2017). Besides, it is frequently associated with CK. Moreover, unbalanced structural aberrations of 3q associated with gain of this region are also frequent in SMZL patients. Owing to the variety of breakpoints and partners involved in 3q rearrangements, these structural aberrations seem to represent secondary events in the pathogenesis (Salido et al., 2010).

Translocations involving the IG heavy (IGH) or light (IGL) chain loci are controversial, although most authors agree that they are uncommon (Matutes et al., 2008; Salido et al., 2010). Partner chromosomes include 3q27 (*BCL6*), 6p21 (*CCND3*), 7q21 (*CDK6*), 9p13 (*PAX5*), 11p11 and 10q24, and may occur either as primary or secondary cytogenetic abnormalities. Conversely, translocations associated with MALT lymphoma, such as t(11;18)(q21;q21), t(14;18)(q32;q21) or t(3;14)(p14;q32), have not been found in SMZL (Matutes et al., 2008). Additionally, deletions of 17p13 (*TP53*) have been observed in 3-17% of cases (Parry et al., 2015).

WES studies have revealed an average of 30 alterations/case in the genomic landscape of SMZL, showing an intermediate genomic complexity degree between aggressive lymphomas, namely DLBCL (90 nonsilent mutations per case) and other indolent lymphomas, such as untreated CLL (12 nonsilent mutations per case) (Fabbri et al., 2011; Pasqualucci et al., 2011; Rossi et al., 2012). The genetic

alterations associated with this lymphoma are predominantly involved in the signaling pathways that regulate marginal zone differentiation, including NOTCH pathway (30%), via the alternative mutation of multiple genes (*NOTCH2, NOTCH1, SPEN, DTX1*, and *SMYD1*), with *NOTCH2* being the most frequently mutated gene (20%); NF-κB pathway, via mutation of *TNFAIP3/A20, BIRC3,* and *TRAF3*; and the B-cell receptor pathway, via mutation of *CARD11* (Figure 7) (Robinson & Cutucache, 2017; Spina & Rossi, 2017). All *NOTCH2* mutations observed in SMZL cause disruption of the protein inhibitory PEST domain, which activate *NOTCH2* signaling, and seem to be relatively specific for SMZL (Rossi et al., 2012). More recently, *KLF2* mutations have been identified as the most frequent somatic changes in SMZL (up to 20%) and appear to comprise a subset with a distinct genotype (Clipson et al., 2015; Campos-Martín et al., 2017). In particular, *KLF2* mutations that cluster in the C2H2 domain are an independent poor prognostic factor in SMZL, and are early clonal events usually found in association with del(7q) (53% with *KLF2* mutation vs 11% without mutation), unmutated IGHV (UM-IGHV) (50% vs 7%), IGHV1-2*04 usage, and other gene mutations including *NOTCH2, TNFAIP3*, and *ARID1A*, suggesting a possible cooperation between genetic abnormalities and B-cell receptor configuration in contributing transformation (Parry et al., 2015). Monoallelic *KLF2* deletions have also been described in 11% of SMZL cases, and it seems that mutations and deletions are mutually exclusive, revealing a predominant allele distribution (Piva et al., 2015; Campos-Martín et al., 2017). Additional recurrently mutated genes have been described, including mutations in *MYD88* (13%), and *KTM2D/MLL2* (4.7%). Del(7q), UM-IGHV, *TP53* inactivation, and *NOTCH2* and *KLF2* mutations have been associated with histological signs of progression and to a worse outcome in SMZL (Peveling-Oberhag et al., 2015; Campos-Martín et al., 2017). In addition, *TNFAIP3* mutations have been associated with an increased risk of high-grade transformation. UM-IGHV, *KLF2*, and *NOTCH2* mutations have been related to shorter time-to-first treatment, whereas *TP53* and *MYD88* mutations are predictors of short and long overall survival, respectively. Nevertheless, *NOTCH2* and *TP53* mutations remain the only alterations that appear as independent markers of poor survival (Parry et al., 2015; Campos-Martín et al., 2017).
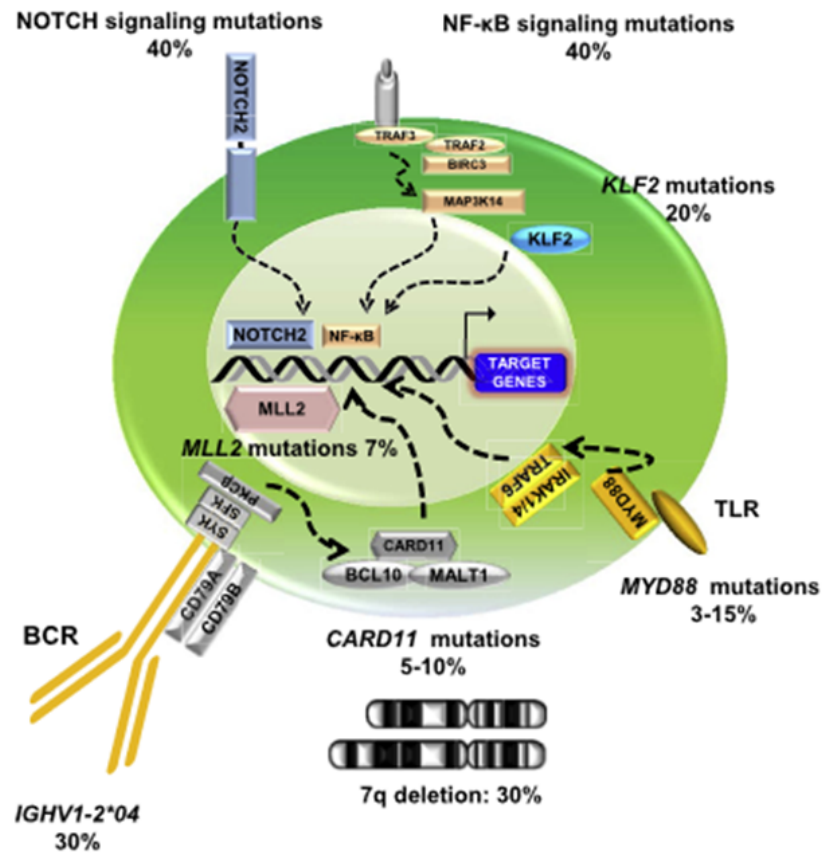
Figure 7. Key molecular alterations in splenic marginal zone lymphoma. (Adapted from Spina & Rossi, 2017)

Clinical significance of the main genetic biomarkers in splenic marginal zone lymphoma is shown in Table 3.

Table 3. Clinical significance of the main genetic biomarkers in splenic marginal zone lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
7q21-q33 deletion	Telomeric processing, and cytoskeleton reorganization	30%-40% of all cases	Considered as a primary diagnostic marker: the most common chromosomal abnormality in SMZL and often detectable as a single lesion Associated with *KLF2* mutation and UM-IGHV	Intermediate/poor (intermediate) Associated with histological signs of progression and worse outcome	
3q gain	Transcriptional regulation, and NF-κβ signaling Suggested to be involved in the development and/or disease progression of MZ lymphomas	20%-30% of all cases	Structural variants are also frequent, with a result of gain of the regions. Considered as a secondary event in the pathogenesis of SMZL. Frequently associated with complex karyotypes	Unknown	
17p13 (*TP53*) deletion	Loss of 1 copy of the *TP53* gene	3%-17% of all cases	Associated with histological signs of progression	Poor (strong) Independent marker of poor survival. Strong predictor of worse outcome and short OS, PFS	
KLF2 mutation/deletion	NF-κβ signaling, downregulating most of the NF-κβ-mediated activities, and TNFR signaling; cell differentiation, proliferation, activation, and trafficking	>20% mutations/11% deletions	Associated with histological signs of progression and with adverse prognostic factors: 7q deletion, UM-IGHV, IGHV1-2*04 usage, and mutations in *NOTCH2*, *TNFAIP3*, and *ARID1A* genes Mutations and deletions are mutually exclusive	Poor (strong) Associated with worse outcome and shorter TTT	

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
NOTCH2 mutation	NOTCH signaling, marginal zone differentiation	20% of all cases	Associated with histological signs of progression	Poor (strong) Strong predictor of worse outcome and shorter TTT, OS, PFS Independent marker of poor survival	
TP53 mutation	DNA repair, cell cycle control, apoptosis, senescence, DNA repair	16% of all cases	Associated with histological signs of progression	Poor (strong) Independent marker of poor survival. Strong predictor of worse outcome and short OS, PFS	
TNFAIP3 mutation	NF-κβ signaling	~10% of all cases	Associated with *KLF2* mutations	Poor (intermediate) Increased risk of high-grade transformation	
MYD88 mutation	NF-κβ signaling, TLR signaling	13% of all cases		Good Longer OS	

Abbreviations: IGHV: immunoglobulin heavy-chain variable region genes; UM-IGHV: unmutated immunoglobulin heavy-chain variable region genes; OS: overall survival; PFS: progression-free survival; TTT: time to first treatment.

(Adapted from Onaindia et al., 2017)

2.3. Hairy cell leukemia

HCL is a rare indolent B-cell malignancy accounting for 2% of lymphoid leukemias. The median age at diagnosis is ~50 years, with a strong male predominance and a substantially higher incidence in caucasian populations. Typical clinical features of HCL include splenomegaly, pancytopenia and monocytopenia, with few circulating neoplastic cells, and marked susceptibility to infections (Swerdlow et al., 2016).

Hairy cell leukemia comprises the clonal hematologic malignancies of classical (cHCL) and variant (vHCL). Although this two entities share some pathological features, they display different morphology and immunophenotype. Classical hairy cells are small to medium-sized lymphoid cells with an oval or kidney-shaped nucleous that is typically presented without nucleoli, and abundant and pale blue cytoplasm with circumferential "hairy" projections. The immunophenotypic profile includes the co-expression of CD11c, CD25, CD103, CD20, CD22, CD123, CD200, TBX21, annexin A1, FMC7, and CCND1. Conversely, vHCL appears with prominent nucleoli, blastic/convoluted nuclei, and/or absence of circumferential "hairy" contours, and have a negative expression of CD25, annexin A1, TRAP, CD123, CD24, and HC2 markers (Matutes et al., 2003). In addition, these two entities also show differences in their clinical course. While cHCL has a median survival of 20 years, vHCL presents with a chronic clinical course with a median estimated survival of 9 years, and 6% of progress to a more aggressive disease (Matutes et al., 2003). Besides, IGHV4-34 usage has been shown to be a poor prognostic factor in both types of HCL, and is related to poorer OS and lack of response to treatment (Arons et al., 2011).

A WES study carried out in 2011 revealed the presence of *BRAF* V600E mutation in almost all cHCL cases, but not in vHCL nor in cHCL with IGHV4-34 usage or other small B-cell neoplasms, providing an easily testable diagnostic marker (Tiacci et al., 2011). However, rare exceptions have been reported, with this mutation found in splenic lymphoma (Raess et al., 2013), single HCL cases lacking this mutation (Langabeer et al., 2013), and other HCL cases with different *BRAF* mutations (Tschernitz et al. 2014). *BRAF* is an important member of the mitogen activated protein kinase (MAPK) pathway and is the most frequently mutated protein kinase in cancer (Pakneshan et al., 2013; Dankner et al., 2018). Besides, a WES study revealed

a high prevalence (48%) of activating mutations in *MAP2K1* gene (another essential component of the *MAPK* signal transduction pathway, kinase *MEK1*, which is downstream of *BRAF*) in both vHCL and HCL IGHV4-34, but not in cHCL, suggesting a relationship between these subtypes and their distinction from cHCL (Waterfall et al., 2013). Although activating *MAPK* mutations are critical for both cHCL and vHCL, additional shared cooperating alterations, as well as disease-specific alterations have been described as novel drivers of HCL, including *BRAF, CDKN1B* (16%), and *KMT2C* (15%) in cHCL and *MAP2K1, TP53* (38%), *KMT2C* (25%), *CCND3* (13%), and *U2AF1* (13%) in vHCL (Durham et al., 2017). These findings may have therapeutic relevance, since it is suggested that *CCND3* mutations confer sensitivity to CDK4/CDK6 inhibitors, whereas mutations in *U2AF1* confer sensitivity to spliceosome inhibitors (Schmitz et al., 2012; Shirai et al., 2017).

Regarding cytogenetic alterations, the most recurrent are deletions of chromosomes 7q and 13q, and gains of chromosome 5. Two MDRs at 7q31.33-q33 have been identified in both cHCL and vHCL, and contain a total of 33 known genes. Deletion 7q [del(7q)] are rare in other lymphoproliferative disorders, with the exception of SMZL, a disease that shares some features with HCL (Matutes et al., 2003). The occurrence of del(7q) in these disorders suggests that deregulation of key tumor suppressor genes that reside within this region may be linked to their disease pathogenesis. As seen in CLL, recurrent del(13q) include the tumor suppressor *RB1* and the *MIR15A* / and *MIR16*-1 microRNA cluster at 13q14.3 (Durham et al., 2017). Gains on chromosome 5 have been reported to be highly frequent in both cHCL and vHCL. This alteration can appear as trisomy or gain of 5q, in which a minimal gain region (MGR) of approximately 19 Mb has been described at 5q34-q35.3 *locus*, encompassing over 200 known genes, including *TLX3, FGF18, BNIP1, MSX2* and *GFGR4* (Hockley et al., 2011). Although no distinct chromosomal abnormality has been described for vHCL, CKs that include structural abnormalities on 8q24 and deletions and/or mutations of *TP53* were more frequently identified in this entity, the latter being associated with poor outcome (Hockley et al., 2011; Matutes et al., 2015). Table 4. sumarizes the Clinical significance of the main genetic biomarkers in hairy cell leukemia.

Table 4. Clinical significance of the main genetic biomarkers in hairy cell leukemia and hairy cell leukemia variant.

Biomarker	Biologic function	Frequency	Clinical presentation	Prognostic value (evidence)	Predictive value
7q deletion	Telomeric processing, and cytoskeleton reorganization	20-30% of all cases		Unknown	
13q14 deletion	Loss of *DLEU2/miR-15a/16-1 locus* which has a tumor-suppressor role and modulate de proliferation			Unknown	
Trisomy 5/Gain 5q	Unknown	~30% of all cases		Unknown	
17p13 (*TP53*) deletion	Loss of 1 copy of the *TP53* gene	~30% of all cases Higher incidence in vHCL	Associate with adverse prognostic factors: UM-IGHV status, CK, and *TP53* mutations	Poor (strong) Poor outcome Increased risk of progression to aggressive B-cell lymphoma transformation	Poor response to chemoimmunotherapy. Resistance to purine analogs
TP53 mutation	DNA repair, cell cycle control, apoptosis, senescence, DNA repair	~30% of all cases Higher incidence in vHCL	Associate with adverse prognostic factors: UM-IGHV status, CK, and *TP53* deletions	Poor (strong) Poor outcome Increased risk of progression to aggressive B-cell lymphoma transformation	Poor response to chemoimmunotherapy. Resistance to purine analogs

Biomarker	Biologic function	Frequency	Clinical presentation	Prognostic value (evidence)	Predictive value
BRAF mutation (V600E)	MAPK signaling	~100 cHCL	Expression of phosphorylated MEK and ERK, which indicates a constitutive activation of the MAPK pathway Also found in HSCs in HCL patients	Unknown	
MAP2K1 mutation	MAPK signaling	~48% of vHCL and HCL IGHV4-34 cases	Constitutive activation of the MAPK pathway	independently related to poorer overall survival	lack of response to treatment

Abbreviations: IGHV: immunoglobulin heavy-chain variable region genes; UM-IGHV: unmutated immunoglobulin heavy-chain variable region genes; CK: complex karyotype; cHCL: classic hairy cell leukemia; vHCL: variant hairy cell leukemia; HSCs: hematopoietic stem cells

(Adapted from Onaindia et al., 2017)

2.4. Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue

MALT lymphoma is a low-grade B-cell lymphoma that accounts for 7-8% of all B-cell lymphomas, most cases occurring in adults with a median age of onset of 70 years (Swerdlow et al., 2016). It is composed of morphologically heterogeneous small B cells including marginal zone (centrocyte-like) cells, cells resembling monocytoid cells, small lymphocytes, and scattered immuno-blasts and centroblast-like cells. In some cases, there is also a plasmacytic differentiation. Altogether, MALT lymphoma cells share the same cytological and immunophenotypical (CD20+, CD21+, CD35+, IgM+, and IgD−) features as marginal zone (MZ) B-cells, what gave rise to its name (Swerdlow et al. 2016). MALT lymphoma represent a heterogeneous group of lymphoid neoplasms that arise at a wide range of different extranodal sides, including the stomach (70%), lung (14%), ocular adnexa (12%), thyroid (4%), and small intestine (1%) (Isaacson et al., 2004). Although they share many characteristics, MALT lymphomas also present site-specific differences with respect to etiology, morphology, molecular cytogenetic abnormalities and clinical course (Kuper-Hommel & Van Krieken, 2012).

Regarding cytogenetic alterations, four main recurrent chromosomal translocations, which frequency depends on the primary site, have been associated with the pathogenesis of MALT lymphomas: t(1;14)(p22;q32), t(11;18)(q21;q21), t(14;18)(q32;q21), and t(3;14)(p14.1;q32) (Troppan et al., 2015).

Other cytogenetic abnormalities include trisomies 3, 12, and/or 18. They are usually present as a single abnormality in 22% of cases, but are often associated with one of the four major translocations described above (Streubel et al., 2004). Trisomy 3 has been described as the most common aberration in gastrointestinal MALT lymphomas, with a frequency of up to 35% (Taji et al., 2005). Although its contribution to lymphomagenesis has not yet been experimentally addressed, several genes appear to be involved in lymphomagenesis, such as the protooncogene *BCL6*, the transcription factor *FOXP1*, and the chemokine receptor *CCR4*, which are differentially overexpressed in MALT lymphomas with trisomy 3 (Deutsch et al., 2008).

Regarding genetic alterations, most affect NF-*κ*B signal pathway-related genes, causing its constitutive activation (Kuper-Hommel and Van Krieken, 2012). In fact, *BIRC3* and *TRAF3* are mutated in about 5% of cases and are responsible for non-canonical NF-*κ*B pathway activation. Furthermore, inactivation of *TNFAIP3* (*A20*) by deletion and/or mutation has also been reported, which promotes NF-*κ*B activation by abolishing auto-negative feedback.   Besides, MALT lymphomas are also characterized by frequent lesions affecting chromatin remodeling, BCR and NOTCH pathways (Cascione et al., 2019). However, lymphomas from different anatomic sites exhibit a different spectrum of genetic lesions, and further, none of the mutations described has been shown to have an impact on the management of these patients.

Clinical significance of the main genetic biomarkers in MALT lymphomas is shown in Table 5.

Table 5. Clinical significance of the main genetic biomarkers in MALT lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value	Predictive value
t(11;18)(q21;q21)	*BIRC3-MALT1* fusion gene	15%-40% of all cases	Considered as diagnostic marker In most cases appears as the sole chromosomal alteration.		
t(14;18)(q32;q21)	*IGH-MALT1* fusion gene	15-20% of all cases	Considered as diagnostic marker Frequently associated with other cytogenetic abnormalities.		
t(1;14)(p22;q32)	*IGH-BCL10* fusion gene	1-2% of all cases	Considered as diagnostic marker Frequently associated with other cytogenetic abnormalities.		
t(3;14)(p14.1;q32)	*IGH-FOXP1* fusion gene	-	Considered as diagnostic marker Frequently associated with other cytogenetic abnormalities.		
Total/partial trisomy 3	Transcriptional regulation, and NF-κB signaling Suggested to be involved in the development and/or disease progression of MZ lymphomas	35% of all cases	Often associated with one of the four major translocations. However, in 22% of cases is found as a single abnormality.		

Abbreviations: MZ: marginal zone

(Adapted from Onaindia et al., 2017)

2.5. Follicular lymphoma

FL is a GC derived B-cell lymphoma that accounts for 20% of all new cases of non-Hodgkin lymphoma, representing the second most common non-Hodgkin lymphoma in the western world (Horwitz et al., 2016; Swerdlow et al., 2016). The median age at diagnosis is 59 years with a male to female ratio of 1:1.7, and a higher incidence (2-3 times) in caucasian than in black populations (Groves et al., 2000). FL is considered an indolent lymphoma characterized by a slow progression over many years. However, the clinical course of FL patients can be surprisingly variable, with up to 30% of cases transforming to more aggressive lymphomas (usually DLBCL), reflecting the biologic heterogeneity of the disease (Swerdlow et al., 2016).

FL is typically composed of the two types of B-cells normally found in GC, such as centrocytes and centroblasts. In fact, FL cases are divided into three main histologic grades based on the relative proportion of centrocytes and centroblasts. Grade 1 and 2 FL cases are predominantly composed of centrocytes with low proliferation rates, whereas grade 3 FL cases (which are further subdivided into grades 3a and 3b) have increasing numbers of actively proliferating centroblasts with higher proliferation rates (Swerdlow et al., 2016) (Table 6). Regarding the immunophenotype, tumour cells are usually positive for surface immunoglobulin (lgM with or without lgD, lgG, or rarely lgA) and B-cell associated antigens (CD19, CD20, CD22, and CD79a), in addition to bcl-2, bcl-6, and CD10 and are negative for CDS and CD43. Some cases, most commonly grade 3B FL, lack CD10, but retain bcl-6 expression (Pittaluga et al., 1996; Lai et al., 1999; Ott et al., 2002; Karube et al., 2007). Bcl-2 overexpression is the hallmark of FL and is expressed by a variable proportion of the neoplastic cells. In fact, it is overexpressed in 85-90% of cases with grade 1-2 FL, but in <50% of cases with grade 3 FL (Menendez et al., 2004). In some cases, the apparent absence of bcl-2 protein is due to mutations in the *BCL2* gene that eliminate the epitopes recognized by the most commonly used antibody (Schraders et al., 2005; Masir et al., 2009).

Table 6. Follicular lymphoma grading.

Grading	Definition
Grade 1-2 (low grade)	0-15 centroblasts per HPF
1	0-5 centroblasts per HPF
2	6-15 centroblasts per HPF
Grade 3	> 15 centroblasts per HPF
3A	Centrocytes present
3B	Solid sheets of centroblasts
Reporting of pattern	**Proportion follicular**
Follicular	> 75%
Follicular and diffuse	25-75%
Focally follicular/predominantly diffuse	< 25%
Diffuse	0%

Number of centrobasts corresponds to the abbsolute number of centroblasts per high-power (40x objective, 0.159mm^2) microscopic field (HPF) (Swerdlow et al., 2016)

FL lymphomagenesis is a complex multistep process that occurs through several stages of B-cell differentiation. The acquisition of the t(14;18)(q32;q21), detected in nearly 90% of patients, is usually considered the first oncogenic hit (Huet et al., 2018). This early event places *BCL2* gene, located at 18q21, under the transcriptional control of IGH regulatory regions, at chromosome 14q32, which leads to constitutive expression of the anti-apoptotic protein bcl-2 (Tsujimoto et al., 1985). The rearrangement between *BCL2* and *IGH* is probably owing to the proximity of both loci during VDJ recombination mediated by the VDJ recombination-activating protein (RAG) (Tsujimoto et al., 1985; Kenter et al., 2013). Breaks in *BCL2* are likely due to the inherent fragility of CpG sites and are localized at the 3'-UTR of *BCL2* or downstream the gene, mainly clustered in the major breakpoint region (mbr). In fact, 60-70% of FL cases carrying *BCL2* translocation involves the mbr (Cotter et al., 1991). In addition, smaller clusters of breakpoints have also been described in the intermediate breakpoint cluster region (icr) and the minor breakpoint cluster region (mcr) (Cleary et al., 1986; Batstone & Goodlad, 2005; Weinberg et al., 2007) (Figure 8). Although, t(14;18) provides a survival advantage, there is accumulating evidence that other genetic hits are required for complete transformation to FL, as B-cells

bearing this translocation are also detected at very low levels in the blood in 50–70% of healthy individuals (Schüler et al., 2009).

BCL2 (18q21)
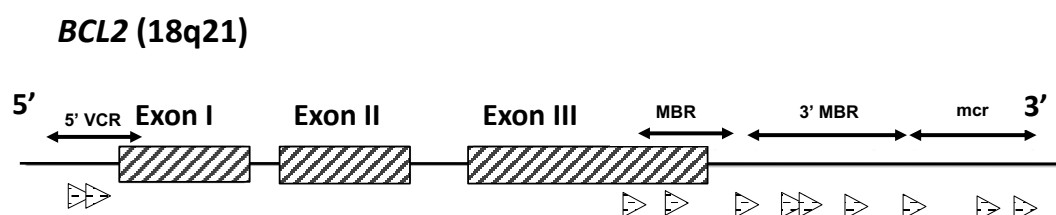
Figure 8. Schematic diagram of the the breakpoints in the *BCL2* gene. Triangles below the gene represent the location of the breakpoints reported in the literature (Adapted from Espinet et al., 2008).

Advancements in NGS technologies have led to remarkable understanding of the genomic landscape of FL (Huet et al., 2018). Chromatin modifier genes, such as histone methytransferases (*KMT2D* and *EZH2*) and acetyltransferases (*CREBBP* and *EP300*) are the hallmark of GC derived tumors and, accordingly, a high prevalence of gain or loss of function mutations in these genes have also been reported in FL. These mutations modify chromatin conformation and disrupt gene transcription in networks that are critical for GC and post-GC B-cell fate. *EZH2* plays a central role in this process and is considered the epigenetic gate keeper of the GC reaction (Béguelin et al., 2013). In addition, *EZH2* mutations are thought to be a common early event in pathogenesis and are potentially targetable with selective *EZH2* inhibitors (McCabe et al., 2012; Bödör et al., 2013). Interestingly, mutations in *KMT2D* and *CREBBP* are also early oncogenic events, and are believed to be present in the so-called common progenitor cell (CPC) population, which contains tumour-initiating cells and acts as a disease reservoir in relapse and histological transformation (Okosun et al., 2014; Horton et al., 2017). In addition, *BCL6*, a key transcriptional repressor that orchestrates the GC reaction, is dysregulated by a variety of alterations (Basso & Dalla-Favera, 2015; Huet et al., 2018). Mutations targeting *MEF2B* are also identified in 15% of cases, and lead to increased transcriptional activity of *MEF2B*, resulting in increased expression of *BCL6* and *MYC*

oncogene, and thus, increase cell migration. Furthermore, mutations in *ARID1A, ARID1B, BCL7A* and *SMARCA4* are also frequent events in FL (Krysiak et al., 2017). Overall, these different mechanisms converge to lock B-cells in the GC phenotype, thus sustaining proliferation programmes and genetic instability in a cell already resistant to apoptosis due to bcl-2 over expression. And, while these mutations are unlikely to be sufficient to initiate lymphoma, they could act in concert with dysregulated bcl-2 expression to induce lymphomagenesis (García-Ramírez et al., 2017). Besides, alterations leading to the constitutive activation of key signalling pathways, such as BCR, JAK–STAT, mTORC1 and NOTCH (Figure 9), are also frequent events that may participate in sustaining the proliferation and survival of tumour cells (Bouska et al., 2017).
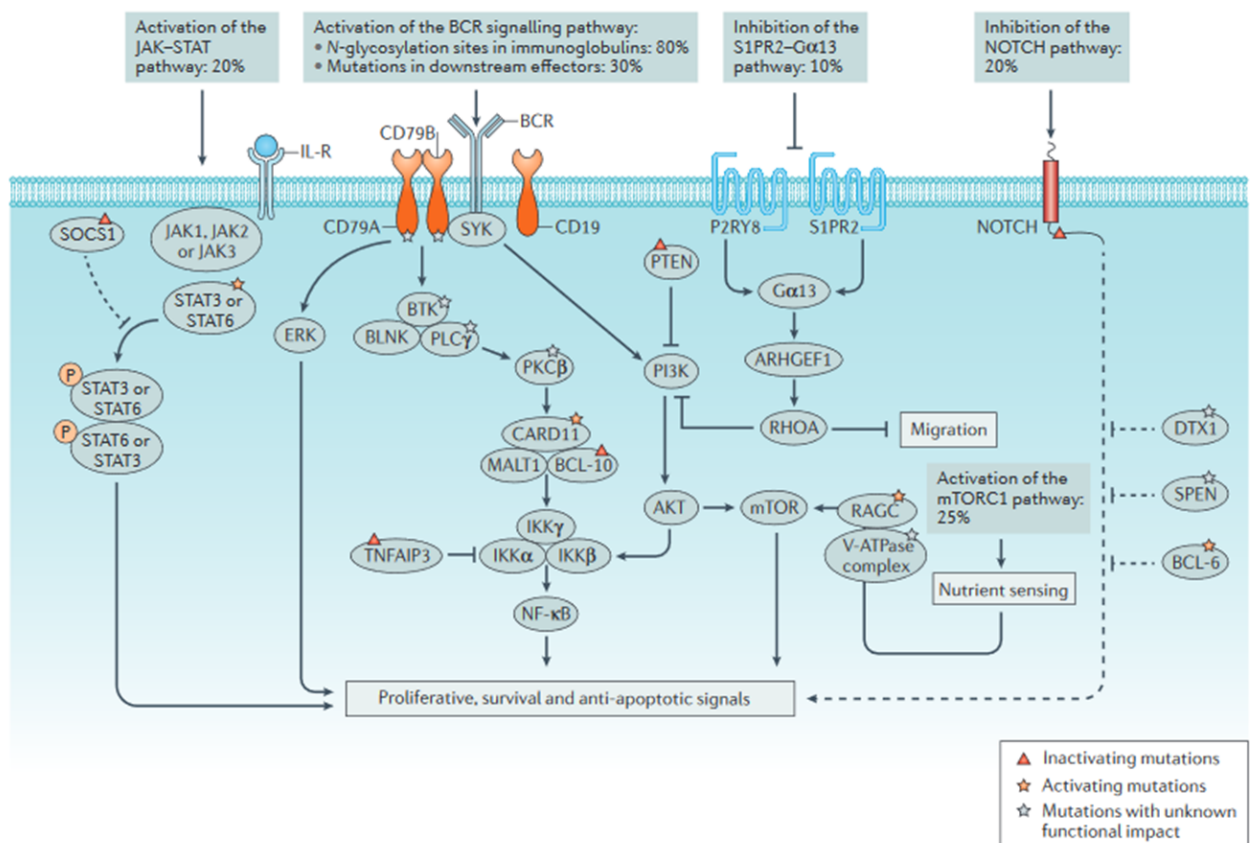
Figure 9. Signals driving proliferation and survival of FL cells. (Adapted from Huet et al., 2018)

Regarding their clinical impact, *Pastore et al.* (Pastore et al. 2015) analyzed the complete coding sequence of 74 genes that mutate recurrently in lymphomas in two independent cohorts of patients who received first-line immunochemotherapy, and established a clinical-genetic risk model (m7-FLIPI) that improves risk stratification by integrating the mutation status of seven genes: *EZH2, ARID1A, MEF2B, EP300, FOXO1, CREBBP* and *CARD11*. The clinical significance of each of these genes is detailed in Table 7.

On the other hand, studies comparing the genetic landscape of transformed follicular lymphoma (tFL) with that found in the initial FL have shown that an increased prevalence of genomic copy number alterations (CNAs) or gene mutations is presented in tFL, which may affect the activation of some oncogenes (for example, gain of 2p16 (*REL–BCL11A* locus), 3q27 (containing *BCL6*) or 8q24 (containing *MYC*)) and/or the inactivation of specific tumour suppressor genes (for example, *TP53* mutations or deletion of *CDKN2A*) (Okosun et al., 2014). In addition, 1p36 deletions [del(1p36)] have been reported in >50% of cases, and are associated with poor prognosis (Cheung et al., 2007). Furthermore, NGS technologies have enabled the identification of new hits in genes involved in signalling pathways (for example, *PIM1, SOCS1, STAT6, MYD88, TNFAIP3* and *ITPKB*), cell cycle (for example, *CCND3*), B-cell development (for example, *EBF1*), GC B-cell dissemination (*GNA13, S1PR2* and *P2RY8*) or immune evasion (for example, *CD58* and *B2M*) (Okosun et al., 2014; Pasqualucci et al., 2014; Kridel et al., 2016). However, as the molecular mechanisms leading to FL transformation are complex and variable among patients, target genes leading the transformation have not yet been elucidated.

Clinical significance of the main genetic biomarkers in folicular lymphoma is shown in Table 7.

Table 7. Clinical significance of the main genetic biomarkers in follicular lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
t(14;18)(q32;q21)	*IGH-BCL2* Apoptosis regulator	90% of all FL cases	Diagnostic marker		
Del(1p36)		>50% of cases		Poor prognosis; associated with tFL	
t(8;14)(q24;q32)	*IGH-MYC* Cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis	Rare in FL; but in 25% of tFL			
EZH2 mutations	Chromatin modification	89% of all FL cases		Low risk of a failure-free survival event after immunochemotherapy	Potentially targetable with selective *EZH2* inhibitors
CDKN2A/B mutations	DNA damage response, cell cycle regulation	46% of tFL	Biallelic loss of *CDKN2A/B* is specifically acquired during follicular lymphoma transformation	Identify a subgroup of patients with high-risk FL	
MYC mutations	Cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis	33% of tFL		Identify a subgroup of patients with high-risk FL	
TP53 mutations	DNA repair, cell cycle control, apoptosis, senescence	6% of FL and 17.9% of tFL	In FL, associated with older age at diagnosis, and higher IPI score.	Identify a subgroup of patients with high-risk FL with shorter OS, and PFS	

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
MEF2B mutations	Chromatin modification	15% of all FL cases		Included in m7-FLIPI	
ARID1A mutations	Chromatin modification	9% of all FL cases		correlated with longer failure-free survival in m7-FLIPI	
EP300 mutations	Chromatin modification	14% of all FL cases		associated with shorter failure-free survival in m7-FLIPI	
FOXO1 mutations	Cell survival			associated with shorter failure-free survival in m7-FLIPI	
CREBBP mutations	Chromatin modification	55% of all FL cases		associated with high-risk in m7-FLIPI	pan-HDAC inhibitors would be beneficial
CARD11 mutations	NF-Kβ signaling			inferior overall survival in m7-FLIPI	

Abbreviations: OS: overall survival; PFS: progression-free survival; IPI: International Pronostic Index score; tFL: transformed follicular lymphoma; FL: follicular lymphoma; m7-FLIPI: m7- Follicular Lymphoma International Prognostic Index

(Adapted from Onaindia et al., 2017)

2.6. Mantle cell lymphoma

Mantle cell lymphoma (MCL) is a rare subtype of small B-cell neoplasm that accounts for approximately 6–9% of all malignant lymphomas, with an annual incidence of 1–2/100,000 cases in Western countries (Swerdlow et al., 2016). It predominantly occurs in elderly men with a median age at diagnosis of 65–70 years (Ruan et al., 2015). MCL is an aggressive lymphoma that displays poor prognosis, with median OS of about 3 years, and clinical course characterized by continuous relapses and increasing resistance to chemotherapy (Goy & Kahl, 2011). Nevertheless, minority of patients present an indolent clinical course (Campo & Rule, 2015). MCL is usually composed of monomorphic small to medium-sized lymphoid cells with irregular nuclear contours, dispersed chromatin and inconspicuous nucleoli, that closely resemble centrocytes (Swerdlow et al., 2016). Regarding the immunophenotype, cells express relative intense surface IgM/IgD, more frequently with lambda than kappa restriction and are usually positive for SOX11, cyclin D1, bcl-2, CDS, FMC7, CD43 and sometimes also for IRF4/ MUM1; and are negative for CD10, bcl-6 and CD23 (which can also be weakly positive) (Swerdlow et al., 2016). However, aberrant phenotypes have also been described, including absence of CDS and expression of CD10 and bcl-6, which are sometimes associated with blastoid or pleomorphic variants (Gao et al., 2009; Akhter et al., 2015).

Biologically, the chromosomal translocation t(11;14)(q13;q32) is considered the genetic hallmark of MCL and plays a key role in its pathogenesis (Jares et al., 2012). It juxtaposes *CCND1* to the *IGH* enhancer, resulting in cyclin D1 overexpression, which ultimately induces deregulation of the cell cycle, in addition to other biological processes such as DNA repair or transcriptional regulation (Jares et al., 2012). This translocation involves many different breakpoints. In fact, in *CCND1*, approximately 30–50% of breakpoints are located in a 2 Kb region known as the major translocation cluster (MTC), and more breakpoints have also been localized within the close vicinity of the 5'-end of the *CCND1* gene (Tsujimoto et al., 1984; Greisman et al., 2011) (Figure 10). Besides, *IGH* breakpoints are clustered in a chromosomal region 5' of the joining genes (JH). A part from the translocations, amplifications of the translocated allele have rarely been observed (Beà et al., 2009). Nevertheless, there

are few MCLs (<5% of the cases) that lack the translocation t(11;14) and do not overexpress cyclin D1. Instead, these cases seem to frequently express higher levels of cyclin D2 (*CCND2*) or cyclin D3 (*CCND3)*, and in 55% of these lymphomas the *CCND2 locus* is rearranged, emphasizing the importance of cyclin D overexpression in MCL lymphomagenesis (Salaverria et al., 2013).

CCND1 (11q13)
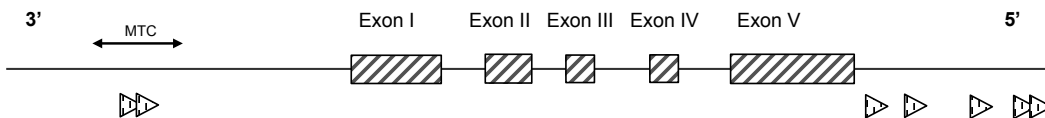
Figure 10. Schematic diagram of the breakpoints in the *CCND1* gene. Triangles below the gene represent the location of the break points reported in the literature

Remarkably, secondary genetic alterations and deregulated pathways are required to initiate and promote MCL lymphomagenesis (Figure 11) (Jares et al., 2012).
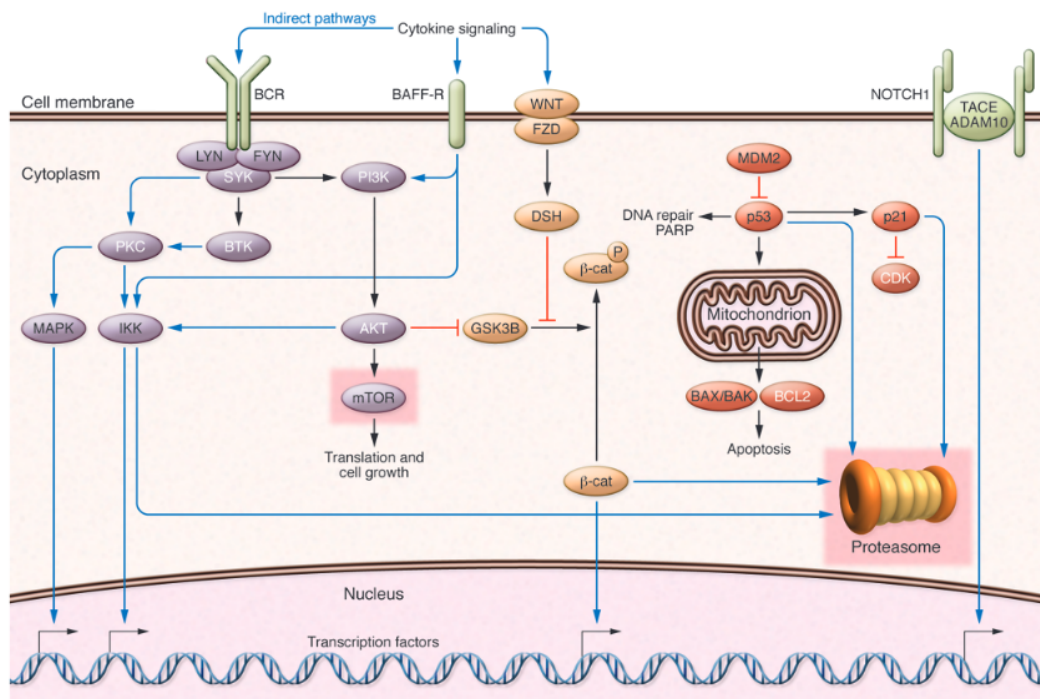
Figure 11. Major aberrant pathways in mantle cell lymphoma. (Adapted from Jares et al., 2012).

ATM and *TP53* genes are among the most commonly mutated which play a pivotal role in DNA damage response and regulation of apoptosis (Beà et al., 2013). Mutations in *ATM* predominantly exist either on both *ATM* alleles or they co-occur with 11q deletions of the wildtype allele. And, whereas *ATM* mutations are considered early events in MCL oncogenesis, *TP53* mutations seem to occur later in lymphomagenesis and confer a dismal prognosis (Beà et al., 2013). In addition to these alterations, other aberrations causing resistance to apoptosis have also been observed, which include deletions of *BCL2L11* and amplification of the genomic locus 18q21, causing a consecutive overexpression of the anti-apoptotic protein bcl-2 (Hartmann et al., 2010). Genetic alterations involving the INK4a/CDK4/RB1 and the ARF/MDM2/p53 pathways are also common mechanisms causing disruption of cell cycle control in MCL (Jares et al., 2012). In addition, the *CDKN2A* locus, encoding the CDK inhibitor INK4a and the positive p53 regulator *ARF*, is frequently deleted in MCL (Hartmann et al., 2010). Besides, some MCL patients also harbor mutations in chromatin modifiers such as *WHSC1*, *MLL2*, and *MEF2B* (Beà et al., 2013; Zhang et al., 2014). Moreover, constitutive activation of both the classical and the alternative NF-κβ pathway has also been detected in MCL (Figure 12) (Vogt et al., 2017). Constitutive B-cell receptor (BCR) signaling is associated with activation of the classical NF-κβ pathway in MCL and different genetic aberrations have been associated with constitutive BCR-NF-κβ signaling, including *SYK* amplifications and overexpression, activating mutations in the coiled-coil domain of *CARD11*, which interestingly seem to occur in relapsed tumors more often than at diagnosis, and loss of the negative NF-κβ regulator *TNFAIP3/A20* (Rahal et al., 2014; Wu et al., 2016). Besides, somatic mutations affecting *BIRC3* and *TRAF2* lead to constitutive activation of the alternative NF-κβ pathway (Figure 12), which are mutually exclusive, pointing out a non-redundant mechanism of action (Vogt et al., 2017). Furthermore, these aberrations are potentially and clinically relevant, as they are associated with poor prognosis and resistance to the BTK inhibitor ibrutinib (Rahal et al., 2014). In addition, some cases with *BIRC3* mutations also harbor deletions of the wildtype allele, suggesting that homozygous *BIRC3* loss might be advantageous for MCL cells (Beà et al., 2013). Besides, some MCL patients harbor mutations in *NOTCH1* and *NOTCH2* genes (Kridel et al., 2012). These mutations are predominantly caused by

truncation or small frame shifting indels, and are associated with a poor prognosis (Beà et al., 2013). Finally, deletions or mutations of the *CCND1* gene have also been observed, resulting in cyclin D1 overexpression, and are associated with a higher proliferation index, shorter overall survival, and ibrutinib resistance (Wiestner et al., 2007; Mohanty et al., 2016).
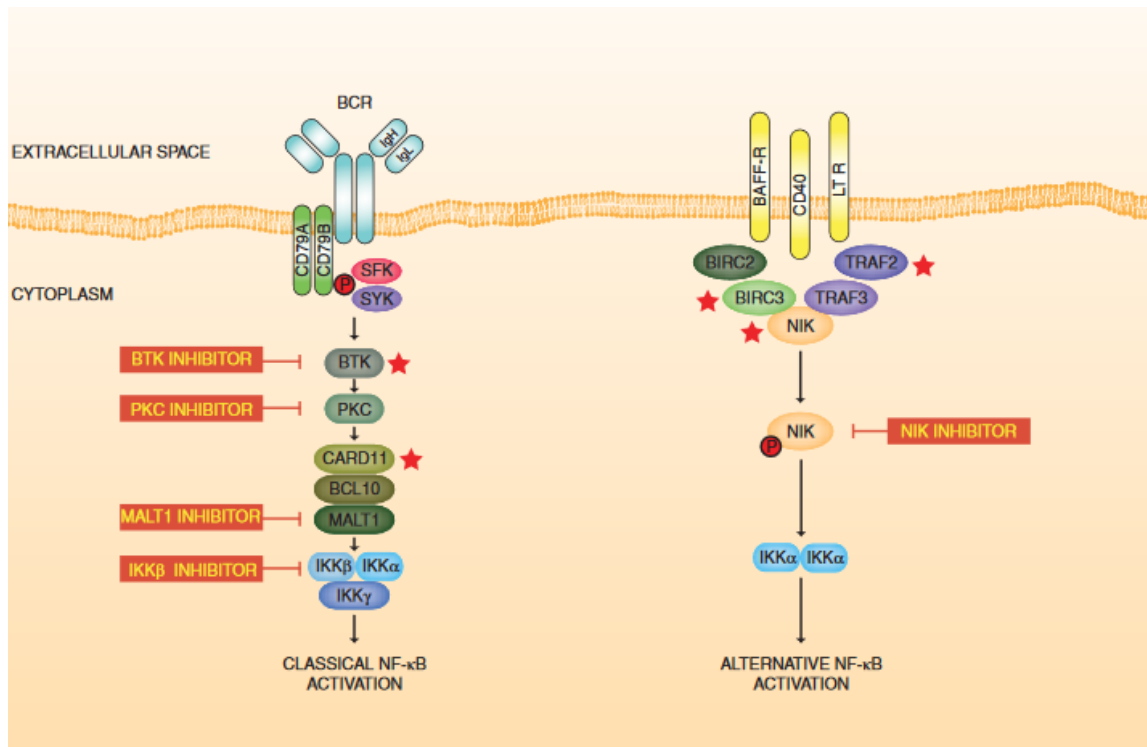
Figure 12. Constitutive activation of NF-κβ pathway in mantle cell lymphoma. Potential targets for small molecule inhibitors are highlighted. Red asterisks indicate SNV that have been detected in primary MCL samples. Abbreviations: IgH: immunoglobulin heavy chain; IgL:immunoglobulin light chain. (Adapted from Vogt et al., 2017)

Clinical significance of the main genetic biomarkers in mantle cell lymphoma is shown in Table 8.

Table 8. Clinical significance of the main genetic biomarkers in mantle cell lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
t(11;14)(q13;q32)	IGH-*CCND1* Cell cycle, transcriptional, and DNA repair process regulation	>90% of all MCL cases	Diagnostic marker		
CCND2 rearrangements	IGK-*CCND2;* IGL-*CCND2;* IGH-*CCND2,* among others	Rare	Diagnostic marker		
CCND3 rearrangements	*IGH-CCND3,* among others	Rare	Diagnostic marker		
TP53 mutations	DNA repair, cell cycle control, apoptosis, senescence	7-20% of cases	More frequent in advanced stages of the disease	Poor prognosis	Poor or no response to firs-line treatments
BIRC3 deletions and/or mutations	Inhibitor of apoptosis	15% of cases		Poor prognosis	Resistance to the BTK inhibitor ibrutinib
TRAF2 mutations	Regulation of the activation of NF-κβ and JNK and regulation of cell survival and apoptosis	Up to 35% of cases		Poor prognosis	Resistance to the BTK inhibitor ibrutinib

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value (evidence)	Predictive value
CCND1 mutations	Cell cycle regulation	Up to 35% of cases	Associated with higher proliferation index	Shorter OS	Resistance to the BTK inhibitor ibrutinib
NOTCH1 mutations	Regulation of cell proliferation, apoptosis and differentiation	12% of cases		Poor OS	
NOTCH2 mutations	Regulation of cell proliferation, apoptosis and differentiation	5.2% of cases		Poor OS	

Abbreviations: OS: overall survival

(Adapted from Onaindia et al., 2017)

2.7. Diffuse large B-cell lymphoma

Diffuse large B-cell lymphoma (DLBCL) is the most common type of lymphoid malignancy worldwide, accounting for about one third of all non-Hodgkin lymphomas (Horwitz et al., 2016). It is more common in the elderly, with a median age of 70 years and is slightly more frequent in men than in women (Swerdlow et al., 2016). DLBCL is a neoplasm of medium or large B-cells with a diffuse growth pattern. Neoplastic cells typically express pan-B-cell markers such as CD19, CD20, CD22, CD79a, and PAXS, but may lack one or more of these. Surface and cytoplasmic immunoglobulin (most commonly lgM, followed by lgG and lgA) are also detected in 50-75% of the cases. In addition, myc protein expression is observed in a high proportion of DLBCL (30-50%) and is associated with concomitant bcl-2 expression in 20-35% of cases, which usually do not carry any chromosomal alteration in *MYC* or *BCL2* and are called "double-expressor (DE) lymphoma" (Karube & Campo, 2015). Several studies point to a worse outcome for DE DLBCL cases, suggesting that double expression of myc and bcl-2 proteins without genetic aberrations should be considered a prognostic indicator in DLBCL (Johnson et al., 2012; Molina et al., 2014).

DLBCL is characterized by an aggressive clinical course, and exhibits marked heterogeneity in clinical, morphologic, and molecular findings. In fact, gene expression profiling studies have identified two distinct molecular subgroups of DLBCL which differ in their cell of origin and present different outcomes and pathogenic mechanisms (Figure 13): GC B-cell (GCB) like (94% 5-year OS), and activated B-cell (ABC) like (68% 5-year OS). However, approximately 10-15% of cases cannot be included in either of these subtypes and remain unclassified (Alizadeh et al., 2000).
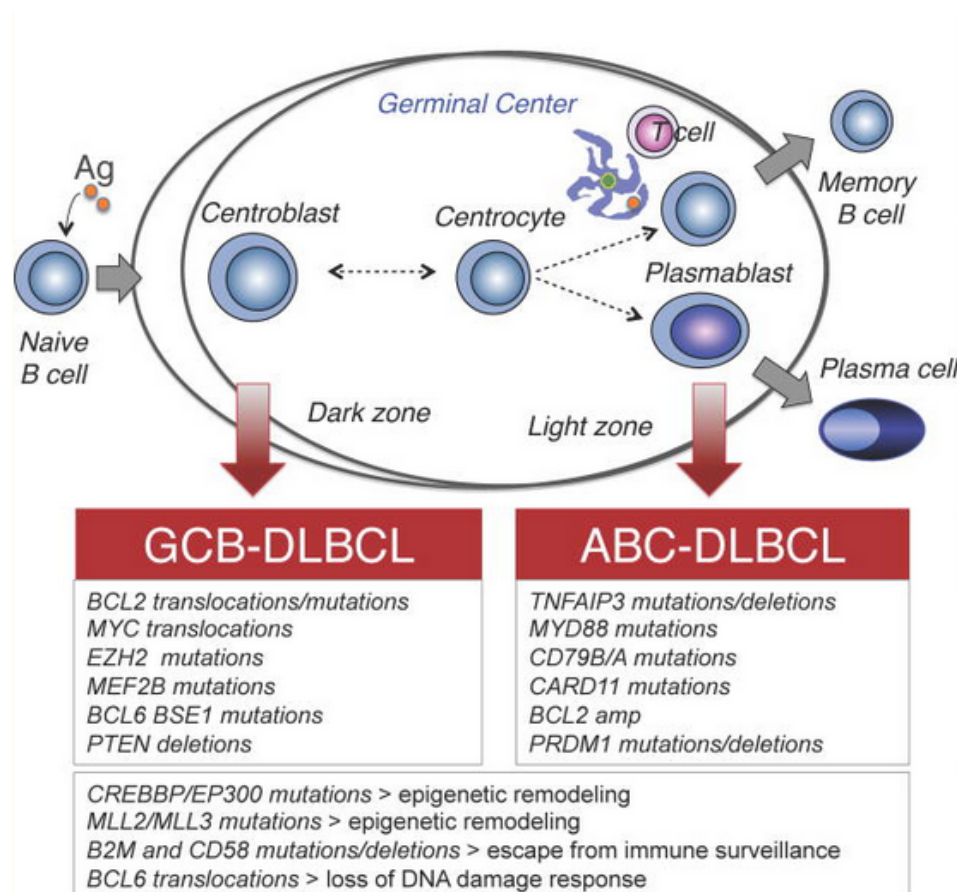
Figure 13. Postulated normal counterpart of major diffuse large B-cell lymphoma subtypes and the genetic lesions most commonly associated with each subtype. (Adapted from Pasqualucci et al., 2013).

Regarding genetic alterations, translocations involving *BCL2* or *MYC* are present in 20-30% and 5–10% of DLBCL, respectively, whereas *BCL6* translocations occur in about 30% of cases (Akyurek et al., 2012). Importantly, the co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double hit (DH) or triple hit (TH) lymphoma, identified a subset of highly aggressive B-cell lymphomas that can have DLBCL-like or Burkitt-like morphology (Rosenthal & Younes, 2017). The 2016 update of the WHO includes a provisional new category designated as high-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements for these neoplasms (Swerdlow et al., 2016).

The detection of translocations involving *MYC* is important for the diagnosis, but mainly for the prognosis of DLBCL patients. However, the wide variation in the

location of the breakpoint within the *MYC* region, which can extend up to one Mb on each side of *MYC*, makes these rearrangements difficult to detect (Figure 14) (Einerson et al., 2006; Bertrand et al., 2007). Regarding the translocation partners, although IG genes have been reported to be the most frequent, non-IG genes have also been described in 35% to 53% of *MYC* rearranged DLBCL cases (Karube & Campo, 2015).

MYC (8q24)
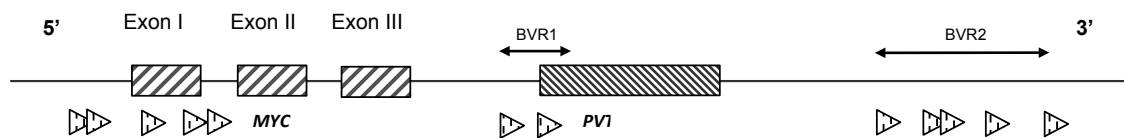
Figure 14. Schematic diagram of the breakpoints in the *MYC* gene for diffuse large B-cell lymphoma. Triangles below the gene represent the location of the breakpoints reported in the literature.

Translocations affecting the *BCL6* gene have been described to occur predominantly in the major breakpoint region (MBR), which encompasses the first noncoding exon and a portion of the first intron of *BCL6* (Yoshida et al., 1999; Ohno, 2006). However, an alternative breakpoint cluster (ABR) located 245–285 kb 5' *BCL6* has also been described (Iqbal et al., 2007) (Figure 15). With respect to translocation partners, *BCL6* can be rearranged with several different translocation partners. In fact, IG-associated BCL6 translocations occur only in 5-10% of DLBCL cases, with more than 20 distinct non-Ig translocation partners having been described to date (Jardin et al., 2007; Flodr et al., 2014; Jarosova et al., 2016).

BCL6 (3q27)
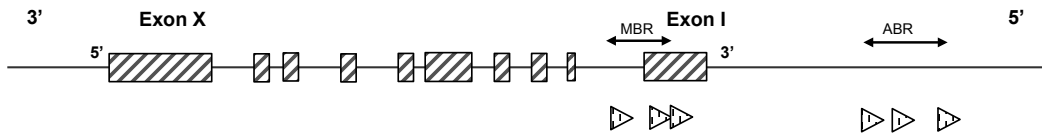
Figure 15. Schematic diagram of the breakpoints in the *BCL6* gene. Triangles below the gene represent the location of the breakpoints reported in the literature.

Recent NGS studies have identified common somatic mutations in all subgroups of DLBCL but also a profile of alterations differentially represented in both GCB and ABC subtypes (Figure 13) (Pasqualucci, 2013). Somatic mutations common in both DLBCL subtypes are inactivating mutations of *TP53* and genes involved in immunosurveillance (*B2M, CD58*), alterations in epigenetic regulators (*CREBBP/EP300*, *KMT2D/C (MLL2/3), MEF2B*) and oncogenic activation of *BCL6* (Reddy et al., 2017). *TP53* mutations have been identified in 20% of DLBCL cases of either GCB or ABC like type and are important because their presence is an independent predictor of poor survival in patients with DLBCL (Young et al., 2007; 2008). However, the domain location of *TP53* mutations plays a critical role in determining clinical outcome. Indeed, mutations in exons 5–9 affecting the DNA-binding domain resulting in a loss of function have been associated with a significant worse overall survival in DLBCL patients (Xu-Monette et al., 2012). Regarding the specific alterations of each subgroup, GCB-DLBCL carries frequent mutations in the cell motility regulator *GNA13*, deletions in the tumor suppressor *PTEN*, which is the major negative regulator of PI3K/AKT, and *BCL2* translocations (Iqbal et al., 2004; Healy et al., 2016). In addition, mutations affecting the histone methyltransferase *EZH2* Y641 residue are also specific for the GCB-DLBCL subtype, occurring in 30% of the cases, and can be targeted with EZH2 inhibitors (Morin et al., 2010). Otherwise, ABC-DLBCL harbors mutations in genes activating the BCR/TLR and NF-κβ pathways, such as *CD79A/B, MYD88, CARD11, PRDM1/BLIMP-1* and *TNFAIP3 (*Figure 16) (Pasqualucci, 2013)*.

45
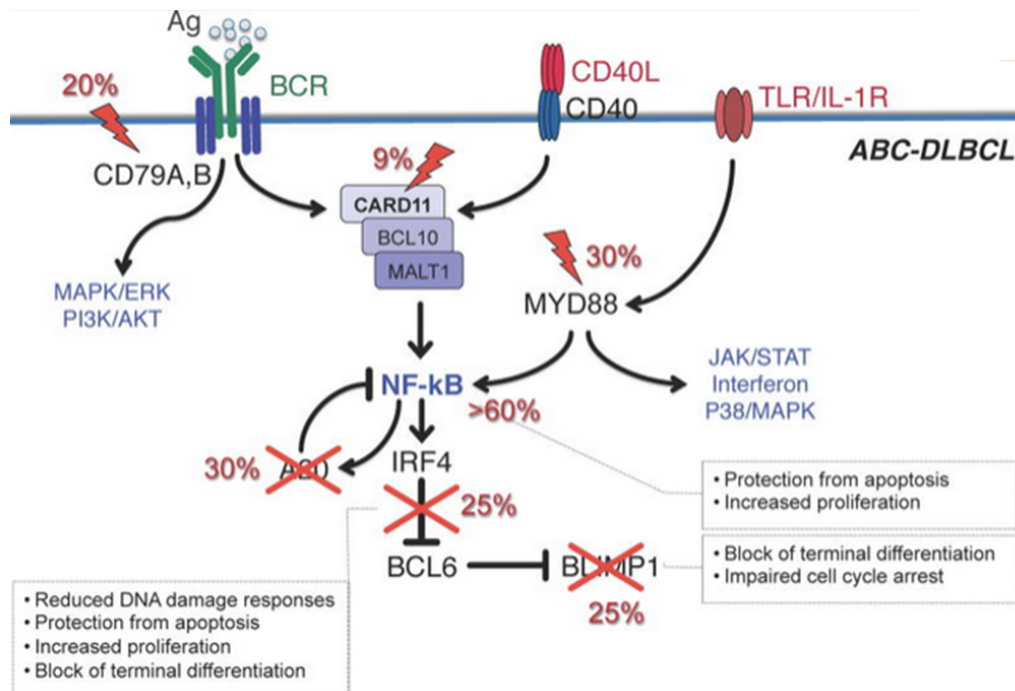
Figure 16. Genetic alterations and pathways affected in ABC-DLBCL. (Adapted from Pasqualucci, 2013)

ABC-DLBCL depend on a chronically active form of BCR signaling, which in >20% of cases is sustained by hotspot mutations of *CD79A* and *CD79B* genes*.* Importantly, knockdown of several BCR proximal and distal subunits is specifically toxic to ABC-DLBCL (Davis et al., 2010). Indeed, use of the BTK inhibitor ibrutinib is proving significantly effective in ABC-DLBCL patients carrying *CD79A/B* mutations (Yang et al., 2012). Besides, *MYD88* mutations occur in 30-40% of DLBCL, more frequently in ABC like subtype, and identify a group of patients with an adverse clinical presentation and poor outcome, thus representing a useful prognostic indicator in DLBCL patients (Rovira et al., 2016). Moreover, as somatically acquired *MYD88* mutations have been shown to promote NF-κB and JAK–STAT3 signaling in ABC like DLBCL, patients with ABC like DLBCL with *MYD88* L265P mutations may benefit from therapies targeting *IRAK4* alone or in combination with agents targeting the BCR, NF-κB, or JAK–STAT3 pathways (Lam et al., 2008; Davis et al., 2010; Milhollen et al., 2010).

The clinical significance of the main genetic biomarkers in diffuse large B-cell lymphoma is summarized in Table 9.

Table 9. Clinical significance of the main genetic biomarkers in diffuse large B-cell lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value	Predictive value
t(14;18)(q32;q21)	IGH-*BCL2* apoptosis regulator	20-30% of DLBCL or HGBL, with *MYC* and *BCL2* and/or *BCL6*	Diagnostic and prognostic marker	The co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double or triple hit lymphoma, identified a subset of highly aggressive B-cell lymphomas	
BCL6 rearrangement	*BCL6*-no partner identified	2% of DLBCL and in 30% of HGBL, with *MYC* and *BCL2* and/or *BCL6*	Diagnostic and prognostic marker	The co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double or triple hit lymphoma, identified a subset of highly aggressive B-cell lymphomas	
t(3;14)(q27;q32)	IGH-BCL6	10% of DLBCL or HGBL, with *MYC* and *BCL2* and/or *BCL6*	Diagnostic and prognostic marker	The co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double or triple hit lymphoma, identified a subset of highly aggressive B-cell lymphomas	
t(8;14)(q24;q32)	*IGH-MYC* cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis	5-10% of DLBCL or HGBL, with *MYC* and *BCL2* and/or *BCL6*	Diagnostic and prognostic marker	The co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double or triple hit lymphoma, identified a subset of highly aggressive B-cell lymphomas	
MYC rearrangement	*MYC*-no partner identified	35%-53% of MYC rearranged cases	Diagnostic and prognostic marker	The co-occurrence of *MYC* translocation with translocations involving *BCL2* or *BCL6* or both, so-called double or triple hit lymphoma, identified a subset of highly aggressive B-cell lymphomas	

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value	Predictive value
CD79A/B mutations	B-cell antigen receptor	3% and 20% of ABC-DLBCL, respectively	Associated with the activated B-cell phenotype		Targeted therapy with BTK inhibitor ibutrinib
MYD88 mutations	Involved in the Toll-like receptor and IL-1 receptor signaling pathway	30-40% in ABC-DLBCL and 10% of GCB-DLBCL	Associated with adverse clinical presentation in ABC-DLBCL: Presentation at older ages, more advanced stage, with frequent extranodal involvement, with higher International Prognostic Index scores	Associated with a shorter overall survival and worse outcome in ABC-DLBCL	
TP53 mutations	DNA repair, cell cycle control, apoptosis, senescence	21% of DLBCL	Younger age at diagnosis, high serum LDH, bulky tumors and high International Prognostic Index risk group	Shorter median 5-year survival in DLBCL patients Stratification of GCB-DLBCL in two different subgroups upon *TP53* mutational status, predicting for worse outcome in the *TP53* mutated subgroup	
EZH2 mutations	Chromatin conformation	*EZH2* Y641 residue in 30% of the GCB-DLBCL			Targeted therapy with *EZH2* inhibitor

Abbreviations: DLBCL: diffuse large B-cell lymphoma; ABC-DLBCL: activated B-cell like diffuse large B-cell lymphoma; GCB-DLBCL: germinal center B-cell like diffuse large B-cell lymphoma; HGBL: high grade B-cell lymphoma

(Adapted from Onaindia et al., 2017)

2.8. Burkitt lymphoma

Burkitt lymphoma (BL) is a highly aggressive MBCN that often presents in extranodal sites or as an acute leukemia. It comprises three epidemiological variants (endemic, sporadic, and immunodeficiency-associated BL) which mainly differ in their geographical distribution, clinical presentation, molecular genetics, biological features and subtle morphological aspects. Endemic BL occurs in equatorial Africa and in Papua New Guinea, being the most common childhood malignancy with an incidence peak among children aged 4-7 years and with a male-to-female ratio of 2:1. Besides, sporadic BL occurs worldwide, mainly in children and young adults (with the median age of the adult patients being 30 years), with a low incidence (1-2% of all lymphomas in western Europe and in the USA), and with a male-to-female ratio of 2:3. Finally, immunodeficiency-associated BL is more frequent in the setting of HIV infection than in other forms of immunosuppression, and it ususaly appears early in the progression of the disease (Swerdlow et al., 2016). Regarding morphology, it is composed of monomorphic medium-sized B-cells with basophilic cytoplasm and numerous mitotic figures, that typically express moderate to strong membrane IgM with light chain restriction, B-cell antigens (CD19, CD20, CD22, CD79a, and PAX5), GC markers (CD10 and bcl-6), strong myc expression (in almost all BL), and often also CD38, CD77, and CD43 (Naresh et al., 2011; Tapia et al., 2011). Otherwise, they are usually negative for CDS, CD23, CD138, bcl-2, and TdT. However, aberrant phenotypes, such as CDS expression, lack of CD10, or weak bcl-2 expression, have also been described in some cases (Masque-Soler et al, 2015).

Regarding genetic alterations, approximately 80% of patients with BL harbor the *MYC*-IGH t(8;14)(q24;q32) translocation, whereas the remaining cases have *MYC* translocated to either the IGK (at 2p11) or the IGL (at 22q11) *locus* (Janz, 2006). As a consequence of these translocations, *MYC* expression is strongly upregulated, which leads to a pronounced increase in cell proliferation, resulting in one of the most rapid tumor types. Heterogeneity in *MYC* translocations occurs not only because it can be rearranged with different *loci*, but also because various breakpoints may occur within these *loci*. Indeed, there are three main translocation breakpoints in *MYC*: The class I breakpoints, which are within the exon 1 and first intron of *MYC*;

the class II breakpoints, which are located at the 5' end of the *MYC*, and usually within a few kilobases of exon 1; and the class III breakpoints, which can be more than 100 kilobases away from *MYC* (Figure 17). Moreover, the clinical variants of BL demonstrate different preferential translocation breakpoints in both *MYC* and the partner immunoglobulin genes (Nguyen et al., 2017). Furthermore, EBV status is thought to affect the occurrence of genetic abnormalities in BL, also impacting on the generation of *MYC*-IG translocation. In fact, in the endemic cases, EBV-positive *MYC*-IG breakpoints originate from aberrant somatic hypermutation, while in the EBV-negative sporadic cases, *MYC* translocations mostly involve the Ig switch regions of the IGH locus (Bellan et al., 2005; Guikema et al., 2006). However, a small number of lymphomas have been recently described lacking *MYC* translocation and with 11q aberrations, but share the classic morphology, clinical presentation, and gene expression profile of BL (Salaverria et al., 2014). This group was recognized as a new entity distinct from BL (Burkitt-like lymphoma with 11q aberrations) by the 2016 WHO update (Swerdlow et al., 2016).

Apart from t(8;14)(q24;q32) translocation, the most common chromosomal abnormalities in sporadic Burkitt lymphoma include gains of 1q, 9q, 12q, 13q, 20q, 22q, and Xq and losses of 4q, 13q, and 17p (Salaverria et al., 2008; Richter et al., 2012). In addition, gains in 18q have been found to be associated with *ID3* mutation as 18q gains only occurred in samples with wild type (WT) *ID3* expression (Richter et al., 2012). Moreover, focal deletions of 19p13.3 have also been described, suggesting that loss of *TNFSF7* and *TNFSF9* in this region could contribute to pathogenesis in BL (Scholtysik et al., 2012).

MYC (8q24)
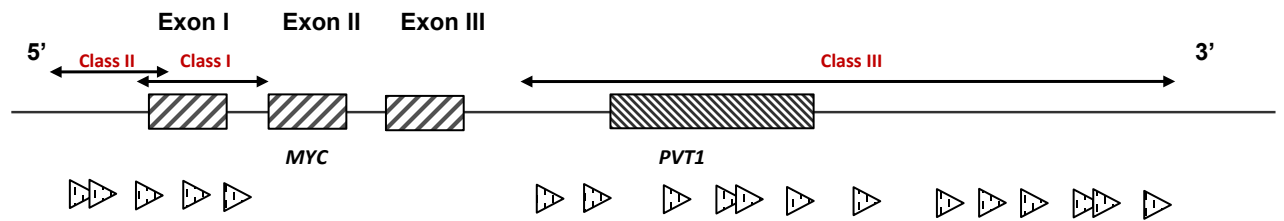
Figure 17. Schematic diagram of the breakpoints in the *MYC* gene for Burkitt lymphoma. Triangles below the gene represent the location of the breakpoints reported in the literature.

Besides, recent studies have demonstrated the existence of regulatory pathways that cooperate with *MYC* in the pathogenesis of BL (Figure 18), highlighting the importance of the BCR pathway in the pathogenic process (Sander et al., 2012). Mutations affecting the transcription factor *TCF3* or disrupting its negative regulator *ID3* (nearly all occurring in the helix–loop–helix functional domain of the protein) have been reported in 70% of sporadic and immunodeficiency related BL and in 40% of endemic cases. These alterations result in a constitutive activation of the BCR/PI3K signaling pathways, promoting survival and proliferation in lymphoid cells (Sander et al., 2012; Schmitz et al., 2012). Moreover, the silencing mutations in *ID3* are particularly interesting in light of previous studies that have identified *ID3* as a direct transcriptional target of *MYC*, suggesting a mutual regulatory relationship between them, which is disrupted in BL through upregulation of *MYC* and silencing of *ID3* (Basso et al., 2005; Seitz et al., 2011). In addition, NGS has identified a number of additional genetic mutations implicating other pathways involved in Burkitt lymphomagenesis (Figure 18). Comparition of the frequency of these mutations with those present in GCB and ABC-DLBCL revealed a stronger overlap between genes mutated in BL and GCB-DLBCL than ABC-DLCBL, further supporting the idea that BL originates from GC B-cells (Schmitz et al., 2012).
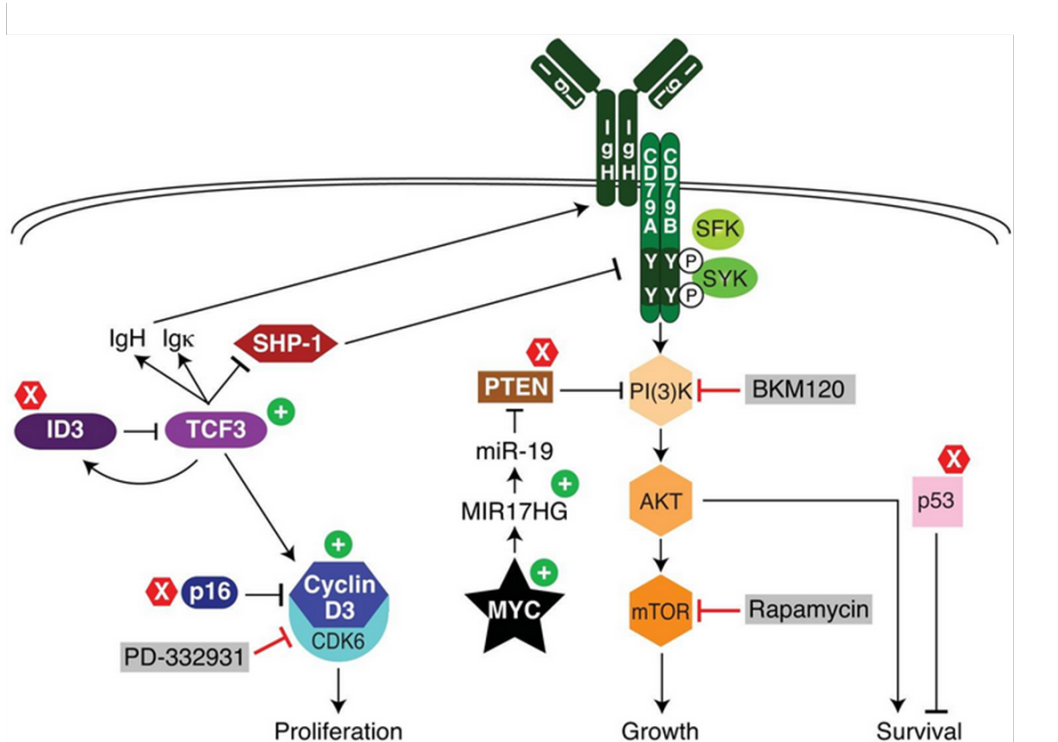
Figure 18. Recurrent oncogenic pathways in Burkitt lymphoma. (Adapted from Schmitz et al., 2012)

Clinical significance of the main genetic biomarkers in Burkitt lymphoma is shown in Table 10.

Table 10. Clinical significance of the main genetic biomarkers in Burkitt lymphoma.

Biomarker	Biologic function	Frequency	Clinical significance	Prognostic value	Predictive value
t(8;14)(q24;q32)	*IGH-MYC* cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis	90% of all BL cases	Diagnostic marker		
t(2;8)(p12;q24)	*IGK-MYC*	Rare	Diagnostic marker		
t(8;22)(q24;q22)	*IGL-MYC*	Rare	Diagnostic marker		
TCF3 mutations	Transcriptional regulator	70% of sporadic and immunodeficiency related BL and 40% of endemic cases		Promote survival and proliferation of lymphoid cells	
ID3 mutations	Transcriptional regulator	70% of sporadic and immunodeficiency related BL and 40% of endemic cases		Promote survival and proliferation of lymphoid cells	

Abbreviations: BL: Burkitt lymphoma

(Adapted from Onaindia et al., 2017)

3. Genetic techniques used in routine management of mature B-cell neoplasms

MBCN can often have overlapping clinical and morphological features, which could represent diagnostic challenges. However, making an accurate diagnosis is essential since the prognosis and treatment of these lymphomas can vary considerably. Several genomic studies have identified molecular biomarkers with a diagnostic, prognostic and predictive value. The integration of some of these molecular biomarkers has shown to be crucial in the diagnosis and management of B-cell neoplasms, as reflected in the WHO classification (Swerdlow et al., 2016) as well as in consensus patient management guidelines. For this reason, routine laboratories perform multiple genetic techniques in order to carry out an integrated diagnosis.

3.1. Chromosome banding analysis

Chromosome banding analysis (CBA) relies on harvesting cells to obtain mitosis, so that the chromosomes can be studied in their most condensed form. This is achieved by treating cells with tubulin inhibitors, such as colchicine or demecolcine (colcemid), which depolymerize the mitotic spindle and arrest cells in metaphase stage. Each chromosome shows a characteristic chromosomal band pattern when stained with a specific dye, with G-(Giemsa), R-(reverse), C-(centromere) and Q-(quinacrine) being the most common banding. G-banding is the most frequently used technique. This banding reflects differences in chromatin structure and base composition between different regions of the genome. Indeed, bands depicted in black are enriched in AT-regions while GC-rich regions are likely clustered in white bands (Korenberg et al., 1988; Bickmore, 2001). Bands are numbered consecutively away from the centromere on both the short (p) and long (q) chromosome arms, and the total number of bands or 'resolution' in the human karyotype depends on how condensed the chromosomes are, and at what stage of mitosis they are in (Francke, 1994; Bickmore et al., 2001). Karyotype definition is based on the number and chromosome band patterns observed in each methapase. Both normal and altered karyotypes are described following the international and standardized nomenclature

ISCN (International System for Human Cytogenetics Nomenclature) (McGowan-Jordan et al., 2016).

The main advantage of CBA is that it provides whole-genome analysis detecting both numerical and structural abnormalities and permits the identification of clonal evolution, as well as the presence of multiple independent clones (Rack et al., 2019). However, some logistic and technical problems hamper the use of this method. These include the need to obtain a relatively large volume of fresh tumor tissue that should be processed within a few hours in order to obtain tumor cells in metaphase after culture. In addition, although the cells are cultured in a favorable media using T or B lymphocyte mitogens like phytohemaglutinin (PHA) or TPA (12-O-Tetradecanoylphorbol-13-acetate) to improve the yield of metaphases, in several lymphomas, in particular indolent lymphomas, the tumor cells may fail to grow in culture and only normal karyotypes may be seen (Kluin & Schuuring, 2011). In this sense, the requirement for metaphases means that it is not applicable for all disease entities. In addition, since the detected abnormalities can be complex, the analysis should be carried out by trained professionals. Nonetheless, alterations could be incorrectly interpreted or missed due to the poor quality of the metaphases obtained in some cases. Furthermore, the resolution offered by banding techniques is low, approximately 5-10Mb, which means that small gains or deletions cannot be detected and some translocation are cryptic by CBA (Prakash et al., 2016).

Despite these limitations, CBA is still a useful method to detect specific primary and secondary abnormalities and to obtain a general assessment of genetic complexity in lymphomas ((Kluin & Schuuring, 2011).

3.2. Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) is a cytogenetic technique that identifies chromosomal abnormalities using molecular technology. It allows the detection and location of specific sequences of nucleic acids (DNA or RNA) on chromosomal preparations, cell extensions or fixed tissue sections (Prakash et al., 2016). FISH technique is based on the ability of a single-stranded DNA, called probe, previously

marked directly or indirectly with fluorescent molecules (FITC or Rhodamine, for instance), to anneal to complementary DNA or RNA. To achieve this, a process of denaturation of the DNA of both the sample and the probe is required. This process is based on the breakage of the hydrogen bridges that bind the double helix of the DNA by using high temperatures (70-80°) or by pH variations, resulting in single-stranded DNA. Incubation at 37° is then performed, which allows the hybridization of the sample DNA with the probe. This process is called renaturalization of the DNA. Both denaturation and renaturalization of the DNA are reversible processes, which occur physiologically in cells, and can be induced by variations in temperature. Finally, the evaluation of the results is carried out using a fluorescence microscope (Gozzetti & Le Beuu, 2000). Besides, when formalin-fixed paraffin-embedded (FFPE) tissue sections are used, a previous pretreatment is an essential step for effective FISH. This step includes a deparaffinization of the FFPE tissue sections, and a proteolytic digestion to permeabilize cell membranes, facilitating the penetration of fluorescently labeled probes. The general stages of the FISH are detailed in the Figure 19.
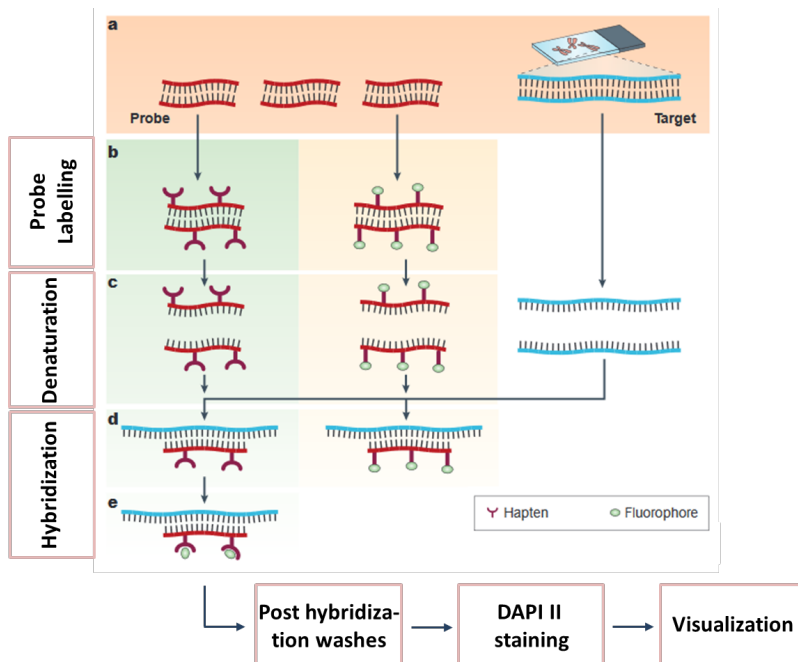
Figure 19. General stages of the FISH technique. (Adapted from Speicher & Carter, 2005)

Early in the 90's, FISH technique began to be used in the hematology laboratories applied to the diagnosis, forecast and follow-up of hematological neoplasms (Chen et al., 1992). The main advantage of FISH analysis is that it is suitable for neoplasms where metaphases are difficult to obtain (due to mitotic limitations or inability to obtain fresh samples) and when rapid diagnostic testing is required. In return, as it is a targeted analysis, it therefore needs to be combined with other testing to provide comprehensive information of chromosome abnormalities. In addition, despite presenting a higher resolution compared with the CBA, its resolution is limited to 100-150 kb, according to the design of the probe. Therefore, critical deletions and/or very small copy number changes should be tested by array or other molecular methods (Rack et al., 2019).

Nowadays it may be used as a sole or first line test for some neoplasms. In fact, FISH is the method of choice when no fresh sample is available, to detect submicroscopic (microdeletions) or cryptic abnormalities by CBA and to assess chromosomal rearrangements that could be missed by other molecular methods due to variable breakpoints (for example some *BCL2*-IGH translocations) or that involve multiple possible partner genes (*MYC* for instance). In addition, FISH is a useful complementary technique, particularly in combination with microarray data for the detection of balanced rearrangements, and to better clarify CBA results.

3.2.1. Probe types

There are a variety of probes used in the diagnostic routine, most of them commercially available. The main types of FISH probes are described below (Figure 20):

- Centromeric probes, which hybridize specifically with the repetitive sequences (α- and β-satellite sequences) of the centromeric region of each chromosome. These probes allow the detection of numerical chromosome alterations (aneuploidies).

- Chromosome painting probes use cloned DNA libraries derived from whole, flow-sorted human chromosomes. Hybridization results in the fluorescent staining or "painting" of the entire chromosome and is particularly useful in characterizing marker chromosomes or unclear structural abnormalities by CBA, but it need to be assessed in metaphase.
- *Locus*-specific probes, which are complementary to specific chromosomal regions or genes. Their sizes vary depending on the nature of the cloning vector. These probes are particularly useful for detecting copy number alterations or structural rearrangements involving a specific *locus* or gene.
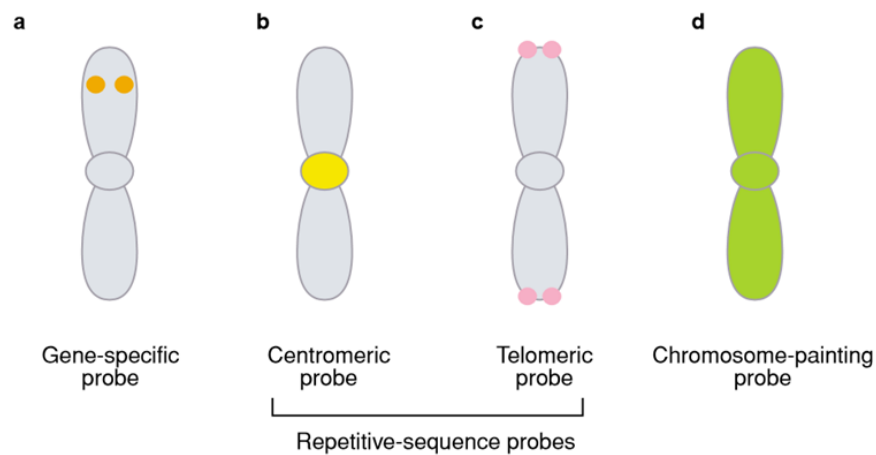
Figure 20. Example of the different types of FISH probes. (Adapted from Mcneil & Ried, 2000)

Within this group of *locus*-specific probes, there are two type of probes designed to detect chromosomal rearrangements (Figure 21):

1. **Dual-fusion probes,** which consist of two probes located in the two genes involved in a chromosomal translocation with different fluorochromes. The detection of two fusion signals allows the identification of a specific translocation. These probes provide information on the two chromosomal regions involved in the translocation, and allow the detection of variant FISH patterns, such as deletions associated with the translocation.

2. **Break-apart probes** are also used to detect chromosomal translocations but especially when the rearrangement could involve multiple partner genes, as is the case for *MYC* or *BCL6*. On this occasion, probes with different fluorochromes target on 3' and 5' of the breakpoint in the oncogene, and the display of the non-merged signals would be indicative of translocation. Hence, the main limitation of these probes is that they do not provide information about the rearrangement partner.
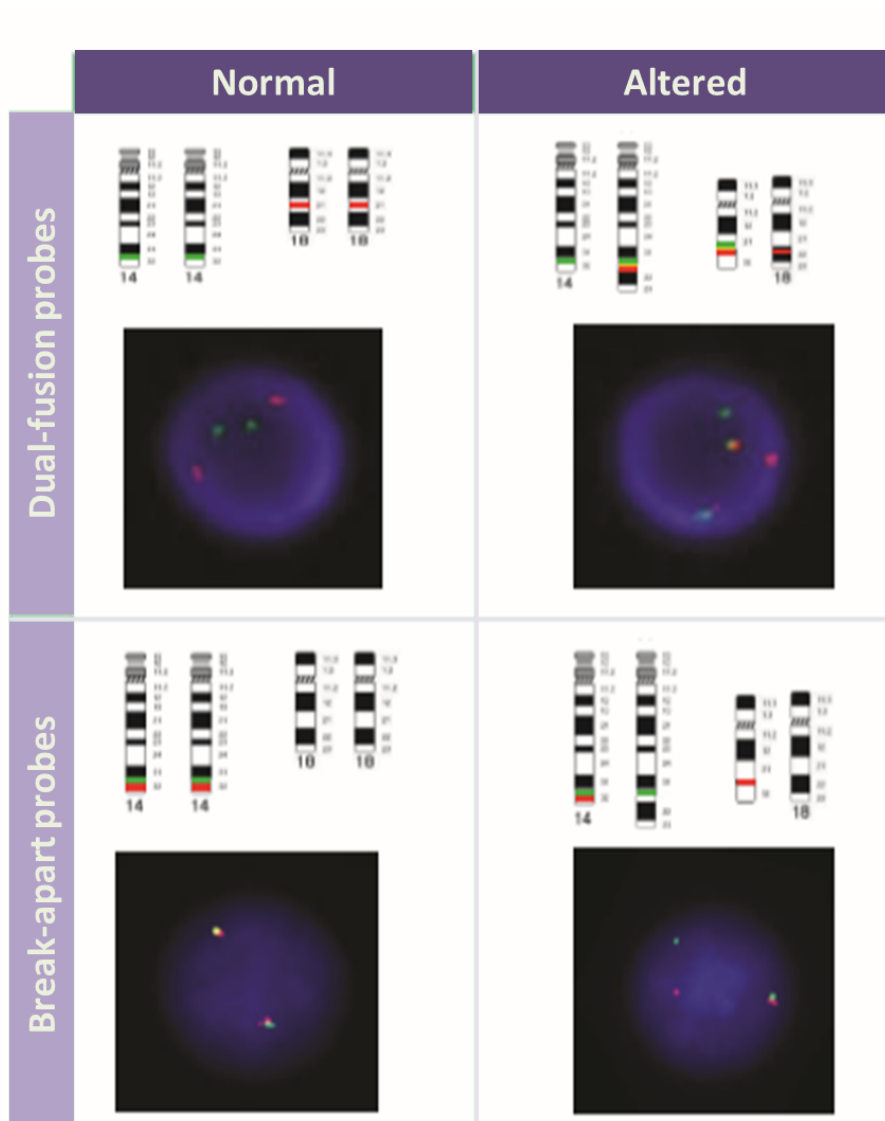
Figure 21. Example of the specific *locus* probe types for the detection of chromosome translocation. For each type of design, dual-fusion and break-apart, normal and altered FISH pattern are shown.

3.2.2. Non-commercial probes

Non-commercial probes can be made from complementary DNA (cDNA) (Von Deimling et al., 1999), plasmid, cosmid (Tsuchiya et al., 2002), P1 clone (Mark et al., 2005), fosmid (Birren et al., 1996), YAC or BAC (Liehr et al., 2002) molecules. The choice of the vector is made based on the length of the DNA to be cloned. However, BACs are the most widely used vector.

BACs are F-plasmids that contain an inserted sequence of the human genome to be cloned, which is approximately 300 kb in size, and an antibiotic resistance gene. BACs are then transfected into Escherichia coli bacteria, which are cultured in LB (Luria-Bertani) medium in the presence of an antibiotic to select those cells that contain the BAC. After that, bacterias are lysed and DNA sequence is joined to its fluorescent marker using the Nick Translation protocol (Rigby et al., 1977). Once the BAC probe has been labeled, it is hybridized on metaphases from cytogenetic cultures of healthy individuals to discard possible cohybridizations with other regions of the genome.

3.3. Genomic arrays

Genomic arrays, also termed 'molecular karyotyping' (Vermeesch et al., 2005) are used to detect changes in DNA copy number. Over the years it has been observed a change from the use of low resolution to higher resolution genomic microarrays for detection of both CNA and SNP variations. Depending on the hybridization system used, two types of genomic microarrays can be distinguished (Figure 22):

1. **The comparative genomic hybridization array (CGH array)** technique was described in the late 1990s and emerged as the evolution of the comparative genomic hybridization (CGH) technique, developed in the early 1990s. The CGH technique is based on the competitive hybridization of a tumor DNA against a control DNA, differentially labelled with two fluorophores, on normal metaphase preparations. CGH resolution ranges from 5-10 Mb, this being the main limitation of the technique. On the contrary, the hybridization of CGH arrays is carried out on a solid support containing thousands of probes and the resolution is defined by the type of probe used (Figure 22). While BAC arrays with relatively low

resolution were initially applied, a shift towards the use of synthetic oligonucleotides ranging from 25 to 75 bases in length over recent years has allowed a dramatic increase in resolution. In general, larger targets such as BACs provide better signal-noise ratios than with smaller oligonucleotide targets. However, the increased genomic coverage and short target length of oligo arrays facilitate more accurate refinement of breakpoints and CNA detection, which may be less than 1 Mb in size. It is noteworthy that some of the current CGH array platforms available combine oligonucleotide probes to detect CNA but also some probes targeting SNPs.

2. **The SNP array** is composed by a combination of oligonucleotide and SNP probes in the solid support. The main difference compared with CGH array is that a single hybridization is performed for the patient's DNA and the signal intensities are then compared with a reference dataset (Figure 22).
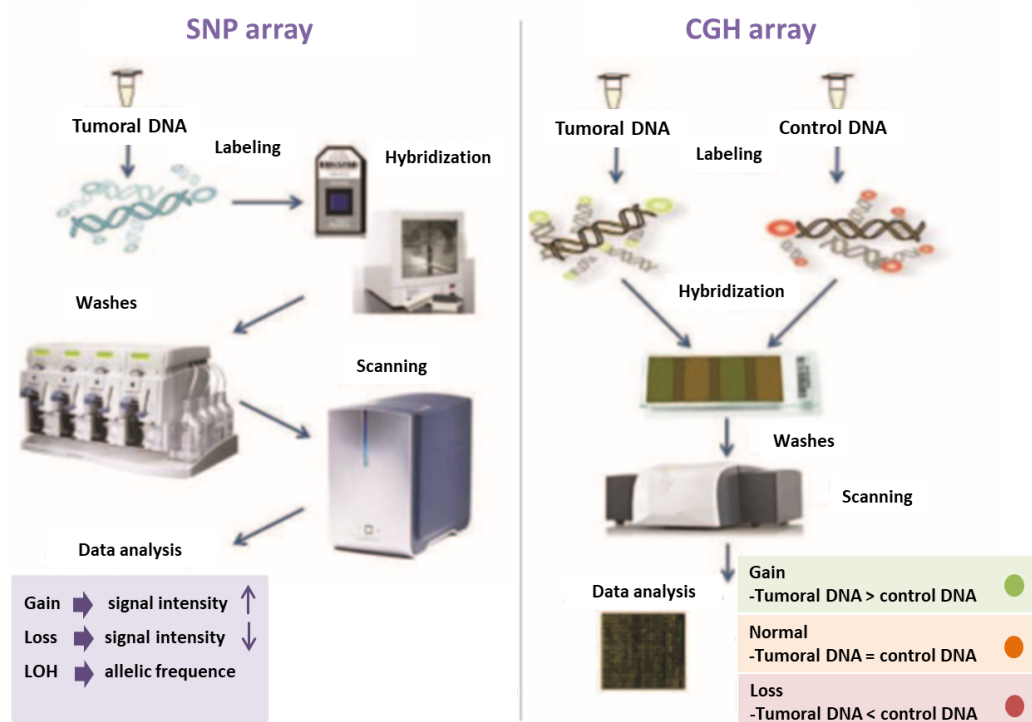
Figure 22. Scheme of the basic process carried out in the two types of array hybridization.

Array analysis allows the accurate detection of CNA along the genome and the identification of genomic instability, including complex genomic alterations such as chromothripsis. In addition, the inclusion of SNP probes in the microarrays enable the detection of polyploidy and acquired copy neutral loss of heterozygosity (aCN-LOH) regions, which often mask point mutations in tumour suppressor genes.

In addition, as no cell culture is required, it is a valuable technique for pathologies where CBA studies could not be performed or show poor results. Furthermore, it is particularly useful for neoplasms where multiple CNA are tested, replacing multiple FISH analyses, and for the detection of very small CNA. However, it is important to note that limitations of microarrays, inherent to the technique, are that they show a limited sensitivity (around 20-30%) and cannot detect individual clones and balanced chromosomal rearrangements (balanced translocations and inversions) (Rack et al., 2019).

3.4. Polymerase chain reaction

The polymerase chain reaction (PCR) was invented in 1985 by Kary B. Mullis (Mullis, 1990) and is a simple enzymatic assay, which allows the amplification of a specific DNA fragment from a complex pool of DNA (Garibyan & Avashia, 2013).

Each PCR assay requires the presence of template DNA, primers, nucleotides, and DNA polymerase. The DNA polymerase is the key enzyme that links individual nucleotides, adenine, thymine, cytosine, and guanine (A, T, C, G), together to form the PCR product. The primers are short DNA fragments with a defined sequence complementary to the extremes of target DNA that is to be detected and amplified, and therefore, they are those that specify the exact DNA product to be amplified (Garibyan & Avashia, 2013).

The reaction consist of three steps (denaturation, annealing, and extension) that are repeated 30-40 cycles, resulting in the exponential accumulation of the specific target fragment, approximately 2^n, where n is the number of cycles (Mullis, 1990). First, the solution is heated above the melting point of the two complementary DNA strands of the target DNA, which allows the strands to separate. This process is

called denaturation. Next, the temperature is lowered to allow the specific primers to bind to the target DNA segments, a process known as hybridization or annealing. The annealing will occur only if they are complementary in sequence. The temperature is then raised again, at which time the DNA polymerase is able to extend the primers by adding nucleotides to the developing DNA strand. Finally, the PCR product is analyzed. There are two main methods of visualizing the PCR products: (1) staining of the amplified DNA product with a chemical dye such as ethidium bromide, which intercalates between the two strands of the helix or (2) labeling the PCR primers or nucleotides with fluorescent dyes (fluorophores) prior to PCR amplification. Besides, the most widely used method for analyzing the PCR product is the use of agarose gel electrophoresis, which separates DNA products on the basis of size and charge. Therefore, in addition to determining the presence of PCR product, it allows defining its size using standardized molecular markers (**Figure 23**).

PCR can be carried out using DNA from a variety of tissues and organisms, but can also be performed using RNA samples, in which case a reverse transcription must be performed before the PCR to obtain cDNA (RT-PCR) (Bachman, 2013).

A variant of PCR is quantitative real-time PCR or qRT-PCR, which provides information beyond mere DNA detection. Unlike conventional PCR, qRT-PCR enables both detection and quantification of the PCR product in real-time, while it is being synthesized (VanGuilder et al., 2008). There are two common methods used to detect and quantify the product, which include (1) fluorescent dyes that non-specifically intercalate with double-stranded DNA and (2) sequence-specific DNA probes consisting of fluorescently labeled reports. The latter allow detection only after hybridization of the probe to its complementary DNA target. In both cases, the quantification of the target DNA is possible since the fluorescence signal increases in direct proportion to the amount of PCR product (Figure 23). Additionally, real-time PCR can also be combined with reverse transcription, allowing RNA samples to be used and quantified.

PCR is a simple and rapid technique with the potential to produce millions to billions of copies of a specific product for detection, sequencing, or cloning. Therefore, it is a

very sensitive technique, since only small amounts of DNA are needed for the PCR to generate enough copies to be analyzed. However, and due to its high sensitivity, any form of contamination of the sample, even by small amounts of DNA, can produce misleading results (Smith & Osborn, 2009).

Besides, the design of the primers requires prior knowledge of the sequence to be amplified. Furthermore, as has been observed in rearrangements, if deletions or variations in the sequence occur, the primer cannot bind (Espinet et al., 2004). Therefore, PCR can only be used to identify the presence or absence of a pathogen or gene with a well-known sequence (Schoenbrunner et al., 2017). Besides, qRT-PCR has the advantage of quantification of the synthesized product. Therefore, it can be used to analyze alterations of gene expression levels in tumors, for example (Iacoangeli et al., 2018).
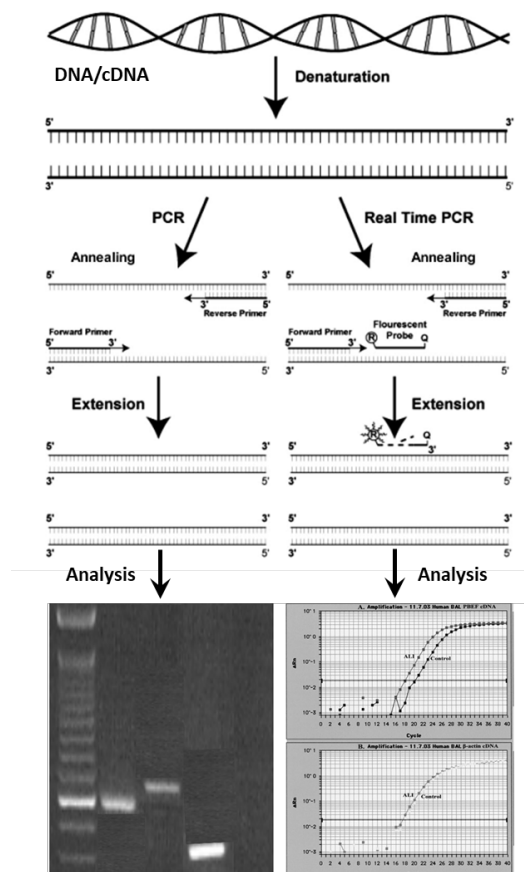
Figure 23. Scheme of the basic process carried out in PCR and real-time PCR. (Adapted from Garcia and Ma, 2005)

3.5. Sanger sequencing

Sanger sequencing, also known as the 'chain termination method', was developed by Frederick Sanger and colleagues in 1977 (Sanger et al., 1977), and allows to determine the nucleotide sequence of DNA.

It is a targeted sequencing technique that uses oligonucleotide primers to search for specific DNA regions. It begins with denaturation of the double-stranded DNA. The primers are then annealed to the single-stranded DNA and elongated the chain using a mixture of deoxynucleotide triphosphates (dNTPs) and chain-terminating dideoxynucleotide triphosphates (ddNTPs) tagged with a fluorescent dye specific for each nucleotide. Since ddNTPs do not contain a 3'-hydroxyl group, the chain cannot be further extended when a ddNTP is added. Moreover, dNTPs and ddNTPs have an equal chance of attaching to the sequence, therefore each DNA fragment will terminate at varying length, resulting in a mixture of DNA fragments of different size all having the same 5'end but with different ddNTP residue at 3' end (Sanger et al., 1977). The extension products are then separated by capillary electrophoresis (CE). Since the rate at which a DNA fragment migrates through the medium is inversely proportional to its molecular weight, CE can separate different DNA fragments by size at one base resolution. In addition, a laser excites the dye-labeled DNA fragments so that the included ddNTP can be idendified as each nucleotide emits light at a characteristic wave length. The software can then interpret the detected signals and translate them into a 'peak' sequence (Figure 24) (Karger & Guttman, 2009).

Sanger sequencing is a robust testing strategy able to determine whether a point mutation or small deletion or duplication is present and it has been widely used for several decades in many settings. However, even though individual Sanger sequencing reactions can be performed to cover any desired region, this approach can be costly when compared with other multiplex testing systems. Therefore, most currently available Sanger sequencing tests are gene-specific or analyze a small subset of genes. In addition, Sanger sequencing is able to identify mosaic mutations including as low as 20% of the cells, but is not precisely quantifiable and additional testing strategies must be used for quantification.
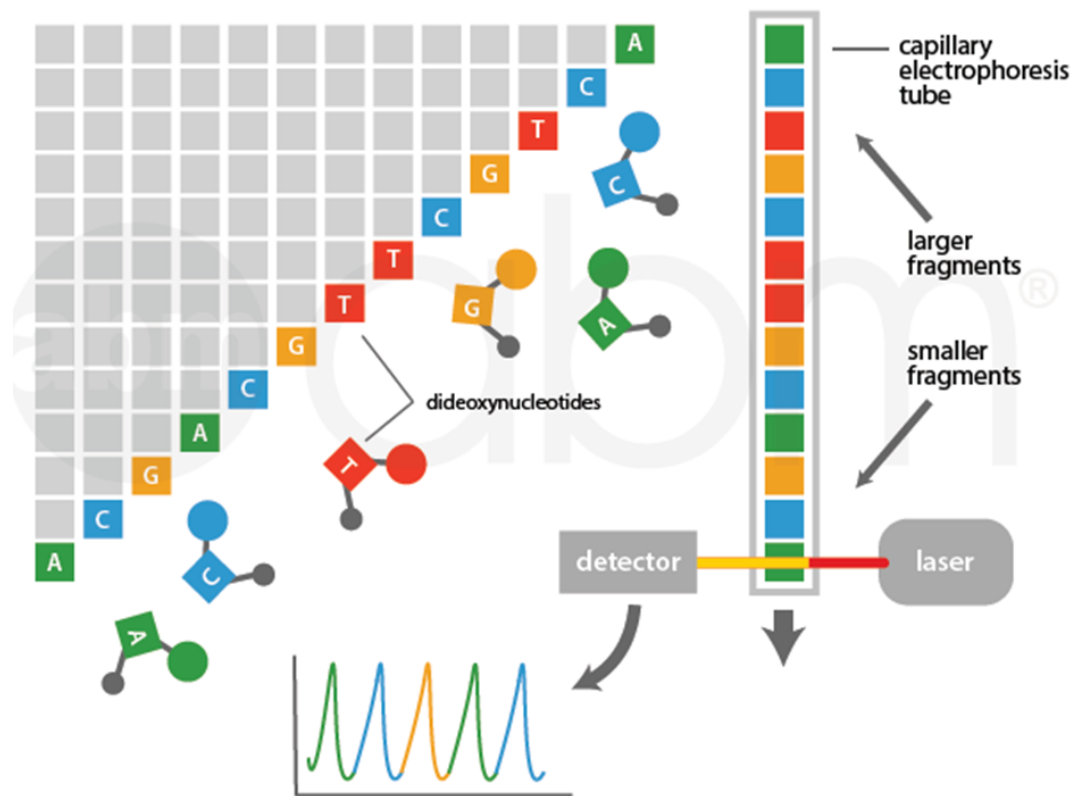
Figure 24. Scheme of Sanger sequencing.

The main carachteristics of the classical genetic thechniques applied in the management of MBCN in the clinical setting are summarizes in Table 11.

Table 11. Comparison of the techniques used in the management of mature B-cell neoplasms.

Technique	Resolution	Sensibility	Need to have dividing cells	Distinction between individual clones	Genome overview	Detection of balanced alterations	LOH detection
Chromosome banding analysis	Low	10%	Yes	Yes	Yes	Yes	No
Fluorescent in situ hybridization	Low	High	No	No	No	Yes	No
CGH arrays	High	20-30%	No	No	Yes	No	No
SNP arrays	High	20-30%	No	No	Yes	No	Yes
PCR	High	High	No	No	No	Yes	Yes
Sanger Sequencing	High	20-30%	No	No	No	Yes	Yes

4. Next-generation sequencing

4.1. Next-generation sequencing applications

For decades, the methodology described by Frederick Sanger has been the gold standard for DNA sequencing (Sanger et al., 1977). Since the completion of the first sequenced human genome (International Human Genome Sequencing Consortium, 2004), there has been a revolution in sequencing techniques. Post-Sanger technologies, which are collectively referred to as next-generation sequencing (NGS), provide a significant increase in sequencing throughput due to parallelization, automation and computerization of the sequencing method (Mardis, 2011; Goodwin et al., 2016). Since 2004, when the first commercial NGS sequencer was available, several NGS sequencers based on different chemistries have been launched into the market (Figure 25) and the costs associated with NGS have decreased continuously, with a notable decline during the last years (Goodwin et al., 2016).

Since the introduction of NGS technology, the production of large number of low-cost reads made the NGS platforms useful for many applications in clinical diagnosis. These include sequencing the entire genome, exons of every protein-coding gene or targeting specific genomic regions (De Koning et al., 2015). **Whole-genome sequencing (WGS)** can be used to determine the complete sequence of an entire DNA sample, including all the coding and non-coding sequences. The major advantage of this approach is that the complete genomic information will be available and the sequencing time is relatively fast. In addition, DNA for WGS is prepared using PCR-free protocols. The purified genomic DNA is fragmented and ligated with indexed adapters with no amplification required. Therefore, this method reduces the bias and gaps associated with PCR preparations, which in turn produces largely uniform depth of coverage of the entire genome. As a result, an average read depth of only ~35X is required to ensure that there are a sufficient number of data points to make a high confidence nucleotide call at nearly all sites (Ajay et al., 2011; Sims et al., 2014). Specifically, in the study of hematological malignancies, WGS has provided a comprehensive view of the neoplastic genome (Morin et al., 2013; Puente et al., 2015). The goal of such analysis is typically to identify somatic variants by comparing the sequence of the neoplastic population with their germline

counterpart. Although this strategy may be attractive, significant computational difficulties are associated with the underlying analysis, mainly due to the very large amount of data that is generated and has to be analyzed. An additional challenge of working with genome scale data is that certain relatively large portions of the genome are still difficult to examine accurately because they are highly repetitive, such as sequences in centromeric and telomeric regions, in addition to other highly repetitive sequences that are scattered throughout the genome (Treangen & Salzberg, 2012). Moreover, although ~35X coverage is adequate to produce genotype calls with acceptably low error rates, reliable identification of variants in highly aneuploid genomes or heterogeneous cell populations, such as those of tumors that may also have subclonal variants and alterations of very low frequency, requires greater depth of coverage (Ajay et al., 2011; Sims et al., 2014). Besides, despite the decreasing costs of NGS technologies, WGS is still too expensive for the clinical practice.

A more specialized approach is to focus only on specific regions of the genome using an approach known as 'targeted sequencing', which can include from a single selected genic region up to the entire set of protein-coding regions, also known as **whole exome sequencing (WES)**. WES is a high throughput genomic technology that focuses on protein-coding regions, which are the most functionally relevant parts of the genome, selectively captured by target enrichment strategies (Hodges et al., 2007). Since the exome represents approximately 1-2% of the genome, WES technology provides higher depth of coverage at a lower cost and in a shorter time than WGS, resulting in a more accessible strategy suitable for molecular diagnosis and for discovering novel disease related genes (Quesada et al., 2012). However, the depth of coverage along exons is less uniformly distributed, contributing to many low coverage regions, which hinders accurate variant calling (Wang et al., 2017). This uneven depth of coverage is mainly associated with the target enrichment strategy, which is achieved by oligonucleotide probes that selectively hybridize, capture and amplify (by PCR) the entire coding region of the genome (Hodges et al., 2007). Capture probes and PCR primers anneal more easily to some regions of the genome than others, therefore, some regions will become amplified many times while others

will be infrequently amplified. Aditionally, WES data has more sparse regions compared with WGS, contributing to more false negative variant calls (Wang et al., 2017). Therefore, to overcome these limitations, samples are generally sequenced at an average read depth of ~100X, which is the depth required to minimize the number of sites with insufficient data points to make a highly trusted nucleotide call. Besides, since this strategy began to be applied to the study of the cancer genome, most large-scale projects, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), have focused exclusively on genomic variation in the coding sequences to characterize the cancer genome (Hudson et al., 2010; Chang, 2013). However, mutations in the non-coding regions of the genome have been identified, some of which with clinical implications, and yet can not be detected by WES approach (Weinhold et al., 2014; Puente et al., 2015; Schmitt & Chang, 2016).

On the other hand, **custom targeted sequencing** has several advantages in clinical diagnosis compared with WES and WGS. This strategy can not only analyze a panel of several selected genomic regions simultaneously, it can also analyze a large number of samples in a single run and in a relatively short processing time, ultimately reducing the cost of sequencing. Since this approach is focused on 'limited' genomic regions, the sequencing read depth obtained can be much higher than in WGS or WES, which also overcomes the limitations derived from the target enrichment strategy. In addition, the analysis of the generated data and the interpretation of variants are less challenging, making the response time significantly shorter (Sikkema-Raddatz et al., 2013). Altogether, targeted NGS panels have emerged as the most widely used NGS methodology for the onco-hematological clinical practice. Several commercial and custom panels have been developed and efforts have emerged to standardize and homogenize their aplication on this field (Li et al., 2017; Palomo et al., 2020). Although in recent years various custom and commercial panels have been developed to study MBCN, these panels have been mostly used at the research level, since in most of the cases the abnormalities assessed are not mandatory in the clinical setting (Sutton et al., 2015; Rodríguez-Vicente et al., 2017). However, a NGS panel that could comprise the distinct genetic alterations present in

the different MBCN could be of great interest in the clinical practice (Sujobert et al., 2019).
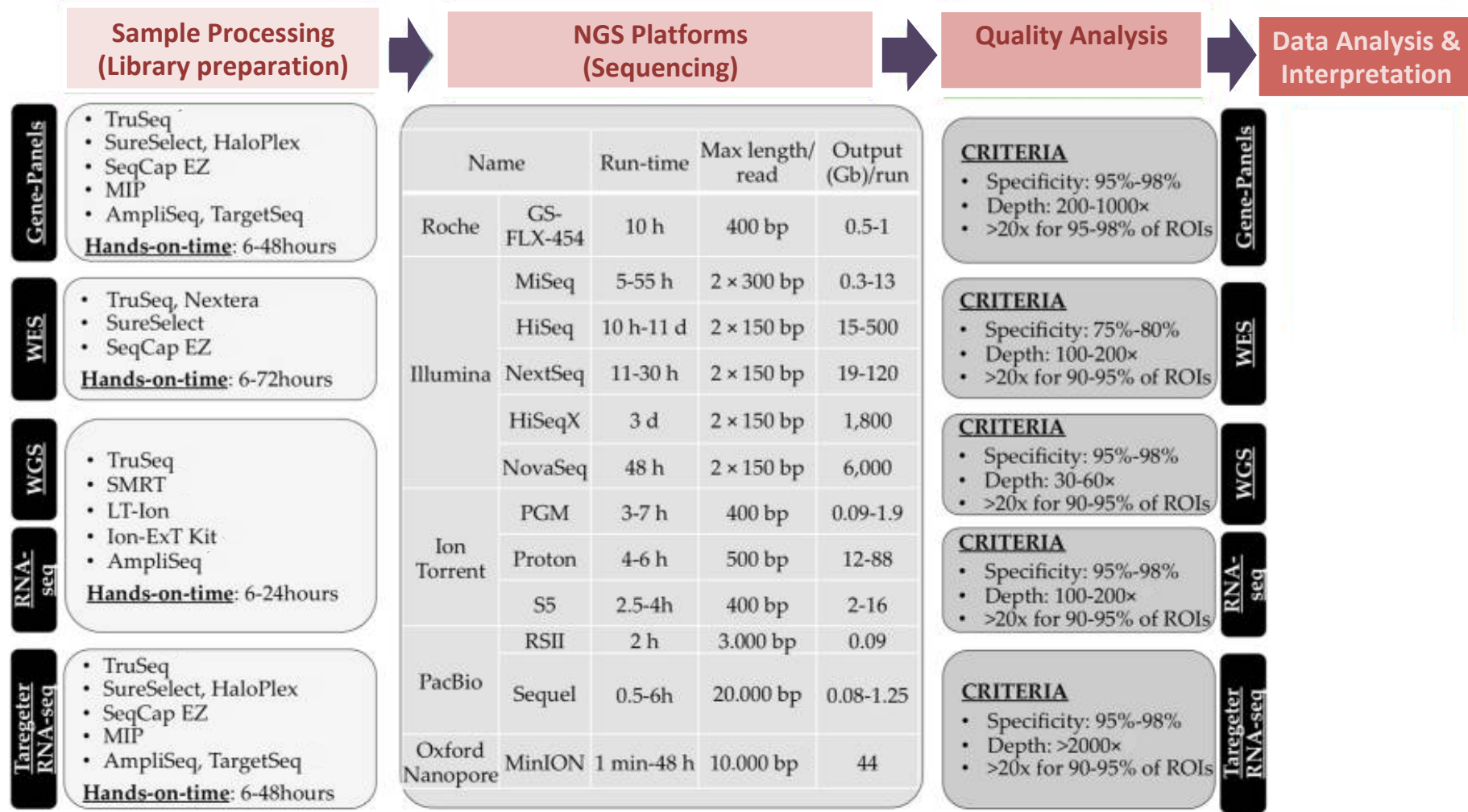
Figure 25. Pipeline illustrating the different commercial NGS sequencers available and their main characteristics. (Adapted from Kamps et al., 2017)

4.2. NGS workflow

Although multiple sequencing chemistries and platforms are currently available, and more are being developed, they generally share a common workflow. Herein, the main steps of targeted sequencing from DNA samples are detailed.

4.2.1. Library preparation

Library preparation refers to the process of preparing DNA for its use on a sequencer. Although there are many methods available, they all present the same core steps: obtaining target DNA molecules of the desired size and assembly of the oligonucleotide adapters to the ends of the target fragments (Head et al., 2014). Indeed, the term library refers to these fragments of DNA with flanking adaptors that are ready for sequencing. Two main strategies have been developed for library preparation (Figure 26):

A) Hibridization-capture based strategies

This strategy includes an initial random fragmentation of the DNA molecules into varying sizes.The size of the target DNA fragments in the final library is a key parameter for NGS library construction. Three approaches are available to fragment nucleic acid chains: physical, enzymatic, and chemical. Nevertheless, physical (acoustic shearing, sonication and hydrodynamic fragmentation) or enzymatic methods (that include digestion by DNase I or Fragmentase) are the most used. Once the starting DNA is fragmented, the ends are blunted and 5' phosphorylated. In addition, the 3' ends are A-tailed to facilitate ligation to sequencing adapters. DNA fragments are then tagged with adaptors that may include molecular barcodes (to allow pooling of patient samples), universal PCR primers, hybridization sequences to bind the DNA fragment to a surface, and recognition sites to initiate sequencing.

An alternative method is the Nextera DNA Sample Prep Kit (Illumina), which prepares genomic DNA libraries by using a transposase enzyme to simultaneously fragment and tag DNA in a single-tube reaction termed "tagmentation" (Adey et al., 2010). The engineered enzyme has dual activity; it fragments the DNA and simultaneously adds specific adapters to both ends of the fragments, which will be used to amplify the insert DNA by PCR. This method brings several advantages, including reduced sample handling and preparation time.

The resulting library undergoes a target enrichment process to capture specific regions of interest to be sequenced, which is based on hybridization to complementary sequences. This hybridization-based strategy employ a genomic DNA denaturation and annealing of different length DNA or RNA single-stranded oligonucleotides (referred to as 'probes' or 'baits') to the region of interest. The probes are generally biotinylated, which allows them to bind to streptavidin-coated magnetic beads and thus capture sheared, complementary DNA in the bead complex.

B) Amplicon-based strategies

PCR-based strategies rely on primer driven amplification of the region of interest and are generally combined with the library preparation step, as the primers that select the regions of interest may also contain the adaptor sequences (Kozarewa et al., 2015).
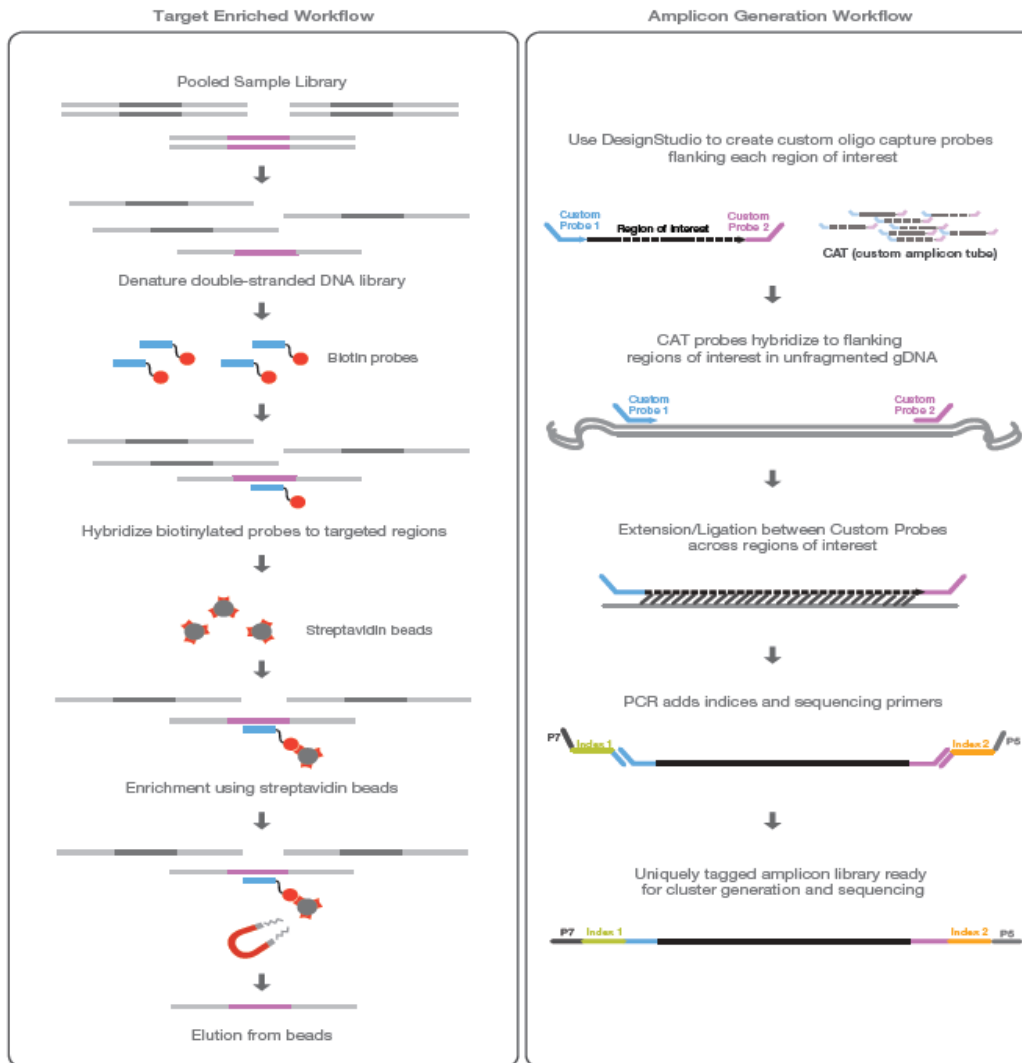
Figure 26. Comparison of target enrichment workflow for amplicon and capture hybridization NGS assays.

Optimizing the enrichment procedure in a diagnostic panel is of major importance. The choice of enrichment strategy is often dictated by the clinical use and the type of alterations that are going to be assessed. While sequence capture is preferred for large genomic regions, PCR is often chosen for smaller regions where greater enrichment is desired (Samorodnitsky et al., 2015). Table 12 summarizes the main differences between both strategies.

Table 12. Comparison between the two main NGS strategies.

	Amplicon-based NGS panels	Hibridization-capture based NGS panels
Basis	DNA is enriched in regions of interest by amplification with region-specific primers	The DNA is enriched in the regions of interest by hybridization with complementary probes to each region and subsequent amplification
Recomended for	Small anels: <500Kb (2-650Kb)	Medium-large panels: 500Kb-15Mb
Type of coverage achieved	Staggered (useful for hot spots)	Bell
Requirements	Sequences must be known for the design of the specific primers: useful for well-known fusion genes	It is not necessary to know the entire sequence of the region to be covered: useful for rearrangements with unknown translocation partners and complex rearrangements with gain and loss of material

(Adapted from Metzker, 2010; Samorodnitsky et al., 2015)

4.2.2. Sequencing

After library preparation and multiplexing of the samples that need to be sequenced, sequencing step is run into a sequencer. Different technical strategies are used among the commercial sequencers available, nonetheless in all of them sequencing include a clonal amplification and a read sequencing approach.

In **clonal amplification** step, each DNA fragment is immobilized and clonally amplified. The two most common methods used are the following:

- Emulsion PCR (emPCR), in which bead emulsion is used to immobilize and clonally amplify the DNA. Beads contain sequences that hybridize to part of the adaptor on the DNA fragments. After adaptor ligation, DNA is separated into single strands and captured onto beads under conditions that favur one DNA molecule per bead. A reaction mixture consisting of an oil–aqueous emulsion is created to encapsulate bead–DNA complexes into single aqueous droplets. PCR amplification is performed within these droplets to create beads containing several thousand copies of the same template sequence. Then, beads are immobilized in a support to continue with sequencing (Mardis, 2013).

- In solid-phase amplification the flow cell contains forward and reverse primers that are complementary to the adaptors on the DNA fragments. After single-stranded DNA (ssDNA) templates are annealed to the anchor oligonucleotides on the surface of the flow cell, DNA molecules are clonally amplified in a modified isothermal PCR reaction termed "bridge PCR", in which the DNA molecules form a "bridge" with an adjacent anchor oligonucleotide. This process results in the generation of more than 100 million spatially separated template clusters, containing over 1000 copies of clonally amplified DNA molecules on the surface of the flow cell. Clusters are then denatured to provide a single-stranded template, and a sequencing primer oligonucleotide is hybridized to the strand initiating the sequencing reaction (Adessi et al., 2000; Fedurco et al., 2006).

With regard to **short-read sequencing** approaches, three different strategies are the most widely used:

- Sequencing by ligation (SBL): involves the hybridization and ligation of labelled probe and anchor sequences to a DNA strand (Drmanac et al., 2010). The probes encode one or two known bases (one-base-encoded probes or two-base-encoded probes) and a series of degenerate or universal bases, driving complementary binding between the probe and template, whereas the anchor fragment encodes a known sequence that is complementary to an adapter sequence and provides a site to initiate ligation. After ligation, the template is imaged and the known base or bases in the probe are identified (Landegren et al., 1988). A new cycle begins after complete removal of the anchor–probe complex or through cleavage to remove the fluorophore and to regenerate the ligation site. This strategy is used by SOLiD and Complete Genomics sequencers.

- Sequencing by synthesis (SBS): is a term used to describe numerous DNA-polymerase-dependent methods in the literature, and there are different mechanisms involved in this approach: cyclic reversible termination (CRT) and single-nucleotide addition (SNA) (Goodwin et al., 2016).

  CRT (Illumina, Qiagen) approaches are defined by their use of terminator molecules that are similar to those used in Sanger sequencing, in which the ribose 3'-OH group is blocked, thus preventing elongation (Guo et al., 2008). During each cycle, a mixture of all four individually labelled and 3'-blocked deoxynucleotides (dNTPs) are added. After the incorporation of a single dNTP to each elongating complementary strand, unbound dNTPs are removed and the surface is imaged to identify which dNTP was incorporated at each cluster. The fluorophore and blocking group are then removed and a new cycle is begun.

  Unlike CRT, **SNA** approaches (454, Ion Torrent) rely on a single signal to identify the incorporation of a dNTP into an elongating strand. For that reason, each of the four nucleotides must be added iteratively to a sequencing reaction to ensure that only one dNTP is responsible for the

signal. The exception to this is homopolymer regions where identical dNTPs are added, with sequence identification relying on a proportional increase in the signal as multiple dNTPs are incorporated.

Despite all the different technologies available on the market, most clinical sequencing is performed on Illumina sequencers (San Diego, California), including the Novaseq, Hiseq, NexSeq, Miseq, Miniseq, iSeq or Ion Torrent sequencers (Thermo Fisher Scientific, Gloucester, UK), which include the IonPGM, IonProton, and IonS5 (Yohe & Thyagarajan, 2017). Figure 27 sumarizes the main steps of targeted sequencing from DNA samples of the different NGS platforms.
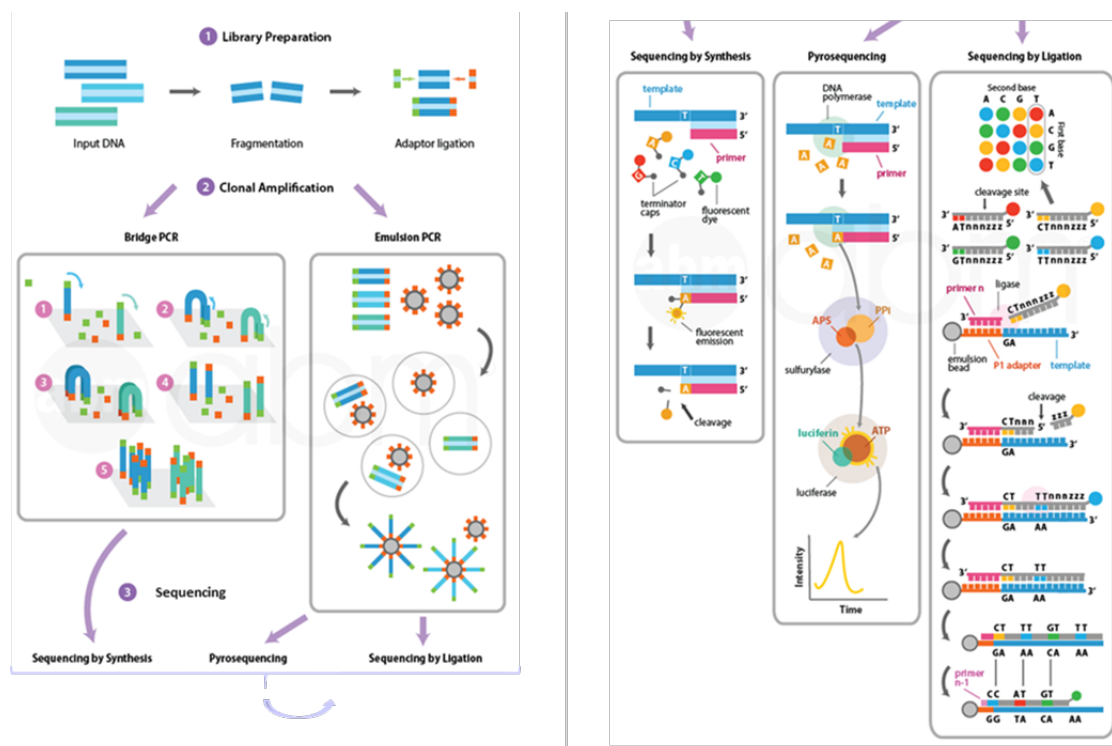
Figure 27. Scheme of the different NGS platforms.

4.2.3. Data analysis

The raw data generated by the different technologies undergo several bioinformatic processes, referred to as a pipeline, to ultimately result in a variant call file (VCF). This process includes demultiplexing, quality analysis, mapping the reads to a reference genome (resequencing), and variant identification and annotation (De Leng et al., 2016). First, in the demultiplexing step, all reads are sorted by bar code/sample before further analysis, generating a file called FASTQ file. After, the individual reads for a sample are mapped or aligned to a reference genome, resulting in the BAM file, and any difference between the reference and the sequencing read is noted. Then, duplicate or identical reads are discarded for capture-based sequencing (also for WGS), but not for amplicon-based sequencing. Later on, BAM files are analyzed to discover all sites with statistical evidence for an alternate allele presence. If multiple reads show the same difference, a variant is called. The quality of signal for an individual base read and the mapping quality are considered when calling a variant. The output file that defines all the variants for a sample and their allelic fractions is referred to as the VCF, and it will contain all variants detected. Additional bioinformatic tools can be used to filter out known sites of variations, machine artefacts, population variants, or variants previously identified as benign, for example, resulting in high-quality genotypes for each sample. Finally, this list of variants undergoes interpretation. It is noteworthy that BAM files can be analyzed with multiple bioinformatics tools to obtain diverse information from the sample. Although in the majority of applications the variants assessed are SNPs and short indels, large CNA or genomic rearrangements can also be detected by NGS sequencing.

4.2.4. Variant Interpretation

Translating the vast amount of data generated by NGS into meaningful clinical information is often a complex and a challenging process. For this reason, a robust and reliable framework is required for curating, interpreting, and reporting actionable variants.

In this sense, the American College of Medical Genetics and Genomics (ACMGG), the Association for Molecular Pathology (AMP), and the College of American Pathologists (CAP) joint efforts and published guidelines for interpretation and classification of germline variants in 2015 (Richards et al., 2015). This recommendation describes a process for classification of variants into five categories of pathogenicity based on various criteria (Figure 28A). More recently, the AMP, the American Society of Clinical Oncology (ASCO) and CAP jointly published guidelines to provide a systematic and consistent approach for the classification, interpretation, and reporting of somatic variants among laboratories (Li et al., 2017). In contrast to previous classification schemes, these guidelines focus on clinical significance and actionability, with less emphasis on the pathobiologic effects of the variants (Figure 28B). The intent of this shift was to highlight and prioritize alterations based on clinical evidence and streamline care decisions.

Taken together, the purpose of these guidelines is to establish uniformity across laboratories, present a reliable algorithm for classification of the less established variants, and provide the most accurate results that will help management decisions.

A

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 — Missense in gene where only truncating cause disease BP1 — Silent variant with non predicted splice impact BP7 — In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 — Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 — Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

B

Tier I: Variants of Strong Clinical Significance
Therapeutic, prognostic & diagnostic

Level A Evidence
FDA-approved therapy
Included in professional guidelines

Level B Evidence
Well-powered studies with consensus from experts in the field

Tier II: Variants of Potential Clinical Significance
Therapeutic, prognostic & diagnostic

Level C Evidence
FDA-approved therapies for different tumor types or investigational therapies
Multiple small published studies with some consensus

Level D Evidence
Preclinical trials or a few case reports without consensus

Tier III: Variants of Unknown Clinical Significance

Not observed at a significant allele frequency in the general or specific subpopulation databases, or pan-cancer or tumor-specific variant databases

No convincing published evidence of cancer association

Tier IV: Benign or Likely Benign Variants

Observed at significant allele frequency in the general or specific subpopulation databases

No existing published evidence of cancer association

Figure 28. Summary table of the classification of variants in the 2015 (A) and 2017 (B) guidelines. (Adapted from Richards et al., 2015; Li et al., 2017)

4.3. Technical considerations and limitations of NGS

The application of NGS technologies in the characterization of human tumours has provided unprecedented opportunities to understand the biological basis of different cancer types, discover new genomic biomarkers, develop targeted therapies, and to guide clinical decision-making regarding the genomic characteristics of patients. Thus, targeted panels are currently used for molecular diagnostic somatic testing in solid tumors and hematological malignancies. However, before the introduction of a new panel in the clinical routine, several issues need to be considered, as they will determine both the quality of the obtained results and the costs of its application:

1) **Target regions included in the panel:** Whether using a commercial or custom panel, the size of the panel may affect sequencing reagent cost, depth of sequencing and laboratory productivity. In multiplex libraries, the number of samples included in each run affects the final coverage to which each sample will be sequenced. In order to obtain cost-effective panels an equilibrium between the number of patients multiplexed and the minimum coverage required should be stablished. Therefore, the use of larger panels will supose either lower multiplexing capacity (higher cost) or diminished depth of sequencing, which ultimately will affect the sensitivity of the panel. It is recommended to include only those genes or genomic regions that have sufficient scientific evidence for the disease diagnosis, prognostication, or treatment (Joseph et al., 2016).

2) **Selection of the target enrichment and sequencing strategy:** As previously mentioned, each target enrichment approach varies in sensitivity (percentage of target bases that are represented by one or more sequence reads), specificity (percentage of sequences that map to the intended targets, known as on-target sequences), uniformity (variability in sequence coverage across target regions), reproducibility (correlation of results obtained from replicate experiments), cost, ease of use, amount of DNA required, and type of alteration that can be detected. Regarding sequencing platforms available, they differ in sequence capacity, sequence read length, sequence run time,

quality and accuracy of the data. Therefore, when choosing a sequencer there are several parameters, such as size of the targeted region, type of variation detected, required depth of coverage, projected sample volume, turnaround time requirements, and costs, that should be considered (Metzker, 2010; Mardis, 2013). Moreover, DNA fragments can be sequenced from one end, single-end, or from both ends matched to each other, paired-ends. Although paired-end reads are twice as expensive, they deliver more than twice the information, namely the read sequences plus their connectivity, and can be used to identify small-scale genomic rearrangements, such as insertions and deletions (indels) and translocations, in addition to aid mapping of repetitive regions of the genome (Ulahannan et al., 2013).

3) **Quantity and quality of samples:** Tumor samples often provide only small amounts of genetic material, particularly when collected as biopsy samples or fine needle aspirates. Such samples may also harbor low levels of tumor cellularity due to infiltration of non-malignant cells, leading to a reduction of detection sensitivity. Moreover, most tumors are processed and stored as FFPE tissue. This process typically leads to the fragmentation and chemical modification of DNA, which might affect the achievable depth of coverage and can elevate certain types of false positive results owing to DNA damage (Williams et al., 1999; Do & Dobrovic, 2015). Fortunately, advances in sample preparation methodologies have led to reductions in the minimum DNA input requirements and have made FFPE samples more amenable to NGS analysis (Van Allen et al., 2014; Einaga et al., 2017). However, low DNA input affects the complexity of the library, which is an important determinant of the quality of the results. In fact, libraries with good enrichment but very little DNA (generally <1 nanogram) may have complexity issues since they do not yield more than a few million distinct reads. Besides, high number of PCR cycles during the amplification step of library construction may also lower library complexity (Berger & Mardis, 2018).

4) **Inherent technical limitations of NGS:** Certain areas of the genome, such as homologous regions, repetitive regions, and GC-rich regions are not reliably interpretable by the current NGS platforms and standard bioinformatics algorithms. Homologous regions, including pseudogenes, are areas with high sequence similarity that may differ from the gene of interest by only a few base pairs. As a result, fragments of DNA from a target gene and homologous regions will have poor mapping quality, and reads belonging to the homolog can be mismapped to the real gene and vice versa, which can ultimately lead to false positive calls (Mandelker et al., 2014).

For repetitive areas, unique sequence flanking the repeat is required to reliably map a sequencing read and determine the size of the repeat. However, although they may be sequenced, the coverage will be less since some reads will not be informative. Conversely, if the repetitive regions are larger than the size of the DNA insert they will not have a flanking sequence and therefore will not be accurately mapped. Finally, GC-rich regions appear to have higher background noise and lower quality of sequencing, and in addition, Illumina sequencers in particular often give substitution errors in areas of high GC content and long G/C homopolymers (Samorodnitsky et al., 2015; Shin & Park 2016).

Besides, although NGS platforms work reasonably well at detection of SNV and small indels, it still have many limitations to detect structural rearrangements or CNA, mainly due to the insert size, which is the size of the DNA fragment between the adaptors (Nord et al., 2011; Pugh et al., 2016). Although some technologies provide longer insert size than others, such as hybrid-capture technologies, this size could not be sufficient to detect large genomic alterations (Amarasinghe et al., 2020).

Finally, there are difficulties associated with sequencing from polymerase amplified fragment populations (Dohm et al., 2008; Potapov & Ong, 2017). These include the early occurrence of polymerase errors during library

construction, as well as preferential amplification of certain fragments of the library that cause them to be differentially over represented compared to others. For this reason NGS technology has evolved towards single-molecule DNA sequencing, also known as third generation NGS. This approaches can eliminate the challenges arising from PCR amplification, and provide longer sequencing reads (long-read sequencing), which in turn eliminates the difficulties associated with small-read sequencing.

5) **Computational analysis and variant analysis:** Beyond important considerations about assay design and scope, the corresponding decisions regarding the computational data analysis pipeline that must accompany the NGS assay are similarly important. Although there are several software programs available to map the genome and call variants, there is a lack of consensus on which tools are the best. On the one hand, programs are continually evolving, so that the existing programs rapidly become out-dated and therefore comparisons are made between asynchronous versions. Moreover, programs may be designed for selected sequencing platforms and their specific characteristics, and may therefore lack the capacity to process data from competing manufacturers. Indeed, some programs from commercial sequencing vendors are not made freely available and the quality of analysis has to be taken 'on trust' (Ulahannan et al., 2013). On the other hand, programs available to detect CNA and SV are relatively new and, although they evolve rapidly, they are currently difficult to apply in clinical practice. Moreover, different bioinformatics pipelines have been shown to generate differences in variant outputs and accuracy (O'Rawe et al., 2013; Ross et al., 2013).

In addition, there is great complexity in translating the large amount of data generated by NGS into meaningful clinical information. Guidelines have been published that can help interpreting and classifying the variants (Richards et al., 2015; Li et al., 2017). However, various studies have observed disagreements in variant classification across different laboratories, which often occur due to differences in the approach used for searching, curating,

selecting, and weighing evidence. In addition, there is not the same evidence published, in terms of quality and availability, for all the variants (Amendola et al., 2016). On the other hand, the emphasis on actionability of the 2017 AMP/ASCO/CAP guidelines improve the clinical interpretation of sequencing reports based on therapeutic, diagnostic, and prognostic information. However, this approach is less useful for classifying somatic variants that may contribute to oncogenesis but whose actionability has not already been tested in clinical trials. Furthermore, the actionability of a variant may change over time, requiring the need for accurate variant designation supported by current evidence. Finally, regarding germline alterations, simultaneous testing of matched patient tumor and normal samples is not usually performed in clinical laboratories, and is generally addressed by a disclaimer or by follow-up testing of a germline sample in select cases (Li et al., 2017).

Patients and Samples

1. Patients

A total of 185 patients with different mature B-cell neoplasms and control individuals were included in the study. However, 14 patients were removed from the analysis due to low coverage obtained in the sequencing and will not be further described. Therefore, 171 patients were finally analysed, with the following diagnoses:

- o Diffuse large B-cell lymphoma: 55
- o Splenic marginal zone lymphoma: 38
- o Chronic lymphocytic leukemia: 37
- o Follicular lymphoma: 14
- o Mantle cell lymphoma: 12
- o Burkitt lymphoma: 5
- o Hairy cell leukemia: 4
- o Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue: 4
- o Splenic diffuse red pulp small B-cell lymphoma (SDRPL): 1
- o Follicular dendritic cell sarcoma (FDCS): 1

Patients were selected for presenting known genetic alterations detected by the different genetic techniques used in the clinical setting in order to be compared with the designed panel. Additional genomic information was also available for some patients included in previous studies. The information available for each entity before performing the sequencing is detailed in Table 13. It is important to note that, although it was not a MBCN, a case diagnosed of follicular dendritic cell sarcoma was selected for presenting an amplification of the *MYC* gene. Therefore, this case will only be taken into account for the CNA analysis.

Although most of the patients were from Hospital del Mar (Barcelona), in order to obtain a representative number of study subjects for each MBCN, especially for those with lower prevalence, samples from other centers included in collaborative studies were also collected. The distribution of patients among different hospitals is detailed below:

- o Hospital del Mar (Barcelona): 51 DLBCL, 16 CLL, 16 SMZL, 12 MCL, 8 FL, 4 BL, 2 MALTL, 1 SDRPL, and 1 FDCS

- o Hospital Clínic (Barcelona): 6 FL, 4 DLBCL, and 1 MALTL

- o Hospital Universitario Cruces (Bilbao): 7 CLL

- o Hospital Universitario Central de Asturias (Oviedo): 6 SMZL

- o Hospital Universitari de Bellvitge (L´Hospitalet de Llobregat): 4 SMZL and 1 HCL

- o Hospital de Sant Pau (Barcelona): 5 CLL

- o Hospital Vall d'Hebron (Barcelona): 4 CLL and 1 SMZL

- o Hospital General Universitario de Valencia (Valencia): 4 CLL

- o Hospital Francesc de Borja (Gandía): 2 HCL

- o Hospital Universitario Basurto (Bilbao): 2 SMZL

- o Hospital Universitario Morales Meseguer (Murcia): 2 SMZL

- o Hospital Germans Trias i Pujol (Badalona): 1 SMZL

- o Hospital Municipal de Badalona (Badalona): 1 SMZL

- o Hospital Txagorritxu (Vitoria-Gasteiz): 1 SMZL

- o Institut Català d'Oncoloia (L´Hospitalet de Llobregat): 1 BL

- o Hospital Universitario de Canarias (Santa Cruz de Tenerife): 1 SMZL

- o Hospital Clínico de Valencia (Valencia): 1 SMZL

- o Hospital Universitario la Fe de Valencia (Valencia): 1 HCL

- o Hospital Universitario Marqués de Valdecilla (Santander): 1CLL

- o Hospital Universitari Mútua de Terrassa (Terrassa): 1 SMZL

- o Hospital Universitari Sant Joan de Reus (Reus): 1 SMZL

- o Hospital Val d'Aran (Vielha): 1 MALT lymphoma

The study was performed in accordance with national and international guidelines (Professional Code of Conduct, Declaration of Helsinki) and approved by the Ethics Committee of Hospital del Mar (2015/6202/I).

Table 13. Genetic information available for each entity previous to perform the sequencing.

Pathology	Number of cases with karyotype information	Number of cases with genomic arrays information	Number of cases with FISH information	Number of cases with mutational study
DLBCL (N=55 cases)	19/55 *6/55	25/55	43/55 with *BCL2* information 37/55 with *MYC* information 34/55 with *BCL6* information 12/55 with 17p information 3/55 with 1p36 and *CCND1* information 1/55 with *CCND2, FOXP1,* and *BCL3* information	6/55 with *MYD88* information 3/55 with *TP53* information 2/55 with *EZH2* information 1/55 with *NOTCH1, NOTCH2, BRAF, MAP2K1, CXCR4, TNFRSF14, KLF2,* and *SF3B1* information
SMZL (N=38 cases)	38/38	20/38	12/38 with 17p information 12/38 with *BCL6* information 8/38 with 7q information 3/38 with information of *+12* probe 3/38 with information of *11q, 13q* probes	8/38 with *MYD88* information 1/38 with *BRAF* information
CLL (N=37 cases)	37/37	30/37	31/37 with information of 11q, 13q, 17p CLL probes 29/37 with information of +12 probe 1/37 with *CCND1* information	29/37 with *TP53* information 11/37 with *NOTCH1, POT1, XPO1*, and *SF3B1* information 10/37 with *BIRC3* information 8/37 with *ATM* information

Pathology	Number of cases with Karyotype inormation	Number of cases with Array information	Number of cases with FISH information	Number of cases with mutational study
FL (N=14 cases)	5/14 *2/14	7/14	13/14 with *BCL2* information 11/14 with *BCL6* and *MYC* information 1/14 with information of 1p36 and 17p probes	
MCL (N=12 cases)	7/12 *4/12	4/12	5/12 with *CCND1* information 2/12 with 17p information 1/12 with 11q information	1/12 with *TP53* information
BL (N=5 cases)	2/5	3/5	3/5 with *MYC* information 1/5 with *BCL2* and *BCL6* information	
HCL (N=4 cases)	4/4	3/4	1/4 with information of *BCL6*, 7q, and 17p probes	2/4 with *BRAF* information 1/4 with *TP53, MYD88, NOTCH1, NOTCH2, EZH2, ID3, MAP2K1, CXCR4, TNFRSF14, KLF2,* and *SF3B1* information
MALTL (N=4 cases)	1/4 *1/4	1/4	1/4 with information of *MYB, BCL6, BCL2, MALT1/IGH,* and *MALT1/API2* probes	1/4 with *MYD88* information
FDCS (N=1 case)	*1/1	1/1	1/1 with information of *MYC* and 17p probes	1/1 with *BRAF* information
SDRPSBCL (N=1 case)	1/1			

* Karyotype was performed but metaphases were not obtained to report the result.

2. Samples

A total of 178 samples from 171 patients were processed in the present study, including samples from peripheral blood (PB), bone marrow or different tissues with documented tumoral infiltration. The vast majority of the sequenced DNA was extracted from fresh samples (n=125) or frozen tissues (n=49), while four formalin-fixed faraffin-embedded tissue samples were included to assess the utility of the panel in this setting as a pilot study. The type of samples is detailed as follows:

1. Fresh samples

o 80 peripheral blood samples collected in a container with EDTA or heparin, of which:

- In 67 cases DNA was obtained from the purified CD19+ fraction selected by immunomagnetic methods (CD19 Microbeads, Miltenyi Biotec, Bergisch Gladbach, Germany).

- In 11 cases DNA was extracted from peripheral blood mononuclear cells (PBMC) isolated using a Ficoll (Lymphoprep, Axis-Shield, Dundee, Scotland) gradient.

- In 2 cases DNA was obtained directly from the peripheral blood.

o 7 bone marrow aspirates, collected in a sterile container with heparin.

o 24 fresh tissue biopsies from the MARBiobanc tissue bank, collected in a sterile container. The samples corresponded to the following tissues:

- Lymph node: 16; Skin: 2; Spleen: 2; Peritoneum: 2; Stomach: 1; Thyroid: 1.

o In 14 cases DNA extracted from fresh samples and stored in a sterile 1.5 mL tube was received from the centre of origin:

- In 8 cases DNA was extracted from peripheral blood.

- In 6 cases DNA was extracted from tissue biopsies, from the following tissues:

  - Lymph node: 4; Skin: 1; Peritoneum: 1.

2. Frozen tissue samples

o 49 samples included in Tissue-Tek® O.C.T™ Compound (Tissue-Tek O.C.T. Compound, Sakura Finetek, Rijn, The Netherlands) obtained from the MARBiobanc tissue bank. A total of 15 serial 10-15 µm cuts for each case were

provided in sterile 1.5 ml tubes. These samples corresponded to the following tissues:

- Lymph node: 32; Spleen: 5; Testicle: 4; Intestine: 2; Bone marrow biopsy: 1; Colon: 1; Stomach: 1; Thyroid: 1; Skin: 1; Adipose tissue: 1.

3. FFPE tissue samples

o 4 samples from FFPE tissues, obtained from the MARBiobanc tissue bank that provided 15 serial 10-15 µm cuts made in the cryostat that were stored in a sterile 1.5 ml tube. The samples corresponded to the following tissues:

- Lymph node: 2; Spleen: 1; Peritoneum: 1

DNA from these samples was sequenced and analyzed using two different panel versions (panel version 1 (PV1) and panel version 2 (PV2). The distribution of the total study cohort in the different panel versions is detailed in the Table 14.

It is important to note that a total of 176 samples from 171 patients were processed. On one hand, two samples from two different tissues were included for the same patient (peripheral blood and spleen), to analyze the abnormalities of the tumor infiltration in both locations. On the other hand, the four FFPE tissues included corresponded to patients of whom a frozen tissue sample from the same biopsy was also incorporated, to sequence them in parallel and assess the performance of the panel analyzing FFPE samples. Moreover, two samples were sequenced with both panel designs to analyze whether the second panel design improved the detection of the alterations for which the design had been optimized. These patients (diagnosed with DLBCL and FL, respectively) were selected because the *BCL2* rearrangement could not be detected in the first panel design, which was previously characterized by FISH.

Table 14. Distribution of the samples sequenced with the two panel designs.

PANEL DESIGN 1		
Diagnosis	**Sample type**	**Tissue**
37 CLL	32 PB	19 CD19+
		11 PBMC
		2 PB
	5 BM	5 BM aspirate
25 DLBCL	18 Fresh tissue	*10 Lymph node
		2 Skin
		2 Peritoneum
		1 Spleen
		1 Stomach
		1 Thyroid
		1 PB
	7 Frozen tissue	6 Lymph node
		1 Testicle
19 SMZL	18 PB	18 CD19+
	1 Frozen tissue	1 BM biopsy
10 FL	9 Fresh tissue	6 PB
		3 Lymph node
	1 Frozen tissue	*1 Lymph node
6 LCM	6 PB	6 CD19+
5 BL	3 Fresh tissue	3 Lymph node
	2 BM	2 BM aspirate
3 MALTL	2 Fresh tissue	1 PB
		1 Skin
	1 Frozen tissue	1 Intestine
3 HCL	3 PB	3 CD19+

PANEL DESIGN 2		
Diagnosis	**Sample type**	**Tissue**
34 DLBCL	26 Frozen tissue	*14 Lymph node 4 Spleen 3 Testicle 1 Skin 1 Adipose tissue 1 Colon 1 Thyroid 1 Intestne
	5 Fresh tissue	**3 Lymph node **1 Spleen **1 Peritoneum
	3 FFPE tissue**	1 Lymph node 1 Spleen 1 Peritoneum
19 SMZL	19 PB	19 CD19+
6 MCL	6 Frozen tissue	5 Lymph node 1 Estomach
6 FL	4 Frozen tissue	**4 Lymph node*
	1 Fresh tissue	1 Lymph node
	1 FFPE tissue**	1 Lymph node
2 SDRPL*	1 PB	1 CD19+
	1 Frozen tissue	1 Spleen
1 HCL	1 PB	1 CD19+
1 MALTL	1 Frozen tissue	1 Lymph node
1 FDCS	1 Frozen tissue	1 Lymph node

PB: Peripheral blood; BM: Bone marrow; PBMC: peripheral blood mononuclear cells; CD19+: purified CD19+ fraction.

* Two patients were sequenced with both panel designs; ** FFPE tissue samples sequenced in parallel to fresh tissue from the same biopsy; *** Two samples from two different tissues belonging to the same patient were sequenced

Methods

1. **DNA extraction**

1.1. DNA extraction from fresh samples

DNA extraction from peripheral blood, bone marrow and lymph node fresh samples was performed using the DNA Gentra Puregene Blood kit (Qiagen, Düsseldorf, Germany). The protocol described is adjusted for 3ml of whole peripheral blood or 20 millions of total cells. For lymph node samples, prior mechanic disgregation with a syringe should be performed in order to obtain a cellular suspension, which will be counted to adjust the number of cells for the extraction.

1.1.1. Cell lysis

1. This first step consisted of a lysis of the red blood cells and was omitted for pellets of selected PBMC and CD19+ samples:

- 9 mL of Red Blood Cell Lysis Solution were dispensed into a 15 mL tube containing the sample, mixed by inverting 10 times and incubated 5 min at room temperature (15–25°C).

- After a centrifuge for 2 min at 2,000g a pellet of cells was obtained

2. The pellet was re-suspended in the remaining residual volume with a vortex and 3 mL of Cell Lysis Solution was added.

3. The solution was mixed well and vortexed until it was homogeneous. If cell clumps were visible, the mix was incubated ar room temperature for one to ten days.

1.1.2. Protein precipitation

1. One mL of Protein Precipitation Solution was added and vortexed to mix it well.

2. Then it was centrifuged for 5 min at 2,000g. A good compact pellet corresponding to the proteins was obtained. If not, the vortex and centrifugation was repeated.

1.1.3. DNA precipitation

1. Three mL of 100% isopropanol (2-propanol) was placed in a new tube.

2. The supernatant obtained after centrifugation was added to the tube with a Pasteur pipette.Then it was inverted about 50 times gently. Visible DNA aggregates should be observed.

3. Centrifuged for 3 min at 2,000g.

4. The supernatant was discarded by pouring it into a beaker taking care not to lose the pellet. The tube was kept upside down for a while to dry. In case that any drops remained attached to the tube, they were dried with paper, taking care not to touch the pellet.

5. Three mL of 70% ethanol was added. The tip of the tube was gently tapped to remove the pellet.

6. Centrifuged for 3 min at 2,000g.

7. The supernatant was discarded again by pouring it into a beaker taking care not to lose the pellet. The tube was kept upside down for 15-45 min to dry. In case that any drop remained attached to the tube, they were dried with paper, taking care not to touch the pellet.

1.1.4. DNA hydration

1. Between twenty and two hundred μL of Buffer TE Low EDTA was added according to the size of the pellet and incubated for 1 hour in the 65ºC stove.

2. It was left at room temperature overnight.

3. DNA was transferred to a 1.5 mL eppendorf.

4. Samples were held at 4ºC or at -20ºC for long storage.

1.2. DNA extraction from frozen tissue samples

DNA extraction from samples included in Tissue-Tek® O.C.T™ Compound was performed using QIAmp tissue kit (Qiagen).

1.2.1. Removal of OCT residues from the sample

1. 1 mL of 1x PBS was added to the eppendorf.

2. A vortex was performed until the tissue was disengaged from the 1.5 mL tube.

3. It was spun at 3,000 rpm for 5 min.

4. Supernatant was removed with a 1 mL Pasteur pipette.

1.2.2. Sample digestion

1. 180 µL of ATL buffer and 20 µL of proteinase K were added and mixed with vortex.

2. The eppendorf was sealed with parafilm and incubated in a 56ºC bath overnight for proteinase K to digest tissue.

1.2.3. DNA extraction

1. Before starting, the dry bath was turned on to 70ºC and an aliquot of sterile water was preheated.

2. The parafilm was removed and the sample was briefly vortexed.

10. Two hundred µLof AL buffer was added and mixed with vortex. It was then incubated in a dry bath at 70ºC for 10 min.

3. Two hundred and ten µL of absolute ethanol were added and mixed with vortex for 15 sec.

4. The mixture was passed through a purification column coupled to a collecting tube and centrifuged for 1 min at 8,000 rpm. The collecting tube was then discarded.

5. The column was placed in a new collection tube and 500 µL of AW1 buffer was added and centrifuged for 1 min at 8,000 rpm.

6. The collection tube was discarded again and the column was placed in a new collection tube. 500 µL of AW2 buffer was added and centrifuged at full speed for 3 min.

7. The collection tube was discarded again and the column was placed in a sterile 1.5 mL tube.

8. Fifty µL of distilled water preheated to 70ºC was added and incubated for 1 min at room temperature to hydrate the column before centrifuging.

9. It was centrifuged for 1 min at 8,000 rpm.

10. The 50 µL were collected and re-filtered through the column to collect any remaining DNA and centrifuged for 1 min at 8,000 rpm.

11. Samples were held at 4ºC or at -20ºC for long storage.

1.3. DNA extraction from Formalin-Fixed Paraffin-Embedded tissue samples

DNA extraction of the 4 FFPE tissue samples was carried out by the collaborating company qGenomics (Esplugues de Llobregat, Barcelona) using the Maxwell® RSC DNA FFPE Kit (Promega, Madison Winsconsin, USA).

2. Panel design

A custom capture NGS panel was designed with NimbleGen SeqCap Ez Choice (Roche Nimblegen, Madison Winsconsin, USA). To make the custom design, the Nimbledesign software, a free online tool that Nimblegen offered to the users, was used. This software allowed the design of the panel in a fast and simple way, by attaching a BED (Browser Extensible Data) file with the coordinate information to be captured. The genomic coordinates of the regions of interest were obtained using the Table Browser tool of The University of California Santa Cruz (UCSC) Genome Bioinformatics Web site (http://genome.ucsc.edu).

To select the regions to be included in the panel, an extensive bibliographic review of the different clinical guidelines, as well as recent publications, was carried out. In this way, a first selection was made of the main mutations and chromosomal alterations with clinical impact described in the different MBCN. These alterations were then divided into three groups according to their frequency, and their diagnostic, prognostic, and predictive value of response to treatment: group of priority I, which comprised the essential alterations to be included in the panel that are currently assessed by other methodologies in routine practice; group of priority II which includeds less frequent alterations and group of priority III which comprised those abnormalities with not well established (or unrecognized) clinical value. As the inclusion of all the alterations represented a large increase in the size of the panel, only the alterations corresponding to groups I and II were finally taken into account. In addition, the mutation hotspots and the rearrangement breakpoints described in the literature and their frequencies, as well as the candidate genes previously suggested as pathogenic for the CNA, were collected. The information taken into consideration to design the region covered for each target alteration is summarized in the Table 15.

With this information a first panel version was performed. PV1 was 4.6Mb in size and covered 11 genes involved in translocations, 33 frequently mutated genes, and 17 regions involved in copy number gain or loss. In addition, 9,111 SNP distributed throughout the genome were also included to obtain a general analysis of it (McKerrell et al., 2016). More in detail, the design included the regions frequently involved at breakpoints, in the case of translocations; all exons and UTR regions of genes involved in single nucleotide variants (SNV) and small indel, although some present mutation hotspots; as well as exons and UTR regions of genes previously suggested as pathogenic in the described cytobands involved in CNA (Table 16).

To optimize the applicability of the panel in the clinical setting, a second version was performed. In this case, the panel size was significantly reduced (1.4Mb) since infrequent target alterations such as *CCND2, CCND3, MALT1, BIRC3, BCL10, FOXP1* and *IRF4* translocations were not included in the design. In addition, the 9,111 SNP included in the previous design were also discarded. Besides, it was considered

appropriate to include only the exons of the genes involved in CNA and SNV, and the UTR regions exclusively for those genes with described mutations in these regions, such as the case of *NOTCH1* and *CCND1*. On the other hand, as the size of the panel decreased considerably, *CCND1* and *CCND3* genes were incorporated into the mutational study, since these genes present mutations described in the literature and were not covered in the PV1 (Schmitz et al., 2012; Beà et al., 2013; Greenough & Dave, 2014). In addition, as some rearrangements were missed with PV1, the design for *BCL2*, *MYC,* and *CCND1* translocation targets was improved as the capture region was expanded, including regions with described breakpoints that had not been incorporated in the previous panel design. In summary, PV2 covers four genes involved in translocations, 35 frequently mutated genes, and 16 regions involved in CNA (Table 16)

The technical parameters for each panel design are displayed in Table 17.

In addition, the Figure 29 show in detail the captured region for the target rearrangements of *BCL2, CCND1, BCL6,* and *MYC* genes for both panel versions.

Table 15. Summary of the data published in the literature that has been taken into account for the design of each target alteration.

TARGET ID	Genes involved/Hotspots described	SNV	CNA	Rearrangement	CLL	DLBCL	SMZL	FL	MCL	BL	MALT	HCL	Bibliography
BCL2	Main breakpoint in exon 3 and 3' *BCL2* (up to 20Kb) Downstream the gen: MBR, mcr, and 3'MBR (ICR); Rarely in 5'MBR (VCR)			X		▨		■					Akasaka et al, 1998; Van Dongen et al, 2003; Espinet et al, 2008
BCL6	88% Major Breakpoint: exon 1 and part of intron 1 Alternative Breakpoints in 245-285Kb 5' *BCL6*			X		▨							Yoshida et al, 2000; Ohno, 2006; Iqbal et al, 2007; Jardin et al, 2007; Flodr et al, 2014; Jarosova et al, 2015
MYC	Very variable: from 1Mb 5'*MYC* to 1Mb 3'*MYC* > 77% of cases: located from 381bp in 5' of exon 1 until 4415bp of intron 1 > 85-92%: breakpoints from 190Kb 5' to 50Kb 3'*MYC*			X	▨	▨		▨		■			Haralambieva et al, 2004; Einerson et al, 2006; Bertrand et al, 2007; Busch et al, 2007; Boerma et al, 2009; Royo et al, 2011; Karube et al, 2015; Nguyen et al, 2017
CCND1	95% MCL with IGH-*CCND1* translocation: >30/50% MTC (1Kb region) >50/70% breakpoints distributed over 120-130 kb 5'*CCND1* Breakpoints in 3'UTR of *CCND1*			X					■				Raynaud et al, 1993; Greisman et al, 2011 ; Wiestner, et al 2007; Menke et al, 2017
CCND2	Region limited by the BACs described			X					■				Royo et al, 2011; Salaverria et al, 2013; Campo et al, 2015
CCND3	Region limited by the BACs described: BACs:RP1-321B9 <-> RP1-242G1 >Breakpoint approximately 0.5Mb centromeric to *CCND3*			X					■				Royo et al, 2011
BCL10	Region limited by the BACs described: RP11-1080I1 <-> RP11-40K4 <-> RP11-1077C10/RP11-36L4 > Breakpoints in 5' *BCL10*			X							■		Willis et al, 1999
FOXP1	Region limited by the BACs described: RP11-154H23 Breakpoint: 38Kb upstream of the first 5' noncoding exon			X							■		Streubel et al, 2005
IRF4	Region limited by the BACs described: RP3-416J7 <-> RP5-1077H22 > Breakpoints downstream of exon 22 and exon 23 of *EXOC2*			X		■							Swerdlow et al, 2016
MALT1	> Breakpoints distributed through the gene. Mainly in exons 2, 5, 7 or 8.			X							■		Baens et al, 2000
BIRC3	> Breakpoints in exon 7			X							■		Baens et al, 2000

104

		Alteration type			Pathology affected by the alteration								
TARGET ID	**Genes involved/Hotspots described**	SNV	CNA	Rearrangement	CLL	DLBCL	SMZL	FL	MCL	BL	MALTL	HCL	**Bibliography**
TNFRSF14		X						X					Kridel et al, 2012
ARID1A	HCL: p.V142fs	X						X		X		X	Love et al, 2012; Weigert et al, 2012; Greenough et al, 2014; Dietrich et al, 2015; Pastore et al 2015
BCL10		X									X		Willis et al, 1999; Du et al, 2000
ID3	Helix 1, helix 2, RGYW motif	X								X			Love et al, 2012; Richter et al, 2012; Greenough et al, 2014
NOTCH2	Variants described in exons 25-34; Hotspot in exon 34	X					X		X				Rossi et al, 2012; Kiel et al 2015
CXCR4	WHIM-like variants (C-ter): 311-345 Aa	X								X			Hunter et al, 2014; Treon et al, 2014; Pastore et al, 2015; Schmidt et al, 2015
SF3B1	Hotspot: exons 14, 15 and part of 16; The most frequents in CLL: p.K700E, p.G742D, p.Y623C	X			X								Baliakas et al, 2014; Mitsui et al, 2015; Puente et al, 2015; Rodríguez et al, 2015; Vollbrecht, 2015
XPO1	The most frequent: p.E571G or p.E571I/K/Q in the exon 15; some in exon 16	X			X								Puente et al, 2012; Rodríguez et al, 2015; Vollbrecht, 2015
MYD88	The most frequent: p.L265P	X			X	X							Puente et al, 2012; Hunter et al, 2013; Xu et al, 2013; Puente et al, 2015; Schmidt et al, 2015; Swerdlow et al, 2015
WHSC1	*MCL: Hotspot in exons 18 and 19; The most frequents (10%): p.E1099K in exon 18, and p.T1150A in exon 19	X							X				Bea et al, 2013; Campo et al, 2015
CCND3		X								X			Schimitz et al, 2012; Greenough et al, 2014

		Alteration type			Pathology affected by the alteration								
TARGET ID	**Genes involved/Hotspots described**	SNV	CNA	Rearrangement	CLL	DLBCL	SMZL	FL	MCL	BL	MALTL	HCL	**Bibliography**
TNFAIP3		X			X							X	Pascualucci et al, 2011; Hunter et al, 2014
ARID1B		X			X							X	Hunter et al, 2014
BRAF	87% p.V600E variant in exon 15 <4% variants in exon 11: p.F468C, p.D449E // *CLL: variants other than p.V600E: p.G469A and p.D638E	X			X							X	Tschernitz et al, 2014
CARD11	Coiled-coil domain (Aa: T110-L442)// *FL: 80% in coiled-coil domain	X				X	X	X					Lanz et al, 2008; Bohers et al, 2014; Pastore et al, 2015
EZH2	Hotspot: Tyr641; also Ala677, but less frequent	X				X		X				X	Kridel et al, 2012; Lunning et al, 2015; Pastore et al, 2015
POT1	76% at OB fold domains: from 1st Aa till 280 Aa. Other variants described till 591 Aa	X			X		X						Landau et al, 2015; Rosenquist et al, 2017
UBR5	Hotspot: exon 58 (clustered in100 bp)	X							X				Meissner et al, 2013; Kutovaya et al, 2016
NOTCH1	Exon 14: del of ΔCT in 7544-7545 is the most frequent Outside exon 14: <1% 3'UTR: p.P2514Rfs*4, p.Q2503*, p.F2482Ffs*2	X			X				X				Baliakas et al, 2014; Puente et al, 2015; Vollbrecht et al, 2015
TRAF2	Throughout the gene	X							X				Campo et al, 2015
ATM		X			X				X				Campo et al, 2015; Landau et al, 2015
BIRC3	Hotspot: exon 9	X			X				X				Bea et al, 2013; Campo et al, 2015; Landau et al, 2015
CCND1	Indels, dups and SNV in **3'UTR** of *CCND1* (first 320bp of 3'UTR); Alterations in N-terminus of *CCND1* and exon 1	X			X				X				Wiestner et al, 2007; Bea et al, 2013
MLL2	Throughout the gene	X			X				X				Pascualucci et al, 2011; Kridel et al, 2012; Bea et al, 2013; Campo et al, 2015; Testoni et al, 2015

TARGET ID	Genes involved/Hotspots described	Alteration type			Pathology affected by the alteration								Bibliography
		SNV	CNA	Rearrangement	CLL	DLBCL	SMZL	FL	MCL	BL	MALTL	HCL	
FOXO1	Hotspot: ~90% exon 1	X				X		X		X			Trinh et al, 2013; Pastore et al, 2015
CHD2	Throughout the gene	X				X		X		X			Landau et al, 2015
MAP2K1	Hotspot: exons 2 and 3	X				X		X		X			Waterfall et al, 2014; Nicola et al, 2016
CREBBP	Throughout the gene	X				X		X		X			Pascualucci et al, 2011; Kridel et al, 2012; Pastore et al, 2015
TP53	90% in exons 4-10	X			X	X	X	X	X	X		X	Salaverria et al, 2007; Royo et al, 2011; Salaverria et al, 2013; Campo et al, 2015; Puente et al, 2015
CD79B	The most frequent: Y196 (at ITAM domain, exons 5 and 6)	X				X	X						Schwaenen et al, 2009; Kim et al, 2013; Bohers et al, 2014
CD79A	ITAM domain (exons 4 and 5)	X				X	X						Kim et al, 2013; Bohers et al, 2014
KLF2	Throughout the gene	X					X						Clipson et al, 2014; Parry et al, 2015; Piva et al, 2015
MEF2B	Hotspot: 80% at exons 2 and 3 (at MADS or MEF2 domains) *MCL: p.N49S (at exon 2)	X				X		X	X	X			Morin et al, 2011; Kridel et al, 2012; Pastore et al, 2015
EP300	Throughout the gene	X				X		X					Morin et al, 2011; Pascualucci et al, 2011; Kridel et al, 2012; Pastore et al, 2015
TCF3		X								X			Greenough et al, 2014; Swerdlow et al, 2016

		Alteration type			Pathology affected by the alteration								
TARGET ID	Genes involved/Hotspots described	SNV	CNA	Rearrangement	CLL	DLBCL	SMZL	FL	MCL	BL	MALT	HCL	Bibliography
Loss 1p36	*TNFRSF14; CDC2L1 (or CDK11B); TNFRSF9; TNFRSF18; TNFRSF25; PRDM16; DFFB; ID3*		X			X		X		X			Cheung et al, 2009; Schwaenen et al, 2009; Kridel et al, 2012
Gain 2p24-2p12	*REL; BCL11A; MYCN; MSH2*		X		X	X		X					Nedomova et al, 2013; Testoni et al, 2015
Gain chr 3/3q	*FOXP1; NFKBIZ*		X			X	X		X		X		Salido et al, 2010; Nedomova et al, 2013; Testoni et al, 2015
Gain chr 5/5q			X									X	Durham et al, 2017
Loss 6q13-q26	*PRDM1; TNFAIP3; EPHA7; PERP*		X			X	X	X	X	X	X		Henderson et al, 2004; Cheung et al, 2009; Kridel et al, 2012; Nedomova et al, 2013; Nguyen-Khac et al, 2013; Testoni et al 2015
Loss 7q22.1-q35	*IRF5*		X					X			X	X	Salido et al, 2010
Loss 8p21.1-p23.1	*TNFRSF10A; TNFRSF10B; KCTD9; BIN3*		X		X								Schwaenen et al, 2009; Puiggros et al, 2014
Gain 8q24 (*MYC*)	*MYC*		X										Schwaenen et al, 2009; Hunter et al, 2014; Puiggros et al, 2014
Loss 9p21	*CDKN2A; CDKN2B; MTAP*		X			X		X	X				Salaverria et al, 2007; Schwaenen et al, 2009; Royo et al, 2011; Salaverria et al, 2013; Testoni et al, 2015
Loss 11q22.1-q24	*ETS1; FLI1; PAFAH1B2; ZNF259 (or ZPR1); ATM; BIRC3*		X		X			X	X				Schwaenen et al, 2009; Puiggros et al, 2014; Vollbrecht et al, 2015
Gain chr12	*MDM2; CDK4; IGF1; ELK3; EPS8; IFNG; P27 (or CDKN1B); HIP1R; MYF6*		X		X	X			X	X	X		Schwaenen et al, 2009; Salido et al, 2010; Braggio et al,2012; Testoni et al, 2015
Loss 13q14	miR-15a; miR16-1; *DLEU2; RB1*		X		X								Parker et al, 2011
Loss 16p13 (*CREBBP*)	*CREBBP*		X			X		X			X		Pascualucci et al, 2011; Kridel et al, 2012; Pastore et al, 2015
Loss 17p13 (*TP53*)	*TP53*		X		X	X	X	X	X	X	X	X	Salaverria et al, 2007; Royo et al, 2011; Salaverria et al, 2013; Campo et al, 2015; Puente et al, 2015
Gain chr18	*BCL2*		X			X		X		X	X		Cheung et al, 2009; Schwaenen et al, 2009; Kridel et al, 2012
Loss 19p13 (*KLF2*)	*KLF2*		X				X						Clipson et al, 2014; Parry et al, 2015; Piva et al, 2015
Loss 22q13 (*EP300*)	*EP300*		X			X		X					Morin et al, 2011; Pascualucci et al, 2011; Kridel et al, 2012; Pastore et al, 2015

The type of alteration caused by each target region is indicated with an; Diagnostic markers are indicated in dark yellow; Prognostic markers are indicated in light yellow; Therapeutic markers are indicated in yellow; Abbreviations: Aa: Amino acid

Table 16. Summary of the regions covered for each target alteration in each version of the panel

TARGET ID	SNV	CNA	Alteration type Rearrangement	PV1 Included	Region	PV2 Included	Region
BCL2			X	YES	30Kb 3'*BCL2* + exon 3 chr18:60769579-60795992	YES**	50Kb 3'*BCL2* + *BCL2* + 50Kb 5'*BCL2* chr18:60740579-61036613
BCL6			X	YES	Exon 1 + intron 1 + 300Kb 5'*BCL6* chr3:187452677-187763513	YES	Exon 1 + intron 1 + 300Kb 5'*BCL6* chr3:187452677-187763513
MYC			X	YES	100Kb3'*MYC* + *MYC* + 100Kb 5'*MYC* chr8:128648316-128853680	YES**	*MYC* + 200Kb 5' + 50kb 3' + BVR1 and BVR2 regions at 3'*MYC* chr8:128548315-128803680 chr8:128893680 -128904680 chr8:129144680-129398680
CCND1			X	YES	122Kb 5'*CCND1* chr11:69333873-69455873	YES**	122Kb 5'*CCND1* + *CCND1* chr11:69325873-69469242
CCND2			X	YES	chr12:4203622-4443393	NO	
CCND3			X	YES	chr6:42351409-42475549	NO	
BCL10			X	YES	chr1:85692230-85802513	NO	
FOXP1			X	YES	chr3:71633140-71733818	NO	
IRF4			X	YES	chr6:213735-835761	NO	
MALT1			X	YES	chr18:56348402-56348568 chr18:56376610-56376788 chr18:56378153-56378185 chr18:56381315-56381341 chr18:56381315-56381341 chr18:56383167-56383199	NO	
BIRC3			X	YES	chr11:102201730-102201972 chr11:102206697-102206951	NO	

TARGET ID	SNV	CNA	Rearrangement	PV1 Included	Region	PV2 Included	Region
TNFRSF14	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
ARID1A	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
BCL10	X		*	**YES**	**Covered for translocation (exons+introns)**	**YES**	All the EXONS
ID3	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
NOTCH2	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CXCR4	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
SF3B1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
XPO1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
MYD88	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
WHSC1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CCND3	X		*	**NO**	**NOT covered (breakpoints outside the gene)**	**YES**	All the EXONS
TNFAIP3	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
ARID1B	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
BRAF	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CARD11	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
EZH2	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
POT1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
UBR5	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS

	Alteration type			PV1		PV2	
TARGET ID	SNV	CNA	Rearrangement	Included	Region	Included	Region
NOTCH1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS + UTR
TRAF2	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
ATM	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
BIRC3	X	*	*	**YES**	All the EXONS + UTR	**YES**	All the EXONS
CCND1	X		*	**NO**	**NOT covered (breakpoints outside the gene)**	**YES**	All the EXONS + UTR
MLL2	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
FOXO1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CHD2	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
MAP2K1	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CREBBP	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
TP53	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
CD79B	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
CD79A	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
KLF2	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
MEF2B	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS
EP300	X	*		**YES**	All the EXONS + UTR	**YES**	All the EXONS
TCF3	X			**YES**	All the EXONS + UTR	**YES**	All the EXONS

		Alteration type		PV1		PV2	
TARGET ID	SNV	CNA	Rearrangement	Included	Region	Included	Region
Loss 1p36		X		YES	Genes + SNP	YES	Genes
Gain 2p24-2p12		X		YES	Genes + SNP	YES	Genes
Gain chr 3/3q		X		YES	Genes + SNP	YES	Genes
Gain chr 5/5q		X		YES	SNP	NO	
Loss 6q13-q26		X		YES	Genes + SNP	YES	Genes
Loss 7q22.1-q35		X		YES	Genes + SNP	YES	Genes
Loss 8p21.1-p23.1		X		YES	Genes + SNP	YES	Genes
Gain 8q24 (*MYC*)		X	*	YES	*MYC* gene (the region included for translocation + exons + introns) + SNP	YES	*MYC* gene + the region included for translocation
Loss 9p21		X		YES	Genes + SNP	YES	Genes
Loss 11q22.1-q24		X	*	YES	Genes + SNP	YES	Genes
Gain chr12		X		YES	Genes + SNP	YES	Genes
Loss 13q14		X		YES	Genes + SNP	YES	Genes
Loss 16p13 (*CREBBP*)		X		YES	Genes + SNP	NO	*Not considered target, but the gene included for SNV
Loss 17p13 (*TP53*)		X	*	YES	Relevant genes + SNP	YES	Genes
Gain chr18		X	*	YES	*BCL2* gene (only the region included for translocation) + SNP	YES	All the exons + the region included for translocation
Loss 19p13 (*KLF2*)		X		YES	Genes + SNP	NO	*Not considered target, but gene included for SNV
Loss 22q13 (*EP300*)		X		YES	Genes + SNP	NO	*Not considered target, but gene included for SNV

The target regions covered in each type of alteration are indicated with an X. In addition, the regions included to analyze other types of alteration are marked with an *
in the corresponding column. ** The capture region for the *BCL2, MYC* and *CCND1* genes was modified in the second panel design.

Table 17. Technical parametres obtained from the NimbleGen design report

Technical parametres	Probe coverage Design 1	Probe coverage Design 2
Target Bases Covered	4,164,801	1,286,215
% Target Bases Covered	84.1	86.4
Target Bases Not Covered:	784,479	202,660
• **Due to Ns**	0	0
• **Due to Repeats %**	602,423	177,495
% Target Bases Not Covered:	15.9	13.6
• **Due to Ns**	0	0
• **Due to Repeats**	12.2	11.9
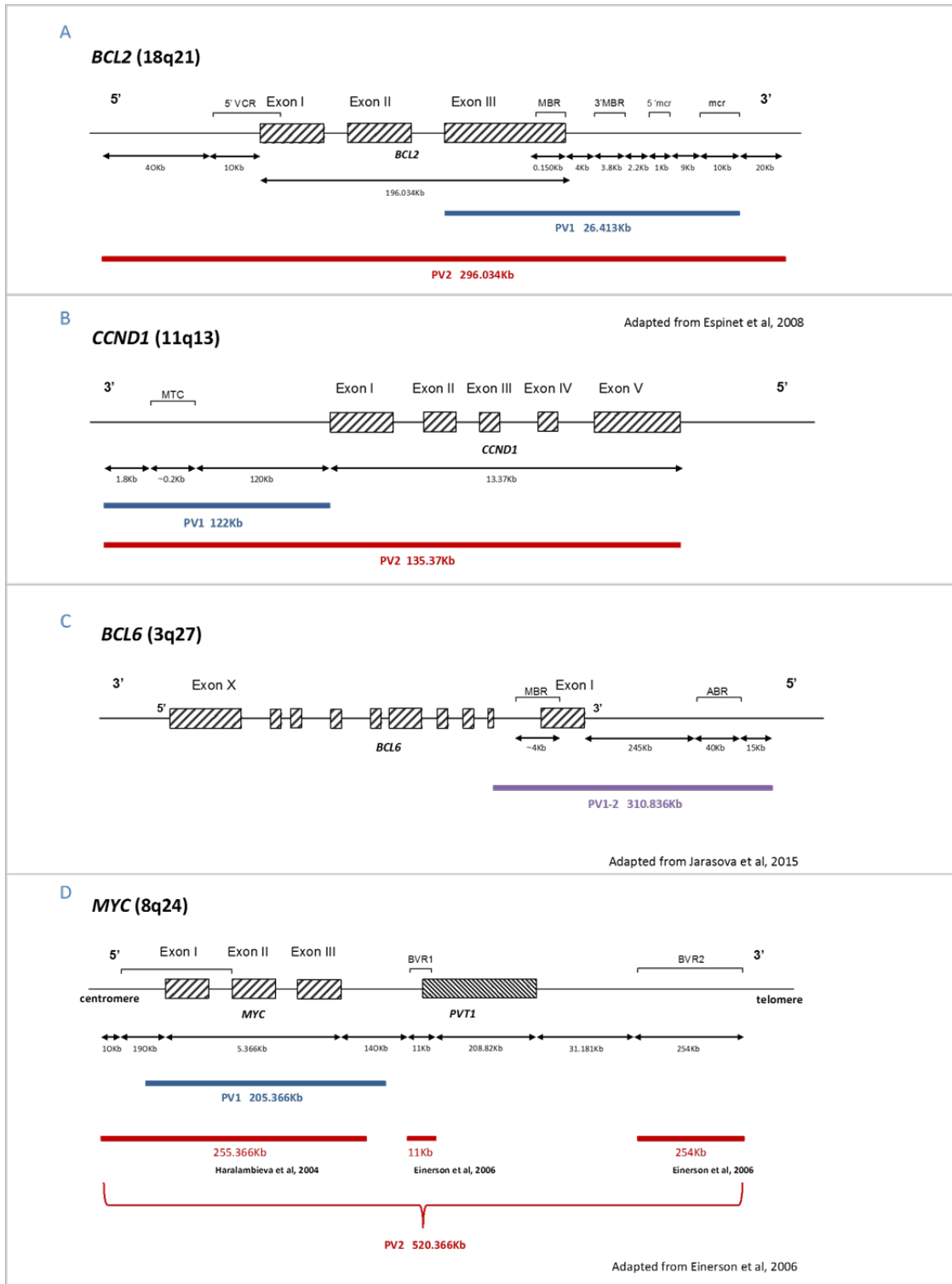
Figure 29. Detail of the region captured in both versions of the panel for each target rearrangement: *BCL2* **(A),** *CCND1* **(B),** *BCL6* **(C) and** *MYC* **(D) genes.** The regions covered by PV1 are shown with a blue bar and by PV2 with a red bar; if the design of PV1 and PV2 is the same, it is displayed with a purple bar.

114

3. Sample library preparation

The preparation of the library was carried out in the collaborating company qGenomics. For this end, KAPA HyperPrep Kit, with KAPA Library Amplification Primer Mix (10X) (Roche), for library preparation, and SeqCap Hybridization and Wash Kits (Roche), for the library hybridization, were applied.

The steps that were carried out for the preparation of the library are detailed below (Figure 30):
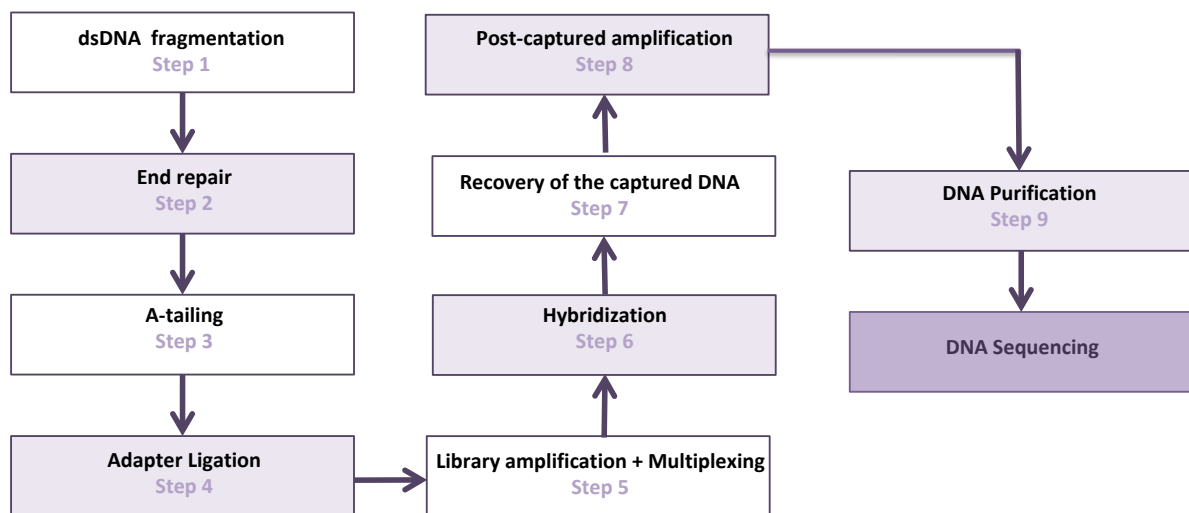
Figure 30. Library preparation workflow

3.1. Double strand (dsDNA) fragmentation:

1. One point one µg of the genomic DNA (gDNA) of interest were pipetted into a 1.5 mL tube.

2. The volume was adjusted to a total of 55 µL using 1x TE low EDTA and transferred to a Covaris microTUBE for fragmentation.

3. The gDNA was fragmented to obtain 200 bp DNA fragments.

3.2. End repair:

1. Fifty µL of the fragmented DNA were transferred to a 0.2 mL PCR tube.

2. To each 50 µL fragmented sample, 20 µL of End Repair Master Mix (Table 18) were added, resulting in a total volume of 70 µL.

Reagent	Volume (µL)
PCR-grade water	8 µL
10X KAPA End Repair Buffer	7 µL
KAPA End Repair Enzyme Mix	5 µL
Total volume	**20 µL**

Table 18. End Repair mix

3. The reaction was mixed by pipetting up and down and incubated at 20ºC for 30 min.

End repair cleanup:

4. To each 70 µL End Repair Reaction, 120 µL of room temperature Agencourt AMPure XP beads were added, resulting in a total volume of 190 µL (Table 19).

Reagent	Volume (µL)
End Repair reaction	70 µL
Agencourt AMPure XP beads	120 µL
Total volume	**190 µL**

Table 19. End Repair Cleanup mix

5. The reaction was mixed thoroughly by pipetting up and down multiple times and incubated at room temperature for 10-15 min to allow the DNA to bind to the beads.

6. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

7. The supernatant was carefully removed and discarded.

8. Keeping the tubes on the magnet, 200 µL of freshly-prepared 80% ethanol were added.

9. The tubes were incubated at room temperature for ≥30 sec.

10. The ethanol was carefully removed and discarded.

11. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

12. The tubes were then incubated at room temperature for ≥30 sec.

13. The ethanol was carefully removed and discarded.

14. The beads were dried at room temperature for 3-5 min, so that all the ethanol evaporated. After, the tubes were removed from the magnet.

3.3. A-tailing:

1. To each tube of DNA plus beads, 50 μL of the A-Tailing Master Mix (Table 20) were added, resulting in a total volume of 50 μL.

Reagent	Volume (μL)
PCR-grade water	42 μL
10X KAPA A-Tailing Buffer	5 μL
KAPA A-Tailing Enzyme	3 μL
Total volume	**50 μL**

Table 20. A-tailing mix

2. The beads were thoroughly resuspended by pipetting up and down multiple times, and incubated at 30ºC for 30 min.

A-tailing cleanup:

3. To each 50 μL of A-Tailing Reaction, 90 μL of thawed, room temperature PEG/NaCl SPRI Solution were added, resulting in a total volume of 140 μL (Table 21).

Reagent	Volume (µL)
A-Tailing Reaction	50 µL
PEG/NaCl SPRI Solution	90 µL
Total volume	**140 µL**

Table 21. A-tailing cleanup mix

4. The reaction was mixed thoroughly by pipetting up and down multiple times and incubated at room temperature for 10-15 min to allow the DNA to bind to the beads.

5. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

6. The supernatant was carefully removed and discarded.

7. Keeping the tubes on the magnet, 200 µL of freshly-prepared 80% ethanol were added.

8. The tubes were incubated at room temperature for ≥30 sec.

9. The ethanol was carefully removed and discarded.

10. Keeping the tubes on the magnet, 200 µL of freshly-prepared 80% ethanol were added.

11. The tubes were then incubated at room temperature for ≥30 sec.

12. The ethanol was carefully removed and discarded.

13. The beads were dried at room temperature for 3-5 min, so that all the ethanol evaporated. After, the tubes were removed from the magnet.

3.4. Adapter ligation:

1. To each tube of beads, 45 µL of the Ligation Master Mix (Table 22) were added, resulting in a total volume of 45 µL.

Reagent	Volume (µL)
PCR-grade water	30 µL
5X KAPA Ligation Buffer	10 µL
KAPA T4 DNA Ligase	5 µL
Total volume	**45 µL**

Table 22. Ligation mix

2. The beads were thoroughly resuspended by pipetting up and down multiple times.

3. Five µL of the SeqCap Library Adapter (with the desired Index for each sample) were added to the tube containing the Ligation Master Mix plus DNA and beads.

4. Then, the reaction was pipetted up and down 10 times and incubated at 20ºC for 15 min.

First post ligation cleanup:

5. To each 50 µL Ligation Reaction, 50 µL of thawed, room temperature PEG/NaCl SPRI Solution were added, resulting in a total volume of 100 µL (Table 23).

Reagent	Volume (µL)
Ligation Reaction	50 µL
PEG/NaCl SPRI Solution	50 µL
Total volume	**100 µL**

Table 23. Post ligation cleanup mix

6. The reaction was mixed thoroughly by pipetting up and down multiple times and incubated at room temperature for 10-15 min to allow the DNA to bind to the beads.

7. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

8. The supernatant was carefully removed and discarded.

9. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

10. The tubes were incubated at room temperature for ≥30 sec.

11. The ethanol was carefully removed and discarded.

12. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

13. The tubes were then incubated at room temperature for ≥30 sec.

14. The ethanol was carefully removed and discarded.

15. The beads were dried at room temperature for 3-5 min, so that all the ethanol evaporated. After, the tubes were removed from the magnet.

16. The beads were thoroughly resuspended in 50 μL of PCR-grade water and incubated at room temperature for 2 min to allow the DNA to elute off the beads.

Second post ligation cleanup:

17. To each 50 μL purified and adapter-ligated DNA with beads, 50 μL of thawed, room temperature PEG/NaCl SPRI Solution were added, resulting in a total volume of 100 μL (Table 24).

Reagent	Volume (μL)
Purified, adapter-ligated DNA with beads	50 μL
PEG/NaCl SPRI Solution	50 μL
Total volume	**100 μL**

Table 24. Post ligation cleanup mix

18. The reaction was mixed thoroughly by pipetting up and down multiple times and incubated at room temperature for 10-15 min to allow the DNA to bind to the beads.

19. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

20. The supernatant was carefully removed and discarded.

21. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

22. The tubes were incubated at room temperature for ≥30 sec.

23. The ethanol was carefully removed and discarded.

24. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

25. The tubes were then incubated at room temperature for ≥30 sec.

26. The ethanol was carefully removed and discarded.

27. The beads were dried at room temperature for 3-5 min, so that all the ethanol evaporated. After, the tubes were removed from the magnet.

28. The beads were thoroughly resuspended in 25 μL of PCR-grade water and incubated at room temperature for 2 min to allow the DNA to elute off the beads.

29. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

30. The supernatant was transferred to a new tube.

3.5. Library amplification:

1. In a new PCR tube, 30 μL of Pre-Capture LM-PCR Master Mix were added, resulting in a total volume of 30 μL per tube (Table 25).

Reagent	Volume (μL)
KAPA HiFi HotStart ReadyMix (2x)	25 μL
Pre LM-PCR Oligos 1 & 2, 5 μM*	5 μL
Total volume	**30 μL**

Table 25. Pre-Capture LM-PCR mix

2. Twenty μL of sample library (or PCR-grade water for negative control) were added to the PCR tube containing the LM-PCR Master Mix.

3. The reaction was thoroughly mixed by pipetting up and down five times (vortex was avoided) and centrifuged briefly.

4. Samples were then amplified using the following cycling protocol :

98ºC 45 sec

98ºC 15 sec

60ºC 30sec   x7

72ºC 30 sec

72ºC 1 min

Hold at 4ºC

Library amplification cleanup:

5. To each 50 μL library amplification reaction product, 50 μL of thoroughly vortexed Agencourt AMPure XP Beads were added, resulting in a total volume of 100 μL (Table 26).

Reagent	Volume (μL)
Library amplification reaction product	50 μL
Agencourt AMPure XP Beads	50 μL
Total volume	**100 μL**

Table 26. Library amplification cleanup mix

6. The reaction was mixed thoroughly by pipetting up and down multiple times and incubated at room temperature for 10-15 min to allow the DNA to bind to the beads.

7. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

8. The supernatant was carefully removed and discarded.

9. Keeping the tubes on the magnet, 200 μL of freshly-prepared 80% ethanol were added.

10. The tubes were incubated at room temperature for ≥30 sec.

11. The ethanol was carefully removed and discarded.

12. Keeping the tubes on the magnet, 200 µL of freshly-prepared 80% ethanol were added.

13. The tubes were then incubated at room temperature for ≥30 sec.

14. The ethanol was carefully removed and discarded.

15. The beads were dried at room temperature for 3-5 min, so that all the ethanol evaporated. After, the tubes were removed from the magnet.

16. The beads were thoroughly resuspended in 50 µL of PCR-grade water and incubated at room temperature for 3 min to allow the DNA to elute off the beads.

17. The tubes were placed on a magnet to capture the beads. Tubes were incubated until the liquid was clear.

18. The supernatant was transferred to a new tube.

19. The concentration, size distribution and quality of the amplified library was determined using The Qubit fluorometer (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and the TapeStation system (Agilent Technologies, Santa Clara, California, USA).

3.6. Hybridization of the amplified sample library:

Before starting:

1. A heat block was turned on to 95ºC.

2. The appropriate number of 4.5 µL SeqCap EZ probe pool aliquots (one per hybridization) were removed from -20ºC and brought to room temperature placed on ice.

3. The lyophilized oligo tubes were given a brief spin to allow the contents to pellet at the bottom of the tube.

4. One hundred twenty µL of cold PCR-grade water was added to the SeqCap HE Universal Oligo tube (1,000 µM final concentration)

5. Ten µL of cold PCR-grade water was added to each SeqCap HE Index Oligo tube (1,000 µM final concentration)

6. The primers plus PCR-grade water was vortexed for five seconds and spun down.

Multiplexing DNA sample library pool:

7. Each of the uniquely indexed amplified DNA sample libraries to be included in the multiplex capture experiment were thawed on ice.

8. Equal amounts (by mass) of each of these amplified DNA sample libraries were mixed together to obtain a single pool with a combined mass of at least 1.25 µg (this mixture is referred to as the 'Multiplex DNA Sample Library Pool').

9. The resuspended SeqCap HE Universal Oligo (1,000 µM) and each resuspended SeqCap HE Index oligo (1,000 µM) that matches a DNA Adapter Index included in the Multiplex DNA Sample Library Pool were thawed on ice.

10. The SeqCap HE Index oligos were mixed together so that the resulting Multiplex Hybridization Enhancing Oligo Pool contained, by mass, 50% SeqCap HE Universal Oligo and 50% of a mixture of the appropriate SeqCap HE Index oligos. The total combined mass of the Multiplex Hybridization Enhancing Oligo Pool should be 2,000 pmol, which is the amount required for a single Sequence Capture experiment.

Hybridization:

11. Five µL of COT Human DNA (1 mg/mL) were added to a new 1.5 mL tube.

12. One µg of Multiplex DNA Sample Library was added to the 1.5 mL tube containing 5 µL of COT Human DNA.

13. Two thousand pmol (or 2 µL) of the Multiplex Hybridization Enhancing Oligo Pool (1 µL of 1,000 pmol SeqCap HE Universal Oligo and 1 µL of the 1,000 pmol SeqCap HE Index Oligo pool) were added to the Multiplex DNA Sample Library Pool plus COT Human DNA, as detailed in Table 27.

Reagent	Amount	Volume (µL)
COT Human DNA	5 µg	5 µL
Multiplex DNA Sample Library Pool	1 µg	~50 µL
SeqCap HE Universal Oligo	1,000 pmol	1 µL
SeqCap HE Index Oligo pool	1,000 pmol	1 µL
Total volume		**~57** µL

Table 27. Sample preparation mix for hybridization

14. The tube's lid was closed and made a hole in the top of the tube's cap with an 18 to 20 gauge or smaller needle to suppress contamination in the DNA vacuum concentrator.

15. The Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool was dried in a DNA vacuum concentrator on high heat (60ºC).

16. To each dried-down Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool, the following was added (final volume: 10.5 µL):

- 7.5 µL of 2X Hybridization Buffer

- 3 µL of Hybridization Component A

17. The hole in the tube's cap was covered with a sticker.

18. The tubes were then vortexed for 10 sec and centrifuged at maximum speed for 10 sec.

19. Next, samples were placed in a 95ºC heat block for 10 min to denature the DNA and after, centrifuged at maximum speed for 10 sec at room temperature.

20. The solution was transferred to the 4.5 µL aliquot of SeqCap EZ probe pool, previously prepared, and vortexed for 3 sec.

21. Samples were then centrifuged at maximum speed for 10 sec and incubated in a thermocycler at 47ºC for 16-20 hours (the thermocycler's heated lid was turned on and set to maintain 47ºC, 10ºC above the hybridization temperature).

3.7. Washing and recovery of the captured DNA sample:

Buffer preparation:

1. Wash Buffers (I, II, III and Stringent) at 10X and 2.5X Bead Wash Buffer were

Reagent	Volume of concentrated Buffer (µL)	Volume of PCR-grade Water (µL)	Total Volume of 1X Buffer (µL)
10X Stringent Wash Buffer	40 µL	360 µL	**400 µL**
10X Wash Buffer I	30 µL	270 µL	**300 µL**
10X Wash Buffer II	20 µL	180 µL	**200 µL**
10X Wash Buffer III	20 µL	180 µL	**200 µL**
2.5X Bead Wash Buffer	200 µL	300 µL	**500 µL**

Table 28. Preparation of buffer dilutions

diluted as detailed in Table 28, to create 1X working solutions.

2. The following wash buffers were preheated to 47ºC in a water bath:

- 400 µL of 1X Stringent Wash Buffer

- 100 µL of 1X Wash Buffer I

Capture Beads preparation:

3. Capture Beads were brought to room temperature for 30 min prior to use.

4. The beads were thoroughly mixed by vortexing for 15 sec and 100 µL of beads for each capture were then aliquoted into a single 1.5 mL tube.

5. The tubes were place in a magnet and incubated till the liquid was clear.

6. The supernatant was carefully removed and discarded.

7. While the tubes were in the magnet, twice the initial volume of beads of 1X Bead Wash Buffer was added (200 μL for one capture).

8. The tubes were removed from the magnet and vortexed for 10 sec.

9. After, the tubes were placed back in the magnet to bind the beads, and incubated till the liquid was clear.

10. The supernatant was carefully removed and discarded.

11. The steps 7-10 were repeated for a total of two washes.

12. After, tubes were resuspended by vortexing the beads in 1x the original volume using the 1X Bead Wash Buffer (100 μL buffer for one capture).

13. One hundred μL of resuspended beads were transferred to a new 0.2 mL tube (one tube for each capture).

14. The tubes were placed in the magnet to bind the beads.

15. The supernatant was carefully removed and discarded. The Capture Beads were ready to bind the captured DNA.

DNA binding to Capture Beads:

16. The hybridization samples were transferred to the Capture Beads prepared in the previous step and mixed thoroughly by pipetting up and down ten times.

17. The tubes containing the beads and DNA were placed in a thermocycler set to 47ºC for 45 min (heated lid set to 57ºC) to bind the captured sample to the beads. Meanwhile, samples were mixed by vortexing for 3 sec at 15 min intervals to ensure that the beads remaineded in suspension.

Washing of DNA-bound Capture Beads:

18. After the 45-minute incubation, 100 μL of 1X Wash Buffer I preheated to 47ºC was added to the 15μL of Capture Beads Plus Bound DNA.

19. The solution was mixed by vortexing for 10 sec and the entire content was transferred to a 1.5 mL tube.

20. The tubes were placed in the magnet to bind the beads, and incubated till the liquid was clear.

21. The supernatant was carefully removed and discarded.

22. The tubes were removed from the magnet and 200 µL of 1X Stringent Wash Buffer preheated to 47ºC was added.

23. The solution was mixed by pipetting up and down ten times and incubated at 47ºC for 5 min.

24. The steps 20-23 were repeated for a total of two washes.

25. The tubes were placed in the magnet to bind the beads, and incubated till the liquid was clear.

26. The supernatant was carefully removed and discarded.

27. Two hundred µL of room temperature 1X Wash Buffer I was added and mixed by vortexing for 2 min.

28. The tubes were placed in the magnet to bind the beads, and incubated till the liquid was clear.

29. The supernatant was carefully removed and discarded.

30. Two hundred µL of room temperature 1X Wash Buffer II was added and mixed by vortexing for 1 min.

31. The tubes were placed in the magnet to bind the beads, and incubated till the liquid was clear.

32. The supernatant was carefully removed and discarded.

33. Two hundred µL of room temperature 1X Wash Buffer III was added and mixed by vortexing for 30 sec.

34. The tubes were placed in the magnet to bind the beads, and incubated till the liquid was clear.

35. The supernatant was carefully removed and discarded.

36. The tubes were removed from the magnet and 50 µL of PCR-grade water was added to each tube of bead-bound captured sample (there is no need to elute DNA off the beads).

3.8. Amplification of the captured DNA sample:

<u>**Before starting:**</u>

1. The lyophilized 'Post-LM-PCR Oligos 1 & 2' were briefly spun, to allow the contents to pellet at the bottom of the tube.

2. Four hundred and eighty µL of PCR-grade water was added to the tube of centrifuged oligos.

3. The resuspended oligos were then briefly vortexed and spun down to collect the contents.

<u>**Post-captured DNA amplification:**</u>

4. To each PCR tube (one pair per captured DNA sample) 30 µL of Post-Capture LM-PCR Master Mix was added, as detailed in Table 29, resulting in a total volume of 30 µL per tube, or 60 µL per DNA sample.

Reagent	Volume (µL)
KAPA HiFi HotStart ReadyMix	25 µL
Post-LM-PCR Oligos 1 & 2, 5 µM	5 µL
Total volume	**30 µL**

Table 29. Post-Capture LM-PCR Master Mix

5. The bead-bound captured DNA was vortexed to ensure a homogenous mixture of beads.

6. Twenty µL of bead-bound captured DNA was transferred as template into each of the two PCR tubes and mixed well by pipetting up and down.

7. Twenty µL of PCR-grade water was added to the negative control and mixed well by pipetting up and down five times.

8. The PCR tubes were placed in the thermocycler, and the captured DNA was amplified using the following program:

98ºC 45 sec
98ºC 15 sec ⎤
60ºC 30 sec ⎬ x14
72ºC 30 sec ⎦
72ºC 1 min
Hold at 4ºC (up to 72h)

3.9. Purification of the captured and amplified DNA sample:

1. The Agencourt AMPure XP Beads were brought to room temperature for at least 30 min before use.

2. The similar amplified captured Multiplex DNA Sample Libraries were pooled in a 1.5 mL microcentrifuge tube (approximately 100 µL). The negative control was processed in exactly the same way as the library of amplified samples.

3. The beads were vortexed for 10 sec before use to ensure a homogenous mixture of beads.

4. One hundred and eighty µL of Agencourt AMPure XP Beads was added to the 100 µL pooled amplified captured Multiplex DNA Sample library. Then, tubes were vortexed briefly.

5. After, samples were incubated at room temperature for 15 min to allow the DNA to bind to the beads.

6. The tubes containing the bead bound DNA were then placed in a magnetic particle collector, and incubated till the liquid was clear.

7. The supernatant was carefully removed and discarded.

8. Keeping the tubes in the magnet, 200 µL of freshly-prepared 80% ethanol was added, and incubated at room temperature for 30 sec.

9. The ethanol was carefully removed and discarded.

10. The step 9 was repeated for a total of two washes with 80% ethanol.

11. Following the second wash, all the ethanol was carefully removed and discarded.

12. The beads were dried at room temperature with the tube lid open for 30 min.

13. The tubes were removed from the magnetic particle collector.

14. The DNA was resuspended using 52µL of PCR-grade water, and pipetted up and down ten times to ensure that all beads were resuspended.

15. Samples were incubated at room temperature for 2 min.

16. After, the tubes were placed back in the magnetic particle collector, and incubated till the liquid was clear.

17. Fifty µL of the supernatant that now contains the amplified captured Multiplex DNA Sample Library Pool was transferred to a new 1.5 mL tube.

18. The concentration, size distribution and quality of the amplified captured multiplex DNA sample was determined using The Qubit fluorometer (Invitrogen) and the TapeStation system (Agilent). The yield of final, amplified library should be >500 ng, and the library fragment lenght distribution should be 150-500 bp.

4. **Sample sequencing**

The sequencing of the library was carried out in the Genetics unit of the Pompeu Fabra University (Barcelona) and in the collaborating company qGenomics. All runs were sequenced using the NextSeq 500/550 Mid-Output v2.5 Kit (300 cycles) (Illumina, San Diego, Califonia, USA). The detail of the distribution of the samples per run is shown in Table 30.

Table 30. Distribution of samples per run in each version of the panel

	Run	Total samples sequenced per run	Patient samples	Control samples	Samples removed from the analysis	Samples sequenced but not included in this project
PV1	Run1	30	30	-	-	-
	Run2	20	19	-	1	-
	Run3	25	21	3	1	-
	Run4	25	23	2	-	-
	Run5	25	15	10	-	-
PV2	Run1	40	31	6	3	-
	Run2	40	23	5	11	1
	Run3	38	16	10	-	12

5. Bioinformatic analysis

The bioinformatic analysis of the data was carried out by the collaborating company qGenomics, MARGenomics (IMIM) and DREAMgenics (Oviedo), and included the following steps:

5.1. Preliminary quality control: A quality control of raw sequence data was performed to obtain an overview of the quality of the sequencing of the sample. This step is independent of the quality criteria shown in the sequencer, since there may be runs whose quality parameters are very good, but the sample has not been sequenced with the expected quality. This step was carried out using the FastQC tool.

5.2. Preprocessing: This step consisted of removing low quality portions while preserving the longest high quality portion of an NGS read (known as trimming) in addition to discarding those reads that did not have a minimum average quality of their bases. This process was carried out with the programs Trimmomatic, Trim Galore, QcReads, and SeqtrimNEXT. After performing this process, the quality of the sequencing was re-analyzed with FastQC to verify that the quality of the sequence had increased to the desirable parameters.

5.3. Alignment: The set of good quality reads were aligned against the reference genome for which the Genome Reference Consortium Human Build 37 (GRCh37 or hg19) was used. This step consisted of searching within the reference genome the position that best fit each of the reads saved in the fastq file. To carry out this step, the Burrows-Wheeler Alignment tool (BWA) was used, which allows an alignment with a sensitivity greater than 90-95% (Li & Durbin, 2009).

Once the reads were aligned, they were sorted by genomic coordinate using the SortSam tool (Picard).

The alignment ended with the creation of an index file for the bam file (which usually has a .bai extension), which allowed the programs that were executed in the following steps to access the information contained within the sam/bam files more quickly. For this end, the Picard's BuildBamIndex program was applied.

5.4. Post alignment: This step included marking of PCR duplicates, realignment of regions with inserts and/or deletions, and base quality recalibration.

o To remove duplicates generated during PCR, the MarkDuplicates tool (Picard) was used. Although duplicate marking is not recommended for amplicon panels, since, due to the high coverage that samples usually achieve, there is a high probability that the program considers false duplicate reads, for capture panels it is considered convenient to do it.

o The realignment of indel was performed with the RealignerTargetCreator and IndelRealigner (GATK) programs and allowed to correct errors created by the aligners, resulting in more consistent reads in regions with insertions or deletions.

o Finally, the base quality recalibration was carried out with the BaseRecalibrator and PrintReads (GATK) tools and allowed to correct sequencing errors and other experimental artifacts.

5.5. Variant calling: In this step, the reads aligned against the reference genome were compared, looking for discrepancies between the reads and the reference genome, known as mismatch. In those cases where enough reads were found with the same mismatch, to rule out a sequencing or alignment error, the mismatch was reported as a variant. The output file of the variant callers is usually a vcf file, which contains all the discrepancies between the sample reads and the reference genome, in addition to the variant allele frequency (VAF), which corresponds to the percentage in which the variant appears with respect to the total coverage of that base.

Different variant callers were applied for the analysis of the different mismatches:

5.5.1. Variant callers for single nucleotide variants and small indel:

Mutect2 was applied, which is a somatic variant caller that uses local assembly and realignment to detect SNV and indel. Besides, it is based on several probabilistic models for genotyping and filtering that work well with and without a matched normal sample and for all sequencing depths (Benjamin et al., 2019).

In addition, in order to annotate the genetic variants detected with Mutect2, ANNOVAR software was used, which provided all the information regarding the detected variants (in a single text file separated in columns) that was essential for subsequent filtering of the variants.

5.5.2. Variant callers for copy number alterations:

CopywriteR was applied, which exploits the off-target sequence reads from targeted sequencing to extract DNA copy number profiles. This method is based on peak calling using Model-based Analysis for ChIPseq (MACS) algorithm in a matched reference sample or, when no reference is available, in the sample itself (Kuilman et al., 2015). It uses the standard circular binary segmentation method (CBS) to identify changes in copy number (Olshen et al, 2004). CopywriteR is an R package available at bioconductor (Huber et al., 2015) or GitHub (https://github.com/PeeperLab/CopywriteR).

CopywriteR exploits off-target reads from targeted sequencing for CNA detection. Thus, the sequence reads in peaks were removed, and the depth of coverage (DOC) was calculated based on fixed-size windows (bins) of 100kb for PV1, and 200kb and 500kb for PV2. DOC was compensated for peak removal, normalized using loess-based corrections for mappability and GC content, and filtered for regions of extensive germline copy number variation. Then, to call the variants, log2 ratio values were stablished at >2.2 for gains and <1.8 for losses, and those alterations presented in at least two controls were removed. After, the detected alterations were intersected with the alterations previously identified by the routine techniques and the alterations in the regions defined as targets were identified. Finally, target and non-target alterations were divided into two tables, and an Excel file was generated with the variants to be analyzed.

On the other hand, since the program only analyzes the off-target sequences, in PV1 an additional analysis was performed to evaluate the need to include the SNP distributed throughout the genome in the design. For this end, the reads corresponding to SNP regions were bioinformatically removed from the sequencing data.

5.5.3. Variant callers for rearrangements:

Due to the technical challenge of detecting SV (since SV generate multiple alignment signals, including altered sequence coverage within duplications or deletions (read-depth), breakpoint-spanning paired-end reads that align discordantly relative to each other (read-pair), and breakpoint-containing single reads that align in split fashion to discontiguous loci in the reference genome (split-read)), three algorithms were used to call the rearrangements:

- o **LUMPY** analyses both paired-end and split-read data separately and integrates the results along with other posible evidence source, such as read-depth, to generate a set of SV predictions (Layer et al., 2014).

o **Smoove** wraps existing software (LUMPY) and adds some internal read-filtering, to simplify calling and genotyping SV. It excludes:

- Reads where both ends are soft-clips

- Interchromosomal reads with >3 mismatches

- Interchromosomal reads with alternative matches

- Interchromosomal split-reads (unless the split and the mate go to the same general location)

- Reads in exclude regions or chromosomes

- Reads from regions where the depths of split or discordant reads is greater than 1000x, removing regions that contribute to spurious calls)

- Reads that are orphaned (do not have a mate) after all of the above filtering

After applying these filters, up to 80% of the reads are removed from the bam file that is sent to LUMPY.

o **DELLY** also exploits multiple SV detection signals, but first uses one signal (that is, read-pair) to drive discovery and then refines and/or genotypes candidates with a second signal (that is, split-read or read-depth) (Rausch et al., 2012).

o Due to the complexity of the rearrangement analysis, .bam and .bai files corresponding to 92 lymphoma samples with possible translocations (Table 31) were processed in parallel by the company **DREAMgenics** for the analysis of rearrangements. These data were used to validate our rearrangement pipeline. The company detected structural variants by the analysis of paired and split-reads using LUMPY framework (Layer et al., 2014).

Table 31. Samples included in DREAMgenics analysis for each panel version.

PV1	PV2
24 DLBCL	34 DLBCL
11 FL	6 FL
6 MCL	6 MCL
5 BL	-
Total: 46	Total: 46

6. Variant analysis

6.1. Single nucleotide variants and small indel variants analysis

6.1.1. Single nucleotide variants and small indel variants filtering

The list of variants provided by the analysis pipeline is quite extensive due to the detection of polymorphisms, sequencing errors, and variants that are not located within the coding area of the gene, among others. Therefore, before beginning to interpret the candidate variants, those variants that were not going to be interpreted or reported were filtered, following the process detailed below.

It should be noted that the variant filtering criteria are under continuous evaluation, since, depending on the advances made in the field of next generation sequencing, it may be convenient to modify the way of evaluating the variants.

- o **Automatic variant filtering**

This process was carried out almost automatically and consisted of eliminating the polymorphic variants as well as those variants outside the coding regions, in addition to the variants in UTR regions as well as synonymous variants.

- • **Polymorphic variants filtering:** MAF (minor allele frequency) refers to the frequency in which a variant is described in the population. Variants with a total MAF (or MAF in an ethnic subpopulation) greater than 1% are

currently considered polymorphisms (SNP). These variants were not categorized because they do not have clinical utility, based on current knowledge, and may hinder the interpretation.

- **Filtering of variants in non-coding and UTR regions and synonymous variants:** In this step, variants in coding (exonic) and splicing regions (minimum -5 to +5) were preserved, and those variants located in intergenic sites, downstream, upstream, in non-coding RNA and in intronic regions far from splicing sites were removed.

On the other hand, synonymous variants, which are those that do not generate any change in the coding protein, as well as variants found in 3'UTR and 5'UTR regions were filtered, except for those genes with variants described in these regions, *NOTCH1* and *CCND1,* respectively.

- **Variant pathogenicity filtering**

In parallel to the process detailed above, variants described as pathogenic or likely pathogenic by the ClinVar database were independently selected to avoid filtering probably pathogenic variants with a MAF >1%.

- **Manual variant filtering**

Manual filtering of variants allowed discarding sequencing errors that were not previously detected. To perform this filtering, the Integrative Genomics Viewer (IGV) was used, which is available for free and allows to open alignment files (bam or sam), and to visualize and analyze them in a simple way.

All the variants that were not discarded in the automatic filter were analyzed one by one with the IGV, which allowed removing the following sequencing errors:

- Variants in non-uniform coverage regions: Variants located in non-uniform coverage regions, range of +/- 15 bases around the variant studied.

- Variants detected in a single fragment: Variants that were not sequenced by more than one non-identical read.

- Variants with strand bias: Variants with an unbalanced proportion of reads in the forward and reverse chains between the mutated and the reference reads.

- Variants in repetitive regions: Variants in repetitive and homopolymer regions, when the quality was poor and the MAF ≤10%.

6.1.2. Single nucleotide variants and small indel variants categorization and classification

Variants were classified based on the following criteria:

I. Recurrent presence in databases or in the literature, indicating the pathogenicity or recurrence of the identified variant: Variants described by at least two different authors were considered recurrently associated with the pathology under study.

II. Pathology or tissue in which the variant has been described.

III. Variant effect on the protein: Variants not described in the literature were categorized as likely pathogenic, of uncertain significance or likely benign, depending on the effect of the variant on the protein.

IV. Functional studies and computational in silico predictive programs: Functional studies and in silico tools can be powerful in supporting pathogenicity. However, not all functional studies are effective in predicting an impact on the function of a gene or protein and should be carefully considered. On the other hand, the algorithms used by each in silico tool may differ and can include determination of the effect of the sequence variant at the nucleotide and amino acid level including determination of the effect of the variant on the primary and alternative gene transcripts, other genomic elements, as well as the potential impact of the variant on the protein. However, these are only predictions, and their interpretation was used only to support previous criteria, but never as a determining criterion for the classification of variants.

Following these criteria, variants were classified using a five-level classification system proposed by the guidelines published by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology in 2015 (Richards et al., 2015) (Table 32):

- **Class 1 or pathogenic variants:** This classification included variants established as pathogenic in the disease of interest by clinical databases.

- **Class 2 or likely pathogenic variants:** Comprised variants established as likely pathogenic in the disease of interest, or pathogenic/likely pathogenic in other diseases that may or may not be related to the disease of interest by clinical databases.

  This classification also included variants described as pathogenic or likely pathogenic in functional studies, in addition to undescribed variants that cause an effect on the protein (such as, nonsense, stop gain o loss, frameshift and splicing variants).

- **Class 3 or variants of uncertain significance (VUS):** This classification comprised undescribed variants that do not cause any effect on the protein (missense) as well as in-frame variants.

- **Class 4 or likely benign variants:** Included variants established as likely benign in the disease of interest, or benign/likely benign in other diseases that may or may not be related to the disease of interest by clinical databases.

  This classification also included variants described as benign or likely benign in functional studies.

- **Class 5 or benign variants:** This classification included variants established as benign in the disease of interest by clinical databases.

To carry out the classification of the variants, ClinVar disease database and the International Agency for Research on Cancer (IARC) database, which is specific for

TP53 variants, were mainly used, since they are a reliable resource that supports the interpretation of variants (Richards et al., 2015; Yang et al., 2017).

In addition, the Varsome tool was used, which is a search engine, aggregator and impact analysis tool which integrates all the information related to the variant, such as the coding effect for different transcripts, the genomic location and neighboring variants, the genes that may affect, the population frequency, the function of the associated protein, relevant phenotypes, related literature, clinical studies and the predicted pathogenicity, as well as the experience of experts in the classification and interpretation of the variants. Moreover, variant pathogenicity is also reported using an automatic variant classifier that evaluates the submitted alteration according to the ACMG guidelines (Richards et al., 2015). Figure 31 shows a squeme of the filtering process followed for SNV and indel variants.
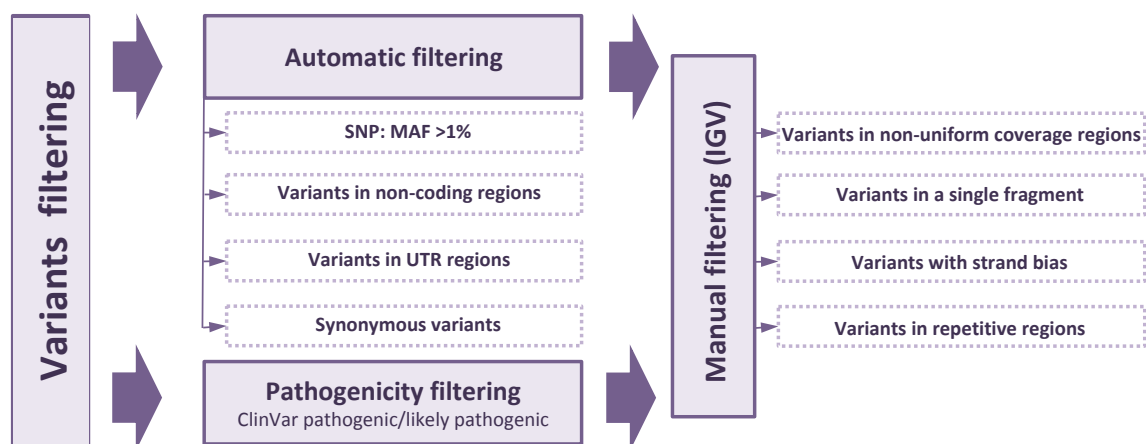
Figure 31. Squeme of the filtering process followed for SNV and indel variants.

Table 32. Single nucleotide variants and small indel variants classification.

Criteria	Criteria I Recurring variant in databases	Criteria II Tissue or pathology where the variant has been described	Criteria III Variant effect on the protein	Criteria IV Functional studies
Class 1 **Pathogenic**	Variant described as pathogenic	In the disease or tissue under study	n/a	n/a
Class 2 **Likely pathogenic**	Variant described as likely pathogenic	In the disease or tissue under study	n/a	n/a
	Variant described as pathogenic/likely pathogenic	In another disease or tissue	n/a	n/a
	Variant not described	Variant not described	Frameshift, nonsense, stop loss/gain, or splicing variant	Variant not described
	Variant not described	Variant not described	n/a	Pathogenic/likely pathogenic effect in functional studies
Class 3 **VUS**	Variant not described	Variant not described	Missense or in-frame variant	Variant not described
Class 4 **Likely benign**	Variant described as likely benign	In the disease or tissue under study	n/a	n/a
	Variant described as benign/likely benign	In another disease or tissue	n/a	n/a
	Variant not described	Variant not described	n/a	Benign/likely benign effect in functional studies
Class 5 **Benign**	Variant described as pathogenic	In the disease or tissue under study	n/a	n/a

(Adapted from Richards et al., 2015)

6.2. Copy number variants filtering

For each PV, the data generated by CopywriteR was analyzed and compared with previously known information to establish additional criteria to identify the definitive target CNAs:

- o A minimum of 5 altered markers were established to be considered as CN alteration

- o Minimum distance to consider altered segments as independent: close alterations at 100 Kb were merged

- o Log2ratio filters based on the false positive and negative detected in each PV: > 2.2 for gains and < 1.8 for losses in PV1; > 2.3 for gains and < 1.7 for losses in PV2.

- o Minimum region affected to call the variants in each target alteration (Table 33).

Table 33. Minimum region affected to call the variants in each target alteration

	Region included in the panel design			Minimum criteria to consider a target region			
TARGET ID	**Cytoband**	**Genomic coordenates**	**Size (Kb)**	**Cytoban**	**Genomic coordenates**	**Size (Kb)**	**Criteria**
Loss 1p36	1p36.33-p36.11	1-28000000	28.000	1p36.33-p36.11	1-28000000	28.000	It has to include: 1p36 cytoband
Gain 2p24-2p12	2p24-p12	16080561-61155306	45.075	2p24-p12	16080561-61155306	45.075	It has to include: one of the genes relevant in these region (*REL, BCL11A, MYCN,* and *MSH2*)
Gain chr 3/3q	3p26.3-q29	1-198022430	198.022	3q11.2-q29	98000000-198022430	132.578	It has to include: 3q
Gain chr 5/5q	5p15.33-q35.3	1-180915260	180.915	5p15.33-q35.3	1-180915260	180.915	It has to include: 5p15.33-q35.3
Loss 6q13-q26	6q13-q26	70000001-164500000	94.500	6q13-q26	70000001-164500000	94.500	It has to include: one of the genes relevant in these region (*PRDM1, TNFAIP3/A20, EPHA7,* and *PERP*)
Loss 7q22.1-q35	7q22.1-q35	98000001-147900000	49.900	7q32.1-q32.3	127100001-132600000	22.009	It has to include: 7q32
Loss 8p21.1-p23.1	8p23.1-p21.1	6200001-28800000	22.600	8p23.1-p21.1	6200001-28800000	22.600	It has to include: one of the genes relevant in these region (*TNFRSF10A, TNFRSF10B, KCTD9*, and *BIN3*)
Gain 8q24 (*MYC*)	8q24.21-q24.21	128748315-128753680	5	8q24.21-q24.21	128748315-128753680	5	It has to include: *MYC*

	Region included in the panel design			Minimum criteria to consider a target region			
TARGET ID	Cytoband	Genomic coordenates	Size (Kb)	Cytoban	Genomic coordenates	Size (Kb)	Criteria
Loss 9p21	9p21.3-p21.1	19900001-33200000	13.300	9p21.3-p21.3	21802636-22009312	207	It has to include: one of the genes relevant in these region (*CDKN2A,CDKN2B, MTAP*)
Loss 11q22.1-q24	11q22.1-q24.3	97200001-130800000	33.600	11q22.2-q22.3	102188181-108239826	6.052	It has to include: ATM and/or BIRC3
Gain chr12	12p13.33-q24.33	1-133851895	133.852	12q13.11-q21.33	46400001-92600000	46.200	It has to include: 12q13-q21 If it is not for CLL, it can be considered target even if it does not include the centromere
Loss 13q14	13q14.11-q14.3	40100001-55300000	15.200	13q14.2-q14.2	50500000-51000000	500	It has to include: miR-15a y miR16-1, and DLEU2
Loss 16p13 (*CREBBP*)	16p13.3-p13.3	3775056-3930121	155	16p13.3-p13.3	3775056-3930121	155	Not included in PV2 for CNA analysis
Loss 17p13 (*TP53*)	17p13.1-p13.1	7571720-7590868	19	17p13.1-p13.1	7571720-7590868	19	It has to include: *TP53*
Gain chr18	18p11.32-q23	1-78077248	78.077	18q11.2-q23	19000001-78077248	59.077	It has to include: 18q
Loss 19p13 (*KLF2*)	19p13.11-p13.11	16435651-16438339	3	19p13.11-p13.11	16435651-16438339	3	It has to include: *KLF2* Target for SMZL

6.3. Rearrangement variants filtering

The output generated by the different callers, in addition to the output provided by the company DREAMgenics, were intersected with the alterations previously identified by the routine techniques and the alterations in the regions defined as targets were identified. Target and non-target alterations were then divided into two tables, and an Excel file was generated with the variants to be analyzed. For each PV, the data generated by each caller was independently analyzed and compared with previously known information to establish additional criteria to identify the definitive target rearrangements:

- Minimum distance between breakpoint location to consider rearrangements as independent: breakpoints located <5kb away from each other in the same rearrangement were considered duplicates and were removed from the analysis.

- Non-specific rearrangements were discarded. These included: rearrangements with frequent rearrangement partners, which appeared repeated between the different patients and target genes, and, on the other hand, rearrangements with specific partners that showed low quality values [genotype quality (GQ) <10000 by DELLY; quality=0.0 by DREAMgenics]. However, the latter could not be filtered directly and visualization of IGV was needed to refine the rearrangements that would be considered.

The output generated from each caller was integrated with the others after their analysis. Only those translocations involving the defined rearrangement target regions were analyzed. Lastly, an analysis was performed comparing the sensitivity and specificity of each caller in the detection of rearrangements.

7. Additional techniques applied for the validation of results

7.1. Fluorescent In Situ Hybridization (FISH) on isolated nuclei

The FISH on isolated nuclei was performed in 20 samples to validate the rearrangements and copy number alterations detected by NGS. The technique was carried out in the Molecular Cytogenetics Laboratory of the Pathology Service from Hospital del Mar, Barcelona.

For this end, samples fixed in carnoy were used, mainly from cytogenetics culture samples (from PB, BM, or lymph node fixed samples) or from fixed PB samples.

7.1.1. Preparation of the extension

1. The tube was centrifugated at 2000 rpm for 8-10 min to concentrate the fixed material and obtain a cellular botton.

2. The slides stored in a box with methanol at 4ºC were dried with hand paper, removing the excess methanol.

Fixed material extensions were performed:

3. The supernatant was decanted and the cell button was resuspended with a few drops of carnoy's solution.

4. The slides were identified with the case number and the probe to hybridize.

5. Extensions were made and dried on the hot plate (50-60ºC).

6. Nuclei concentration and extension quality were assessed using an inverted phase contrast microscope.

7. The area on the slide where the probe solution would be applied was marked with a diamond-tip pencil.

8. Extensions were dried on a hot plate for 4 hours (minimum) or at room temperature for one day.

7.1.2. Hybridization

1. The FISH probes were prepared following manufacturer's protocol:

- <u>Undiluted probes:</u> In an eppendorf 3.5 µL of hybridization buffer + 0.5 µL of probe + 1 µL of distilled H_2O were mixed for each slide to hybridize (these volumes were modified depending on the area to hybridize).

- <u>Probes diluted with hybridization buffer (direct probes):</u> 3-12 µL probe were applied, depending on the selected hybridization area.

2. Sufficient amount of probe was placed to cover the sample on the coverslip (18x18mm, 20x20mm, or 24x24mm).

3. The slide was placed on the coverslip with the probe, making sure that the probe covered the entire area to hybridize

4. The joint was sealed with glue (Fixogum Rubber Cement).

5. The samples were placed in the hybridizer and the following program was established:

   75ºC 5 min

   Hold at 37ºC (maximum 24h)

7.1.3. Post hybridization washes and counterstain

- <u>Post hybridization washes:</u>

1. A coplin was heated with the washing solution (0.4x SSC + 0.3% NP-40) to 73ºC.

2. Another coplin with washing solution (2x SSC + 0.1% NP-40) was prepared and left at room temperature.

3. The preparations were removed from the hybridizer

4. The Fixogum Rubber Cement glue was removed with the tweezers.

5. The preparations were immersed in the washing solution at 73ºC and the coverslip was removed. They were submerged again and incubated for 2 min.

6. Then preparations were immersed in the coplin at room temperature and incubate for 2 min.

- Counterstain:

7. A total of 7 µl of DAPI II Counterstain (Vysis, Downers Grove, Illinois, USA) were placed on a 24x24mm coverslip.

8. The preparations were removed from the washing solution and placed on the coverslip with the DAPI solution on the marked area. The DAPI was gently pressed to spread.

9. The preparations were stored in an opaque plastic box and upright at -20ºC until reading with the fluorescence microscope. The reading was made before 15 days after hybridization to avoid loss of fluorescence intensity.

7.2. Fluorescent In Situ Hybridization (FISH) on FFPE tissue samples

The FISH on FFPE tissues was performed in 22 samples to validate the rearrangements and CNA detected by NGS. It was carried out in the Molecular Cytogenetics Laboratory of the Pathology Service from Hospital del Mar, Barcelona.

7.2.1. Deparaffinization

1. The preparations were heated for 30 min (minimum) on the stove at 65ºC.

2. The slides were immersed in xylol for 10 min. It was repeated three times in different coplins.

3. The slides were then immersed in 100% ethanol for 5 min. It was repeated two times in different coplins.

4. Next, the slides were immersed in 80% ethanol for 5 min.

5. The slides were immersed in 70% ethanol for 5 min.

6. After, the slides were immersed in a coplin with distilled water.

7.2.2. Pretreatment

1. The slides were immersed in a plastic bucket with 1x EDTA solution.

2. Another plastic bucket with water was prepared (to counterbalance in the microwave).

3. The two buckets were heated in the microwave at maximum power for 7 min.

4. The bucket was left at room temperature for 30 min to cool down the EDTA solution.

5. The two buckets were heated again in the microwave at maximum power for 5 min.

6. Steps 4 and 5 were repeated and then the slides were removed from the bucket to dry.

7.2.3. Digestion

1. The pepsin solution (1O mL distilled H_2O + 100 μL 37% hydrochloric acid + 5 μL pepsin) and the "STOP" solution (50 mL 1x PBS + 2.5 mL $MgCl_2$) were prepared.

2. One-two drops of pepsin solution were spread with a plastic Pasteur pipette on 24x40mm coverslips.

3. The coverslips were placed on the slides.

4. They were incubated in a humid chamber at 37ºC for 15-30 min depending on the size and the type of FFPE tissue.

5. The slides were immersed in the "STOP" solution during 5 min.

6. The slides were then immersed in a coplin with distilled water during some sec.

7. The preparations were placed on the stove at 65ºC till they were dried.

8. The degree of digestion was assessed using an inverted phase contrast microscope, and the area to be hybridized was marked with a diamond-tip pencil when necessary.

9. In cases where it was not enough, digestion was repeated, adjusting the time depending on the case.

7.2.4. Hybridization

The same protocol as for FISH on isolated nuclei was followed (**8.1.3. Hybridization**), only the temperature and denaturation time changed:

80ºC 5 min

Hold at 37ºC (maximum 24h)

7.2.5. Post hybridization washes and counterstain

The same protocol as for FISH on isolated nuclei was followed (**7.1.4. Post hybridization washes and counterstain).**

7.3. SNP array

The SNP array technique was performed in 14 samples to validate the CNA detected by NGS. It was carried out in the Laboratoire d'Oncogénomique du Centre Hospitalier Universitaire Vaudois (CHUV) in Lausanne, Switzerland, with the CytoScan HD commercial platform (Affymetrix Inc., Santa Clara, California, USA). This platform comprises a microarray that contains 2.67 million markers for copy number analysis, including 750,000 SNP and 1.9 million non-polymorphic probes.

The SNP array technique consists of a process of DNA digestion with the Nsp1 restriction enzyme to generate a range of fragment sizes, which are then ligated with a common set of Nsp 1 adaptors. A PCR amplification is used to amplify fragments in the 25-125bp range, followed by purification, fragmentation, labeling, and hybridization on the array.

The technical details of each step are specified below.

7.3.1. Enzymatic digestion of DNA

1. Before starting, the thermocycler was turned on and ice was taken in a container, with a metal support inside.

2. All the reagents were brought to room temperature. When thawed, they were given a quick vortex and spin and placed directly on the ice.

3. DNA samples were brought to room temperature. When thawed, they were given little taps and a quick spin and placed directly on the ice.

4. For each sample, up to 250 ng of DNA was diluted in a total volume of 5 μL with low EDTA TE in a new 1.5 mL tube. The tubes were then covered with film, given a quick vortex and spin and placed directly on the ice.

5. For each case, a digestion mix as detailed in Table 34 was prepared:

Reagent	Volume (μL)
Nuclease-Free water	11.55 μL
Nsp I buffer	2 μL
100x BSA	0.20 μL
NSP I enzyme	1 μL
Total volume	19.75 μL

Table 34. Digestion mix

6. The tubes were then covered with parafilm, and given a quick vortex and spin.

7. Samples were subjected to thermal cycling, as follows:

37ºC 2 h

65ºC 20 min

Hold at 4ºC

8. Samples were introduced in the thermocycler when temperature reached 37ºC.

7.3.2. Adaptor ligation

1. Before starting, all the reagents were brought to room temperature. When thawed, they were given a quick vortex and spin, and placed directly on the ice.

2. The digestion products were taken out of the thermocycler, when it had finished, and were given a spin and carefully uncovered to prevent the sample from dropping out.

3. Then, to each digestion product, a ligation mix as detailed in Table 35 was prepared:

Reagent	Volume (µL)
DNA ligase buffer	2.5 µL
Adaptor Nsp I	0.75 µL
DNA ligase enzyme	2 µL
Total volume	25 µL

Table 35. Ligation mix

4. The tubes were then covered with film, and given a quick vortex and spin.

5. Samples were subjected to thermal cycling, as follows:

    16ºC 3 h

    70ºC 20 min

    Hold at 4ºC

6. After the program was finished, samples were stored at -20ºC (they can also be left in the thermocycler at 4ºC overnight).

7.3.3. PCR amplification

1. Before starting the thermocycler was turned on and ice was taken in a container, with a metal support inside.

2. All the reagents were brought to room temperature. When thawed, they were given a quick vortex and spin and placed directly on the ice.

3. Ligated DNA samples were brought to room temperature. When thawed, they were given a quick vortex and spin and placed directly on the ice.

4. Ligated DNA samples (25 µL) were diluted with 75 µL nuclease-free water (final volumen: 100 µL), sealed with parafilm and given a quick vortex and spin.

5. The film was carefully removed and the diluted DNA was transferred into 4 mini eppendorf (10µL/epp). The remaining diluted DNA and the ligated DNA samples (undiluted) were stored at -20ºC.

6. Then, to each sample, a PCR mix as detailed in Table 36 was prepared:

Reagent	Volume (µL)
Nuclease-Free water	39.5 µL
10xTitanium ^{Tm} Taq PCR buffer	10 µL
GC-Melt reagent	20 µL
2.5Mm dNTP mixture	14 µL
PCR primer	4.5 µL
50xTitanium ^{Tm} Taq DNA Polymerase enzyme	2 µL
Total volume	90 µL

Table 36. PCR mix

7. Mini eppendorf were sealed with parafilm and given a quick vortex and spin.

8. Samples were subjected to thermal cycling, as follows:

94ºC 3 min

94ºC 30 sec
60ºC 45 sec   x30
68ºC 15 sec

68ºC 7 min

Hold at 4ºC

9. Samples were introduced in the thermocycler when temperature reached 94ºC.

10. After the program was finished, samples were stored at -20ºC (they could also be left in the thermocycler at 4ºC overnight).

7.3.4. Purification

1. Before starting ice was taken in a container, with a metal support inside.

2. PCR product samples were brought to room temperature placed on the metal support. When thawed, they were given a quick vortex and spin and placed again on the ice.

3. Purification beads were brought to room temperature. After, they were inverted several times to homogenize them, and were placed on the ice (vortex of the purification beads was avoid).

4. Three µL of PCR product from each patient was placed in a new mini eppendorf and stored at 4ºC.

5. Next, the content of the 4 mini eppendorf from the same patient was mixed into a 1.5 eppendorf placed on the ice. After, the eppendorf were removed from the ice.

6. Seven hundred twenty µL of beads were added to each sample, and eppendorf were inverted 10 times to mix them.

7. Samples were incubated with the beads for 10 min.

8. They were then centrifuged for 3 min at 16,100g.

9. The eppendorf were taken out of the centrifuge and placed on a magnetic rack.

10. The supernatant was carefully aspirated to avoid dragging the beads, and then the eppendorf were closed to prevent the beads from drying out.

11. Without removing the eppendorf from the magnetic support, 1 mL of washing buffer was added.

12. The eppendorf were placed on the vortex support and vortexed for 2 min.

13. They were then centrifuged for 3 min at 16,100g.

14. After, they were placed directly on the magnetic rack, and the supernatant was aspirated as previously.

15. Next, the samples were centrifuged for 3 min at 16,100g and placed back on the magnetic holder to remove any remaining supernatant by pipetting.

16. The eppendorf were removed from the magnetic rack, opened and incubated for 10 min.

17. Next, 52 µL of elution buffer was added directly onto the beads, and the eppendorf were placed in the vortex support and vortexed for 10 min (it is important that no lump is observed).

18. After the vortex, samples were centrifuged 3 min at 16,100g.

19. They were then placed directly on the magnetic rack and incubated during 10 min.

20. While incubating, mini eppendorf were placed on the metal support inside the ice.

21. After incubation, and without removing the eppendorf from the magnetic rack, 47 µL of each sample were collected and poured into the mini eppendorf placed on ice.

22. Samples were then sealed with film and given a quick vortex and spin, and placed on the metal support on ice.

7.3.5. Quantification and quality assessment of purified DNA

1. Two µL of purified DNA was collected in another mini eppendorf to which 18 µL of Nuclease-Free water was previously added. Eppendorf were closed and given a vortex and a spin.

2. The quantification and quality assessment of purified DNA was carried out with the spectrophotometer Nanodrop ND-1000. Samples were considered to be of sufficient quality to carry out the labeling and hybridization when their concentration was ≥3ug/µL and the A260/280 ratio value was comprised between 1.8 and 2.0. For those cases with a concentration ≤2.5 ug/µL, the array technique was not continued.

7.3.6. Fragmentation

1. Before starting, the fragmentation centrifuge was set at 4ºC and 2000rpm during 30 min to be tempered.

2. Next, a fragmentation mix as detailed in Table 37 was prepared:

Reagent	Volume (µL)
Nuclease-Free Water	271.2 µL
10x Fragmentation Buffer	348.3 µL
Gene Chips Fragmentation Reagent enzyme	10 µL
Final volume	625 µL

Table 37. Fragmentation mix

3. After, the thermal cycler was set up as follows, and was stopped when the temperature reached 37ºC:

37ºC 35 min

95ºC 15 sec

Hold at 4ºC

4. Next, 10 µL of the fragmentation mix was added to each purified DNA sample as detailed in Table 38:

Reagent	Volume (µL)
PCR purified product	45 µL
fragmentation mix	10 µL
Final Vol	55 µL

Table 38. PCR purified DNAs with fragmentation mix

5. The eppendorf were sealed with film and given a quick vortex and spin.

6. They were then placed in the fragmentation centrifuge, and the start button was held down for a few seconds.

7. After, samples were placed in the thermocycler.

8. Once the thermocycler was finished, samples were placed on the ice.

7.3.7. Fragmentation control with 3% agarose gel:

1. A 3% agarose gel was prepared mixing 3g of agar with 100 mL 0.5% TBE and 5 µL of SYBR Safe DNA gel stain (Thermo Fisher Scientific, Waltham, Massachusetts, USA). It was left to solidify.

2. Twenty eight μL of nuclease-free water and 4 μL of fragmented DNA were added in a mini eppendorf. Samples were then vortexed and given a spin.

3. Once the gel was solidified, it was charged with the samples prepared as follows: 8 μL of diluted and fragmented DNA mixed with 2 μL of Gel Loading Dye Purple (New England Biolabs, Ipswich, Massachusetts, USA) (final volumen: 10 μL). Then, 8 μL of Low Molecular Weight DNA Ladder was charged in the first well.

4. Samples migrated at 130V for 60 min. The fragments obtained were between 25 and 125bp.

7.3.8. Labeling

1. Before starting ice was taken in a container, with a metal support inside.

2. Reagents were brought to room temperature. When thawed, they were given a quick vortex and spin and placed again on the ice.

3. Next, a labeling mix as detailed in Table 39 was prepared:

Reagent	Volume (μL)
TdT Buffer	14 μL
DNA Labeling Reagent	2 μL
TdT enzyme	3.5 μL
Final volume	19.5 μL

Table 39. Labeling mix

4. Nineteen point five μL of labeling mix was added to each sample, and were then sealed with parafilm, vortexed and given a spin.

5. After, the thermal cycler was set up as follows, and samples were introduced when the temperature reached 37ºC:

37ºC 4 h

95ºC 15 min

Hold at 4ºC

7.3.9. Hybridization

1. Before starting, turn on the oven 1 hour before and set it to 50ºC and 60 rpm.

2. The chips were removed from the refrigerator and were labeled and scanned.

3. Reagents were brought to room temperature. When thawed, they were given a quick vortex and spin and placed again on the ice.

4. Next, a hybridization mix as detailed in Table 40 was prepared in the hood:

Reagent	Volume (µL)
Hyb Buffer Part 1	165 µL
Hyb Buffer Part 2	15 µL
Hyb Buffer Part 3	7 µL
Hyb Buffer Part 4	1 µL
Oligo Control Reagent 0100	2 µL
Final volume	190 µL

Table 40. Hybridization mix

5. It was given a long vortex, because the liquid was very viscous, and it was spun.

6. Samples were removed from the thermocycler. Then, 190 µL of hybridization mixture was added to each sample as detailed in Table 41 and given a long vortex.

Reagent	Volume (µL)
Labeled PCR product	70.5 µL
Hybridation mix	190 µL
Final volume	260.5 µL

Table 41. Final volume in each sample

7. After, the thermal cycler was set up as follows:

   95ºC 10 min

   49ºC 2 min

   Hold at 49ºC

8. When the thermocycler finished, the samples were kept inside the thermocycler at 49ºC and the chips were loaded: with the chip upright, a filterless tip was inserted into one of the holes and 200 µL of the reagent was loaded through the other hole, being careful not to generate bubbles. After finishing, both tips were removed and the holes were sealed with stickers.

9. The chip was introduced into the oven after being loaded.

10. This process was repeated for each sample.

11. The chips were hybridized for 16-18 h maximum at 60 rpm and 50ºC.

7.3.10. Post hybridization washes

The washes were carried out using the GeneChip® Command Console® Software (AGCC) that controls the GeneChip® Fluidics Station 450 (FS450).

Before starting:

1. The computer and fluidic stations came on. The following message appeared: "Power-on done NOT PRIMED".

2. The bottles were filled with:

   -Wash A / Wash B (450 mL for 8 chips/ 300 mL for 4 chips)

   -Nanopure water

3. Empty eppendorf were placed in FS450.

4. The waste bottles were emptied.

5. The "Quick registration" was performed, and chips were identified with the barcode.

6. "Affymetrix Launcher" was opened, and later the AGCC Fluidics Control.

Priming:

7. PRIME_450 program was selected, and the ducts were introduced into the solutions (Wash A, Wash B, and water).

8. Empty tubes were loaded to the modules to be used and the program was run.

9. During the Prime 450 protocol, the amber colored eppendorf tubes (light sensitive solutions) were filled with the different "staining" solutions as detailed in Table 42. They can also be prepared the day before and kept at +4°C until use the next day.

Tampons	Volume (µL)
Stain Buffer 1	500 µL
Stain Buffer 2	500 µL
Array Holding Buffer (3)	800 µL

Table 42. Stain solutions

10. At the end of "Prime_450", the needle levers were uploaded and the empty eppendorf tubes were removed. Then, the machine was ready to wash the chips.

Chip washing:

11. CytoScanHD_Array_mapping, CytoScanHD_Array_450, respectively was selected, and the barcodes for each chip were read, and "tough-spot" blocking the openings in the chip were removed. The chips were then inserted into the cartridge.

12. Empty eppendorf were replaced with amber eppendorfs with the "Stain" solutions.

13. The needle lever on the tubes was lowered, starting the protocol (1h 40min). The scanner was then turned on.

14. At the end of the washing protocol, cartridge levers were lowered, chips removed and checked for bubbles (if there are bubbles, the cartridge should be washed again).

15. The two chip openings were closed with tough-spots, and the chips were introduced in the scanner feeder. Then, the AGCC scan control was started.

16. At the end of the cycle, the amber eppendorf tubes containing the "stain" were removed and replaced with the empty eppendorf tubes.

17. The needle levers were then lowered to perform the needles/conduits flush and all tubes were removed. At the end, the message "Protocol ready" appeared.

7.3.11. Scanning

The chips were read using the Affymetrix GeneChip Scanner 30007G, which measures the fluorescence emitted by the chip and translates it into a file (.DAT) that allows the chips to be visualized and their quality assessed. If the file quality is good, a CEL file is generated, used in the ChAS (Chromosome Analysis Suite) software to generate the .cyhd.cychp files, which allow viewing the results of the microarray.

7.3.12. Visualization and analysis of results

ChAS software was used to visualize and analyze the results. Firstly, the quality of the array hybridization was assessed for each case using the quality parameters obtained (QCmetrix):

- **MAPD:** it is a global measure of the variation of all probes on the chip across the genome, and represents the median of the distribution of changes in the log2 relationship between adjacent probes. Since it measures the differences between adjacent probes, it is a short-range noise measurement. Based on an empirical test data set, the provider decided that cases with MAPD > 0.25 are too noisy to provide a reliable number of copies. Thus, a good MAPD value will be **<0.25.**

- **SNPQC:** is a parameter that uses SNP markers to estimate the distribution of homozygous AA alleles, heterozygous AB alleles, and homozygous BB alleles. It is also based on an empirical test data set. However, a value less than the limit set by the provider (**≥15.0**) may be acceptable in cases with many abnormalities.

- **Waviness SD:** this parameter measures variations over a longer range than MAPD (variation between adjacent markers/probes). The measurement of long-range variation is accomplished by calculating the variation in log2 ratios across the entire genome and subtracting the short-range variation, specifically, for the autosomal region probe. This quality control measure should be standard (**≤ 0.120**) in cases with a normal copy number (2 copies). The higher it is, the less reliable the analysis will be, at least for small gains/losses.

However, in tumor DNA samples, the presence of numerous abnormalities generates non-standard values in the SNPQC and waviness SD parameters. Therefore, these parameters were correlated with the result and appearance of the different data on the chip.

7.4. NGS with Ion Torrent

The NGS technique was performed in 23 samples to validate the SNV detected by the designed panel. It was carried out in the Laboratoire d'Oncogénomique du

Centre Hospitalier Universitaire Vaudois (CHUV) in Lausanne, Switzerland, using the Ion S5 XL next-generation sequencing system (Ion Torrent, Thermo Fisher Scientific, Waltham, Massachusetts, USA). To that end, an in-house developed panel was used, which shared the genes detailed in Table 43 with our panel.

On the other hand, 11 samples were validated through the Molecular Biology service Laboratory, Pathology Service of Hospital del Mar, Barcelona, which were sequenced using an in-house developed sequencing panel detailed in Table 43 and the Ion Torrent sequencing platform, following the manufacturer's protocol. Table 43 gathers the genes that were shared with our panel.

Swiss Lausanne panel design	Barcelona panel design		
EZH2	ARID1A	FOXO1	TNFRSF14
TP53	CARD11	ID3	TP53
ATM	CCND3	KMT2D	EZH2
MYD88	CD79B	MEF2B	TNFAIP3
NOTCH1	CREBBP	MYD88	
SF3B1	EP300	TCF3	

Table 43. The genes covered by the two panels used to validate the results

7.4.1. Determination of DNA concentration with Qubit™ Fluorometer

1. All samples were first measured with the spectrophotometer Nanodrop ND-1000. For low volume samples, they were only measured with Qubit™ Fluorometer. According to the concentration of the samples to the nanodrop, two different reagents were used:

- BR > 100 ng/μL

- HS ≤ 100 ng/μL

2. Before using the Qubit, samples were brought to room temperature for 20 min.

3. The reagents were prepared according to the number of samples as detailed in Table 44 (for HS and for BR):

- n x 1 µl Qubit^{TM} Reagent (Dye)

- n x 199 µl Qubit^{TM} Buffer

Reagents	Standard*	Sample
Working solution	190 µL	199 µL
Standard	10 µL	-
Sample	-	1 µL
Final Volume	200 µL	200 µL

Table 44. Preparation of reagents for Qubit^{TM}

* Standard was made twice (2 tubes for Standard 1 and Standard 2 for HS; and 2 tubes for Standard 1/2 for BR).

4. Samples were given a vortex and spin, and incubated at room temperature during 3 min (fluorescence is stable until 3 hours after incubation).

5. The fluorescence of the samples was then measured at Qubit^{TM}, and the obtained values were multiplied by **0.2 for HS,** or by **100 for BR** to establish the concentration of each sample.

7.4.2. Genomic DNA amplification

1. Before starting the thermocycler was turned on and ice was taken in a container, with a metal support inside.

2. All the reagents were brought to room temperature. When thawed, they were given a quick vortex and spin and placed directly on the ice.

3. DNA samples were then diluted to obtain a final concentration of 2.2 ng/µL.

4. Sample ID x9 was diluted (at 20x):

- 3,6 µL Sample ID x9

- 4,4 µL H$_2$O

5. Next, it was mixed with the diluted DNA sample:

- 1,25 µL Sample ID x20

- 13,75 µL diluted DNA sample

6. PCR mix was prepared at room temperature as detailed in Table 45:

Reagents	Volume (µl)	
	Primer Pool 1	Primer Pool 2
5x Ion AmpliSeq™ Hifi Master Mix	4	4
Primer pool	10	10
Mix DNA-ID/cDNA	6	6
Final Volume	20 µL	20 µL

Table 45. PCR mix

6. Samples were given a vortex and spin.

7. After, the thermal cycler was set up as follows:

99ºC 2 min

99ºC 5 sec ⎤
⎥ x16
60 ºC 4 min ⎦

Hold at 10ºC

8. Once the thermocycler was finished, the pools were mixed in a new tube as follows:

- Pool 1: 10 µL

- Pool 2: 10 µL

- Final volumen: 20 µL

7.4.3. Partial digestion of primers

1. Before starting, FuPa reagent was brought to room temperature. When thawed, it was given a quick vortex and spin and placed directly on the ice.

2. Two µL of FuPa reagent was added to each sample (final volume: 22 µL).

3. After, the thermal cycler was set up with the following program:

50ºC 20 min

55ºC 20 min

60 ºC 20 min

Hold at 10ºC (for up to 1 hour)

4. While the barcodes were prepared as detailed in Table 46 (Gloves were changed each time a different barcode was pipetted):

Reagents	Volume (μL)
Nuclease free water	1.10 μL
Ion P1 Adaptor	0.55 μL
Ion Xpress barcodes X	0.55 μL
Final Volume	2.2 μL

Table 46. Barcode mix

7.4.4. Ligation

1. Before starting, switch solution was brought to room temperature. When thawed, it was given a quick vortex and spin and placed directly on the ice.

2. Next, the ligation mix prepared as detailed in Table 47:

Reagents	Volume (μL)
Switch solution	4 μL
Prepared barcodes	2 μL
DNA ligase	2 μL
Final Volume	30 μL

Table 47. Ligation mix

3. The thermocycler was set up as follows:

22ºC 30 min

72ºC 10 min

Hold at 10ºC (for up to 24 hours)

7.4.4. Purification

1. Agencourt™ AMPure™ XP Reagents were brought to room temperature and vortex thoroughly to disperse the beads before use.

2. Forty five μL of Agencourt™ AMPure™ XP Reagent were transferred in a new tube.

3. The total volume of the ligation (30 µL) was added into the tube containing the beads (final volume: 75 µL). It was pipetted up and down 5 times to mix the bead suspension with the DNA thoroughly.

4. The mixture was incubate for 5 min at room temperature.

5. Samples were then placed in a magnetic rack and incubated for 2 min. After, the supernatant was carefully discarded without disturbing the pellet.

6. One hundred fifty µL of freshly prepared 70% ethanol was added, then samples were moved side-to-side in the two positions of the magnet to wash the beads and incubated for 2 min. After, the supernatant was carefully discarded without disturbing the pellet.

7. Step 21 was repeated for a second wash.

8. Keeping the samples in the magnet, the beads were air-dried at room temperature for 5 min.

Library equalization:

9. Tubes with purified libraries were removed from the magnet, and 50 µL of Platinum™ PCR SuperMix HiFi and 2 µL of Equalizer™ Primers were added to each bead pellet. Samples were sealed, vortexed thoroughly, and centrifuged to collect droplets.

10. Tubes were then placed back on the magnet for at least 2 min, then ~50 µL of supernatant from each sample was carefully transferred to a new plate/epp without disturbing the pellet.

11. The plate/epp were sealed with a new clear adhesive film, and loaded in the thermal cycler to amplify. The following program was run:

98ºC 2 min

98ºC 15 sec ⎤ x5

64ºC 1 min ⎦

Hold at 10ºC

First-round purification:

12. Twenty five μL of Agencourt™ AMPure™ XP Reagent was placed in a new 1.5 mL tube. The entire volume of the amplified libraries (~50 μL) was transferred to the tubes containing the beads.

13. The bead suspension was mixed with the DNA thoroughly by pipetting up and down 5 times. After, the mixture was incubated for 5 min at room temperature.

14. The tubes were placed in a magnet for at least 5 min.

Second-round purification:

15. Sixty μL of Agencourt™ AMPure™ XP Reagent was placed in a new 1.5 mL tube.

16. After the incubation in the magnet, the supernatant (~75 μL) was carefully transferred without disturbing the pellet to the tube containing the beads (final volume: 135 μL).

17. The bead suspension was mixed with the DNA thoroughly by pipetting up and down 5 times. After, the mixture was incubated for 5 min at room temperature.

18. The tubes were placed in a magnet for at least 3 min. The supernatant was then carefully removed without disturbing the pellet.

19. One hundred fifty μL of freshly prepared 70% ethanol was added to each tube, and tubes were then moved side to side in the magnet to wash the beads. After, the supernatant was carefully removed without disturbing the pellet.

20. The step 34 was repeated for a second wash.

21. Keeping the tubes in the magnet, the beads were air-dried at room temperature for 2–5 min.

22. Next, the tubes were removed from the magnet, and 50 μL of Low TE was added to the pellet to disperse the beads.

23. Samples were then vortexed thoroughly, centrifuged to collect droplets, and incubated at room temperature for at least 2 min.

24. Without removing the eppendorf from the magnetic support, the supernatant was aspirated manually and very carefully and transferred to a new tube.

25. Samples were then measured with Qubit™ Fluorometer. The reagents were prepared according to the number of samples as detailed in Table 48 (for HS and for BR):

- n x 1 µl Qubit™ Reagent (Dye)

- n x 199 µl Qubit™ Buffer

Reagents	Standard*	Sample
Working solution	190 µL	*195 µL/ **180 µL
Standard	10 µL	-
Sample	-	*5 µL/**20 µL
Final Volume	200 µL	200 µL

Table 48. Preparation of reagents for Qubit™

* For the manual library; **for the automatic library

26. Samples were then diluted at 35 pM, based on the calculated library concentration for the average amplicon size, detailed in Table 49:

Average amplicon size	Concentration in ng/mL (~100 pM)
140 bp	9
175 bp	11
***225 bp**	**15**
275 bp	18
375 bp	24

Table 49. Calculated library concentration for the average amplicon size

* The average amplicon size for our library was 225 bp

27. Libraries were loaded on the chip.

7.4.5. Sequencing with IonS5 TM XL (Ion Torrent)

Preparation of consumables:

1. The Ion S5™ Sequencing Reagent cartridge was unboxed 45 min before use and brought to room temperature. All cartridges and consumables were removed from their packaging, and placed on the bench next to the Ion Chef™ Instrument.

Sequencer initialization:

2. The initialization was carried out one day before performing the sequencing. In the instrument touchscreen main menu, Initialize was selected. The door, chip, and Reagent cartridge clamps were unlocked.

3. When prompted, the Ion S5 ™ Wash Solution bottle was removed to access the waste reservoir, and after, the waste reservoir was removed, emptied and reinstalled.

4. The expended Ion S5™ Sequencing Reagents cartridge was replaced with a new cartridge equilibrated to room temperature.

5. The new Ion S5™ Wash Solution bottle was thoroughly mixed. Then, the red cap was removed and installed.

6. The used sequencing chip from the previous run was properly seated in the chip clamp and the chip clampwas pushed in all the way.

7. If necessary, a new Ion S5™ Cleaning Solution bottle was installed (the Ion S5™ Cleaning Solution bottle contains sufficient reagent to complete 4 cleanings).

8. The door was closed, and Next was pressed. The instrument confirmed that the consumables and chip were properly installed and that the Ion S5™ Cleaning Solution contained sufficient reagent to perform the post-run clean.

9. When initialization was completed (~50 min), Home was pressed.

Loading the Ion Chef S5™ System:

10. The Chef was loaded following the instructions of the provider, as detailed in Figure 32.

① Empty tip rack (move from new Tip Cartridge position)
② Frame Seal v2
③ New Tip Cartridge
④ PCR plate
⑤ Ion 510™ & Ion 520™ & Ion 530™ Chef Reagents cartridge
⑥ Ion S5™ Chef Solutions cartridge
⑦ Recovery Tubes and Recovery Station Disposable Lid v2
⑧ Enrichment Cartridge v2
⑨ Chip Adapter/Chip assemblies
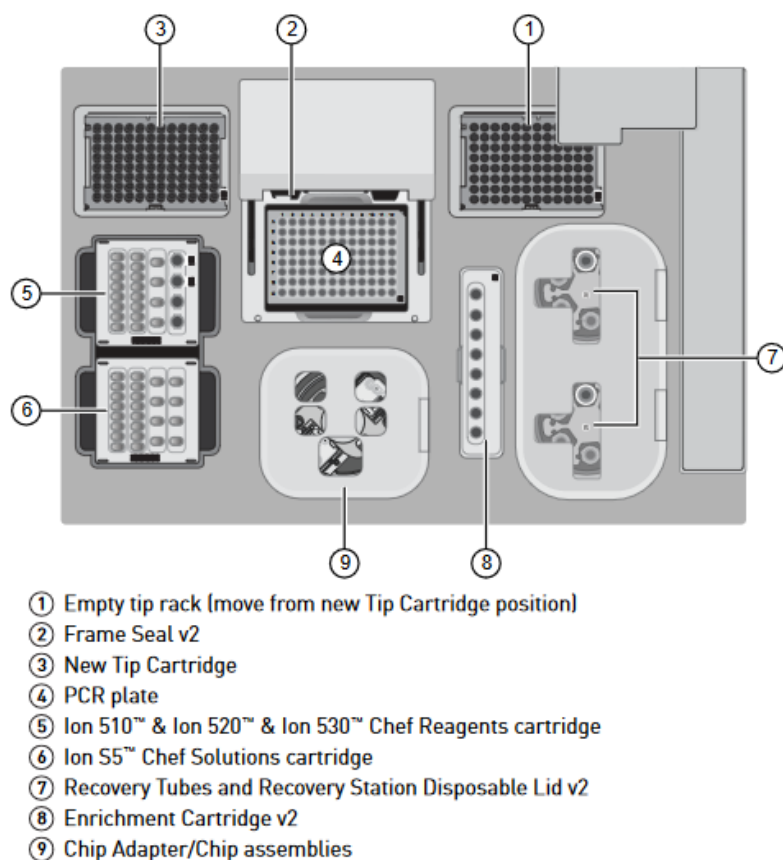
Figure 32. Scheme to load the Ion Chef

Sequencing:

11. After completion of initialization, Run was pressed in the instrument touchscreen. The door and chip clamp were then unlocked.

12. The used sequencing chip was removed, and the new chip was then secured in the chip clamp.

13. The chip clamp was pushed all the way down to engage it, the instrument door was closed and then Next was pressed.

14. It was confirmed that the correct Planned Run was auto-populated, then, Review was pressed.

15. It was confirmed that the remaining pre-populated settings were correct, or changes were made using the buttons and dropdown lists if needed.

171

16. The instrument door was closed, then Start run was pressed to begin the sequencing run.

7.4.6. Analysis of sequencing data

The analysis of the NGS data was performed with the Sequence Pilot software, SeqNext module, from JSI Medical System GmBH. The analysis was performed using fastq files generated by the sequencer, which allowed identifying the SNV presented in the samples.

After sequencing, the following parameters were analyzed:

- o   Read lenght histogram of each panel

- o   If the coverage was correct (≥1000x)

- o   Number of reads and % clonality-polyclonality

- o   Sample ID verifiation (if corresponded to the patient)

- o   Variant identification

Hypotheses and Aims

1. **Hypotheses**

The application of NGS techniques to the study of mature B-cell neoplasms at the research level is providing a large amount of new information, in certain cases with clinical implications, which includes the identification of molecular biomarkers with diagnostic, prognostic and predictive value. The incorporation of some of these molecular biomarkers into clinical practice has shown to be crucial, as it is reflected in the evolving classification of B-cell malignancies (Swerdlow et al., 2016), as well as in several consensus guidelines for patient management. Therefore, the integration of molecular information together with the genetic alterations detectable by NGS is of main relevance to transfer the new genomic knowledge generated to daily clinical practice.

In this sense, in recent years numerous NGS panels have been commercialized for the detection of SNV and small indel variants, mainly, in solid tumors and hematological neoplasms, with a greater focus on myeloid neoplasms. Besides, some of these panels also allow the analysis of CNA and fusion genes by amplicon-based sequencing using RNA. Focusing on the study of MBCN, NGS panels available on the market are scarce and mainly focused on the study of SNV and small indel variants. In this sense, to date, none of the panels available on the market can detect structural chromosomal alterations, such as translocations, which are of special interest in these entities.

On the other hand, mature B-cell neoplasms can often have overlapping clinical and morphological features that could pose diagnostic challenges (Momose et al., 2015; Swerdlow et al., 2016; Wagener et al., 2019). However, most of the panels developed focus on the study of a single entity, while the use of a panel that comprises the genetic alterations involved in different MBCN, could help to make a differential diagnosis, in addition to predict the progression to a more aggressive lymphoma.

Finally, no studies have yet been conducted comparing this technology with conventional genetic techniques in the framework of MBCN.

On these grounds, the hypotheses of the present study are:

1. The development of a targeted NGS panel will allow the study of genetic alterations with diagnostic, prognostic and therapeutic value in mature B-cell neoplasms, currently determined with different techniques, using a single methodology.

2. The simultaneous study of multiple genetic alterations in a single test, as well as the parallel analysis of different patients, will reduce the long term economic cost in diagnostic laboratories.

3. In addition to the study of the alterations included in the current clinical guidelines, the designed panel will allow the inclusion of other relevant genetic alterations, which may have an impact on the diagnostic routine in the near future.

4. Finally, designing a NGS panel that includes genetic alterations of the different mature B-cell neoplasms could help to make a differential diagnosis, in addition to allowing the genetic characterization of less frequent neoplasms for which designing a specific panel would not be cost effective in clinical practice.

2. Aims

The general aim is to develop a targeted sequencing panel for the detection of the main genomic events, including base substitutions, indels, CNA, and chromosomal translocations with clinical value in mature B-cell neoplasms, providing greater efficiency and scalability than the combination of genetic techniques currently used in the diagnosis of these lymphomas.

The more specific aims are:

1. Design a targeted NGS panel: select the main genomic alterations with clinical relevance described in the different mature B-cell neoplasms.

2. Design a bioinformatic algorithm to analyze the results of the sequencing panel.

3. Validate the genomic alterations detected by the designed panel using different conventional genetic techniques (chromosome banding analysis, FISH, genomic microarrays, PCR and Sanger sequencing) in addition to other massive sequencing panels in the case of mutations.

4. Compare detection sensitivity and specificity of the developed panel with respect to conventional genetic techniques performed in routine laboratories in addition to other NGS panels.

Results and Discussion

1. **Sequencing quality analysis**

In this section we aim to describe the primary metrics used to evaluate sequencing performance of the two designed panels: depth of coverage (which is the number of unique reads that include a given nucleotide in the reconstructed sequence), and mapping quality (which is the confidence that the read is correctly mapped to the genomic coordinates).

1.1. Coverage depth

We first examined the mean target coverage depth in the sequenced samples with both panels, PV1 and PV2, in order to define a cut-off to consider that a sample was suitable to be analyzed. In this sense, a minimum mean depth of coverage of 110x was established. Therefore, 14 samples (two in PV1 and 12 in PV2) which did not fulfill this criterion were removed from the study.

Globally, despite showing high values of total reads sequenced in each sample, the mean target coverage depth obtained for PV1 was significantly lower than that observed in PV2 samples (mean: 222x vs 506x, respectively; P<0.001). This increase in the coverage obtained in the second version was expected as a consequence of the panel size reduction (4.6Mb for PV1 vs 1.4Mb for PV2). However, while PV1 showed a great homogeneity among samples, an unexpected variability in the mean depth of coverage observed among samples sequenced in PV2 was found. For this reason, to characterize the uniformity of coverage throughout the capture space, the proportion of bases sequenced to 100x depth in each sample was examined. Despite the variability in the median depth of coverage observed among the samples in PV2, 85.5% of these (n=59/69) presented with ≥90% of their bases sequenced at 100x (with a median of 92.5%). Regarding PV1 samples, their lower mean coverage was associated with less proportion of bases sequenced at a minimum of 100x (median: 83%). Indeed, only 77.8% (n=84/108) of the PV1 samples presented with ≥80% of their bases sequenced at 100x (Table 50).

In addition, coverage data were also assessed to analyze whether there were differences between the samples sequenced within each panel design: samples sequenced in the different runs, and samples with different tissue processing (fresh, frozen and FFPE tissue). Regarding the effect of the different runs, a significant batch effect was discarded for each panel design sequenced. Notably, the first sequencing run of PV1 presented lower values in the mean target coverage depth with respect to the samples sequenced in other runs, although these differences were not significant (Figure 33). These results are explained by the fact that a greater number of samples were included in this initial run compared to the remaining, in which fewer samples were multiplexed to improve the quality of the data. With regard to the type of tissue preparation, no significant differences were observed between the different samples for either of the two panels (Figure 34).
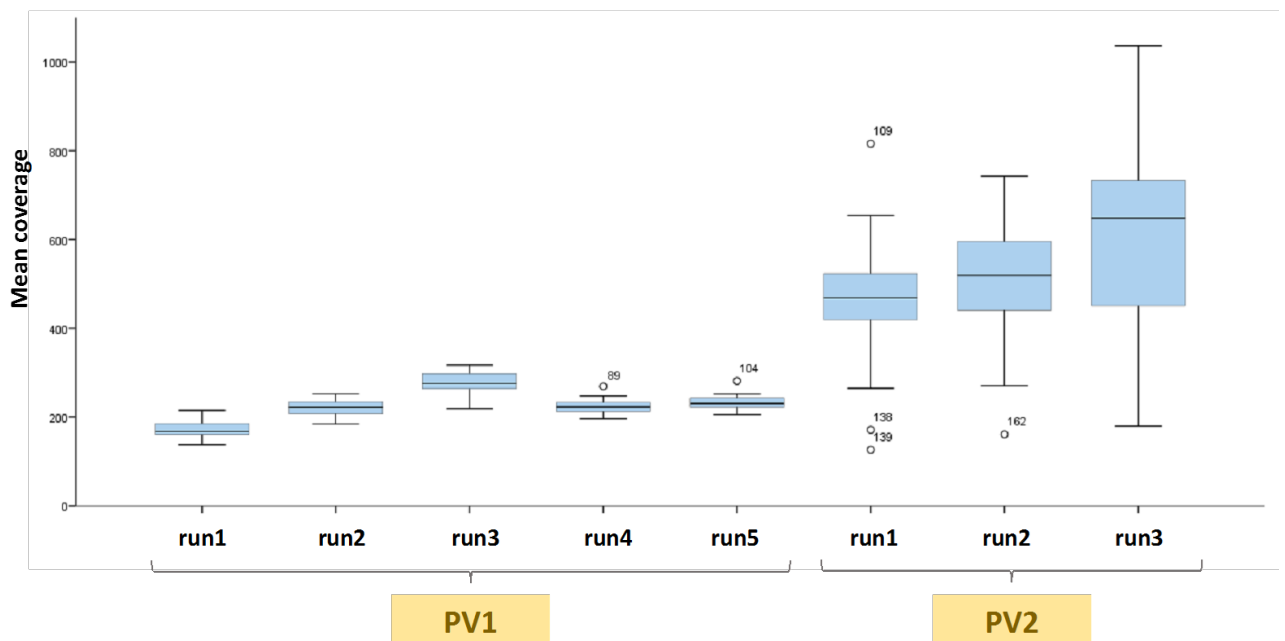
Figure 33. Distribution of the mean depth of coverage obtained for each of the runs carried out for both versions of the panel (PV1 and PV2)
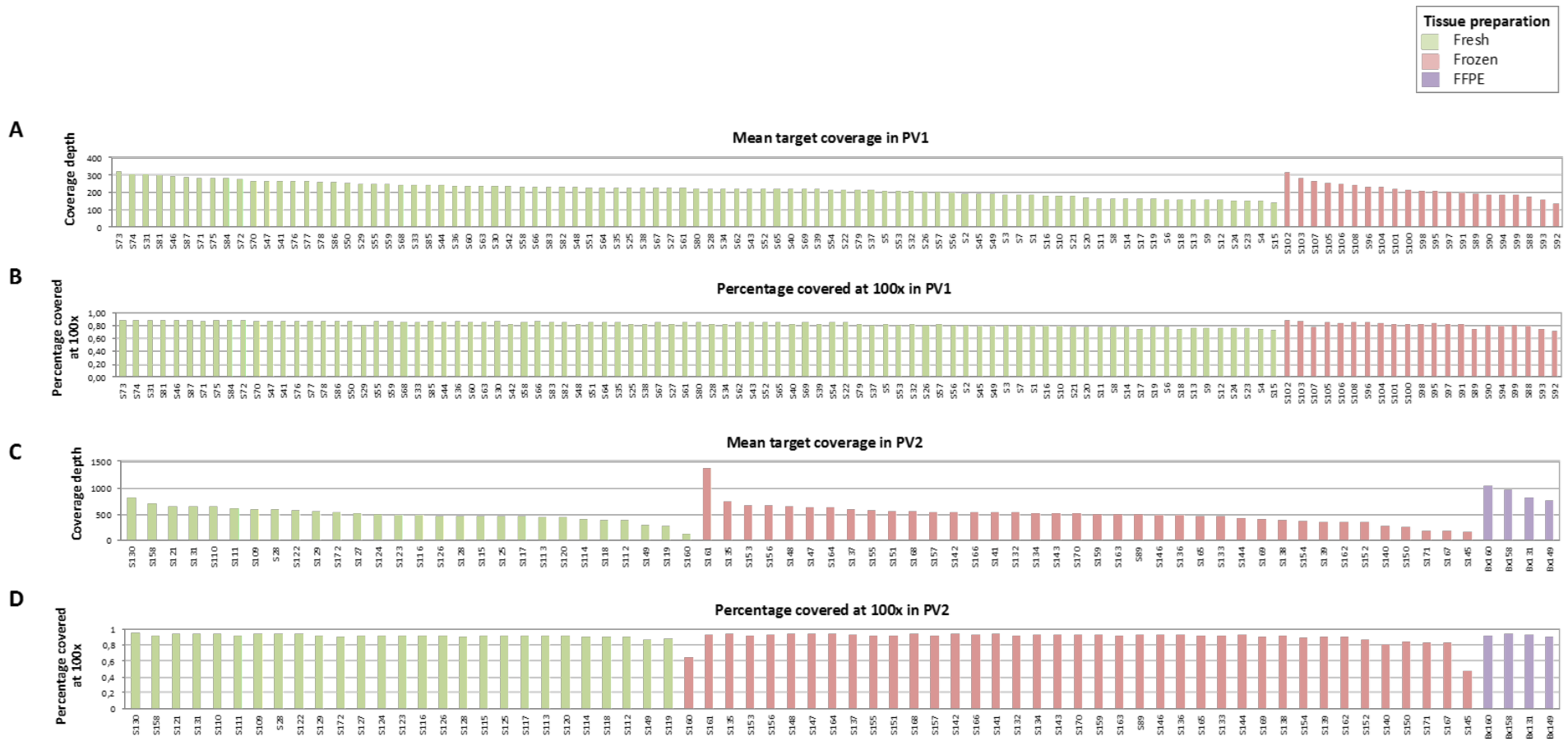
Figure 34. Mean target coverage and percentage of bases covered at 100x for samples sequenced with PV1 (A and B) and PV2 (C and D). Samples from different tissue preparations are displayed in different color (Fresh tissue in green; Frozen tissue in red; and FFPE tissue in purple).

In all, these data show uniformity of coverage throughout the entire capture region which globally reaches at least 100x in almost all the samples. These results are similar to those established by the Spanish Group of Myelodysplastic Syndromes (GESMD) that recently published guidelines for NGS testing in myeloid neoplasms. In these guidelines, the authors recommend that ≥95% of the detected bases should be covered by a minimum of 100 reads (Palomo et al., 2020). However, they suggest a minimum mean coverage depth of 1000x not achieved by our panel. These recommendations are mainly oriented to the use of amplicon-based sequencing panels for the detection of SNV. Indeed, these designs usually have a small size, allowing them to reach a much greater depth of coverage (average depth of 1000x), and therefore also a higher sensitivity (average sensitivity of 2.5% if the coverage is 1000x). Nonetheless, our design does not focus exclusively on the detection of SNV, but also on the detection of CNA and rearrangements. As a consequence, our panel has a larger size and therefore a lower depth of coverage, which reduces its sensitivity. However, since our objective is not the detection of minimal residual disease, nor the detection of subclonal variants, the depth of coverage reached by our panel has been considered sufficient for the detection of the alterations for which it has been designed. In addition, these data are supported by previous studies carried out with hybridization capture panels with a design similar to ours (which also analyze CNA, rearrangements and SNV), which affirm that depth of coverage lower than those achieved in our study are sufficient for the correct detection of these alterations (19x-150x for SNV and CNA, and 19x-170x for rearrangements) (Grossmann et al., 2011; Duncavage et al., 2012; Frampton et al., 2013; He et al., 2016; McKerrell et al., 2016; Prieto-Conde et al., 2020). Comparative data with similar published panels are detailed in section 6 (**Global analysis of the designed panel and its application in clinical practice**).

In this line, other studies have also been published that use targeted panels for the detection of SNV, exclusively, which accept a much lower average depth of coverage (250x) than that recommended in the NGS guidelines (Hung et al., 2018).

1.2. Mapping quality

The enrichment strategies used for targeted sequencing have limited efficacy. As a consequence, a large proportion of the generated sequences are mapped outside the target regions, and produce the so-called off-target sequences. While this proportion is relatively low (~10%) for amplicon panels, it can be up to 40-60% for capture panels (Samuels et al., 2013). For this reason, we examined the specificity of the capture approach for enriching DNA in the target regions of our assay. The median percentage of off-target sequences was 22.2% for PV1 and 10.5% for PV2 (Table 50; Figure 35). These values are in accordance with the recommendations made by NGS guidelines in other hematological malignances, which suggest that the percentage of off-target sequences should be less than 30% (Palomo et al., 2020), indicating a high probe complementarity. Furthermore, the values obtained in our study are better than those reported in previous studies with similar panels, which reported an on taget efficiency of 36.3% in one of the studies (Bolli et al., 2016) and 2.9%-63% in the other (Grossmann et al., 2011).

It is noteworthy that a decrease in the proportion of off-target sequences was observed in PV2 compared to PV1. This decrease could be associated with both the effectiveness of the target enrichment procedures or the changes made in the panel design. On one hand, it has been described that higher capture efficiency would lead to fewer off-target reads (Kuilman et al., 2015). However, as no changes were made in the technical protocol, nor in the enrichment kit used for both panel designs, similar effectiveness may be assumed. On the other hand, although it was not expected, the removal of some target regions and SNP from the design might have had an effect on the distribution of off-target sequences generated in the sequencing runs from PV2. In fact, although these sequences have been reported to be homogeneously distributed (results observed in WES studies; Kuilman et al., 2015), our results suggested the contrary. Thus, in those cases assessed with both panel versions, we observed a higher concentration of these sequences in the same regions when SNP were included in the design (PV1) (Figure 36). Therefore, it is possible that enrichment in certain genomic regions favors the generation of off-target sequences adjacent to those regions.
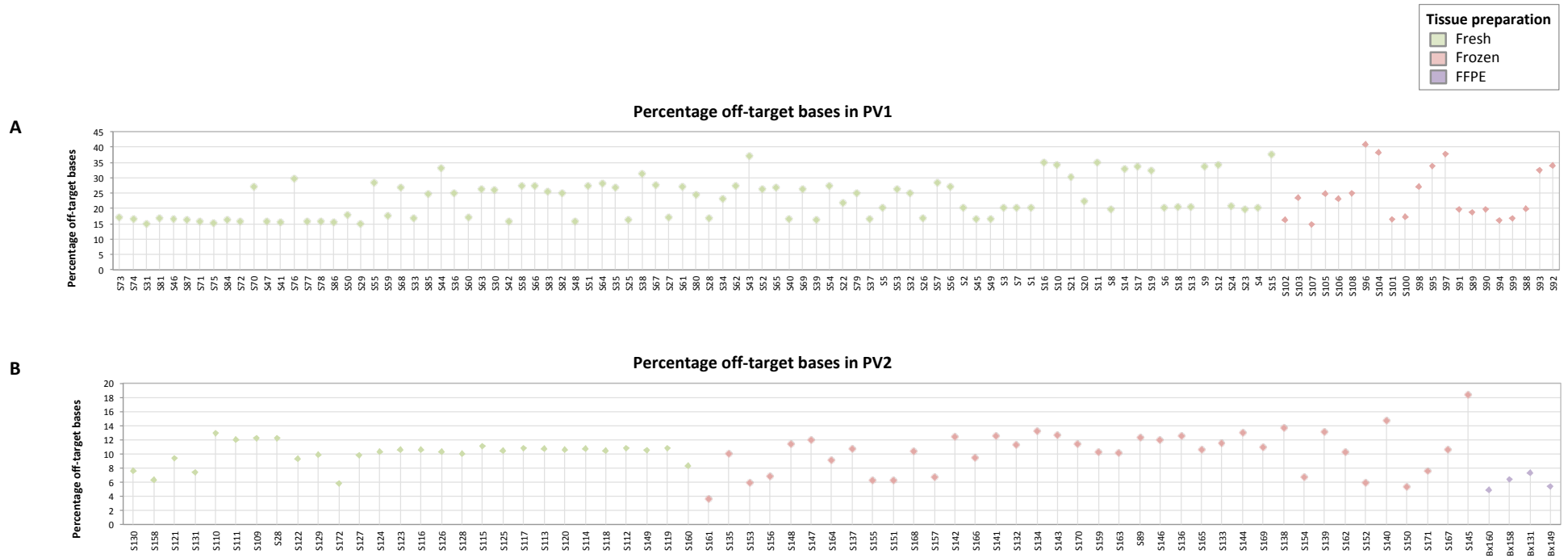
Figure 35. Percentage of off-target bases for PV1 (A) and for PV2 (B). Samples from different tissue preparations are displayed in different color (Fresh tissue in green; Frozen tissue in red; and FFPE tissue in purple).
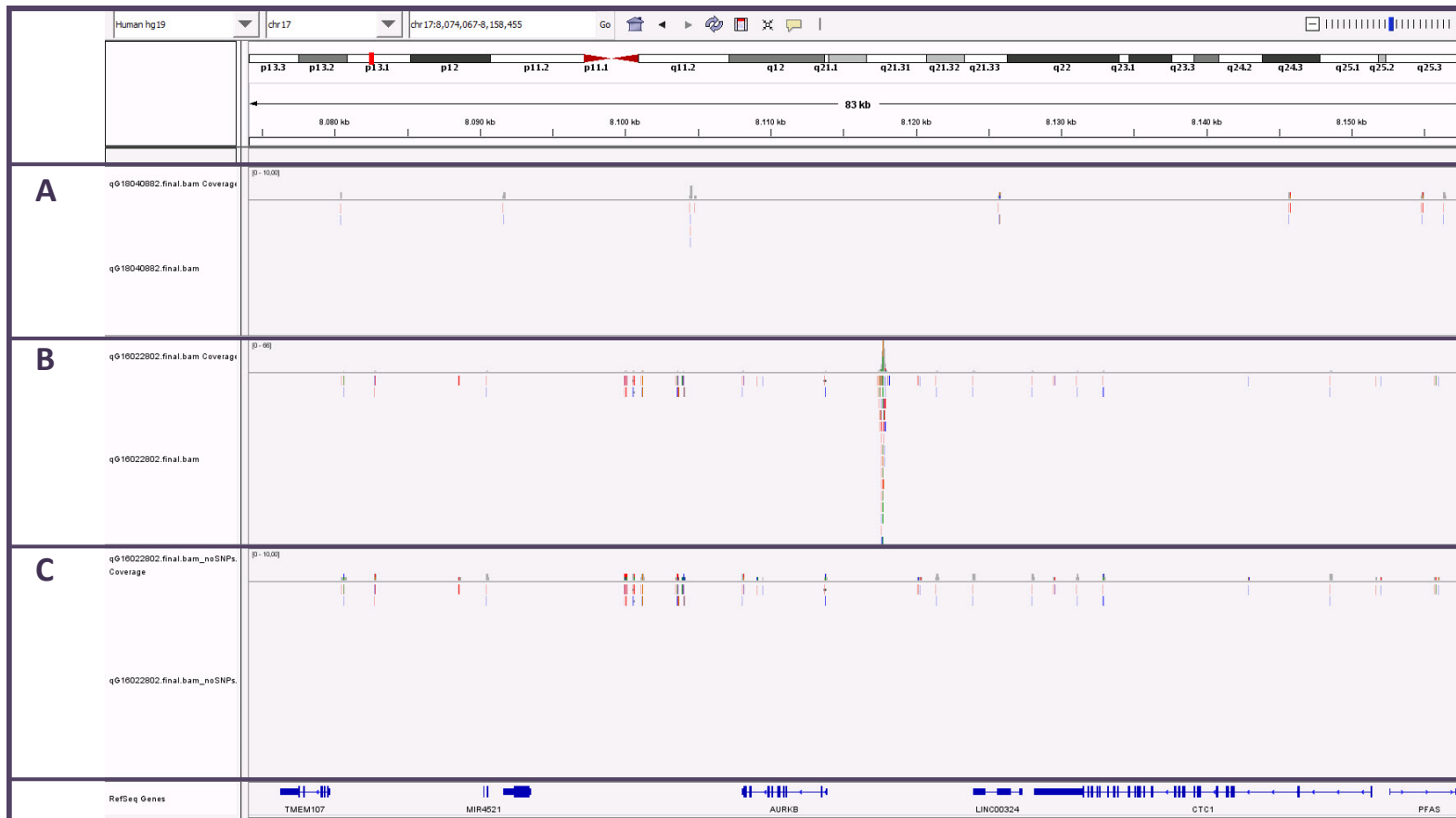
Figure 36. Comparison of the density of off-target sequences observed in a non-target region with the Integrative Genomics Viewer (IGV) for the same patient in (A) PV2, (B) PV1 with SNP and (C) PV1 with bioinformatically removal of SNP

Table 50. Detail of the quality parameters obtained for each panel (PV1 and PV2)

Panel version	Global vs Tissue preparation	Total reads Mean (min-max)	Mean target coverage depth Mean (min-max)	Percent covered at 100x Mean (min-max)	Percent off-target bases Mean (min-max)
PV1	Global	12112799 (7861383-16310248)	221.7 (138-317.3)	83 (72-88)	22.2 (14.8-40.8)
	Fresh	12127881 (7861383-15707019)	222.3 (145.4-317.3)	84 (74-88)	21.9 (14.8-37.4)
	Frozen	11894532 (8625605-16310248)	210.5 (138-314.1)	82 (72-87)	23.1 (14.8-40.8)
PV2	Global	6450000 (1800000-15800000)	506.2 (126.3-1379.7)	92.5 (46.7-94.8)	10.5 (3.6-18.4)
	Fresh	6100000 (1800000-9800000)	480.3 (126.3-815.8)	92.4 (65.1-94.5)	10.5 (6.3-12.9)
	Frozen	6450000 (2200000-15800000)	506.2 (161.3-1379.7)	92.7 (46.7-94.8)	10.6 (3.6-18.4)
	FFPE	10800000 (9700000-12900000)	883.8 (762.6-1036.6)	92.5 (91.2-93.8)	5.9 (4.9-7.3)

2. Single nucleotide variations and indels analysis

2.1 Assessment of sensitivity and specificity

For optimal assay accuracy of SNV and indels variant detection and calling, only the raw data obtained with the MUTECT2 software from samples with an average coverage of >110x were considered and filtered. The first step consisted of comparing the obtained data with those variants previously reported using Sanger sequencing (n=48 patients, 135 variant calls), as well as with other NGS panels (n=30 patients, 294 variant calls) in order to assess the sensitivity, specificity as well as the accuracy on VAF detection from PV1 and PV2 designs.

Comparative analysis with the 429 variants analyzed by reference and validation techniques in 78 patients revealed that no false positive results were detected, so the specificity was 100% for both panels (confirmed normal results in 199 and 122 variant calls in PV1 and PV2, respectively). Moreover, both panels could detect almost all the known variants achieving a sensitivity of 100% (75/75) for PV1 and 97% (32/33) for PV2. The only variant not detected was a null variant (intronic within ±2 of splice site) affecting *TP53* gene (VAF: 47.9)*,* and which position, despite being covered by both panel designs, was not well covered in this patient (with a depth of coverage of 15 reads) (Figure 37). With regard to the VAF and coverage achieved for the validated variants, VAF values measured using the designed panels and the reference/validation techniques showed a high correlation (R^2 =0.8986 for PV1, and R^2 =0.8998 for PV2) (Figure 37). However, the obtained coverage was very variable (range: 25-331x for PV1, and 26-1228x for PV2) and 11 of the 75 variants (14.6%) detected in PV1 presented less than 10 supporting variant reads, which was considered insufficient. Thus, to increase the assay accuracy, a second threshold of ≥10 supporting variant reads was established for all the variants detected, unless they had been validated or re-sequenced for confirmation. As they were validated, all the 11 variants were considered for subsequent analysis although they were not considered in the assessment of the limit of detection of the panel due to their poor quality. In all, the limit of variant frequency detection for each design was established, based on the lowest validated VAF, at 6.4% for PV1 and 4.2% for PV2.
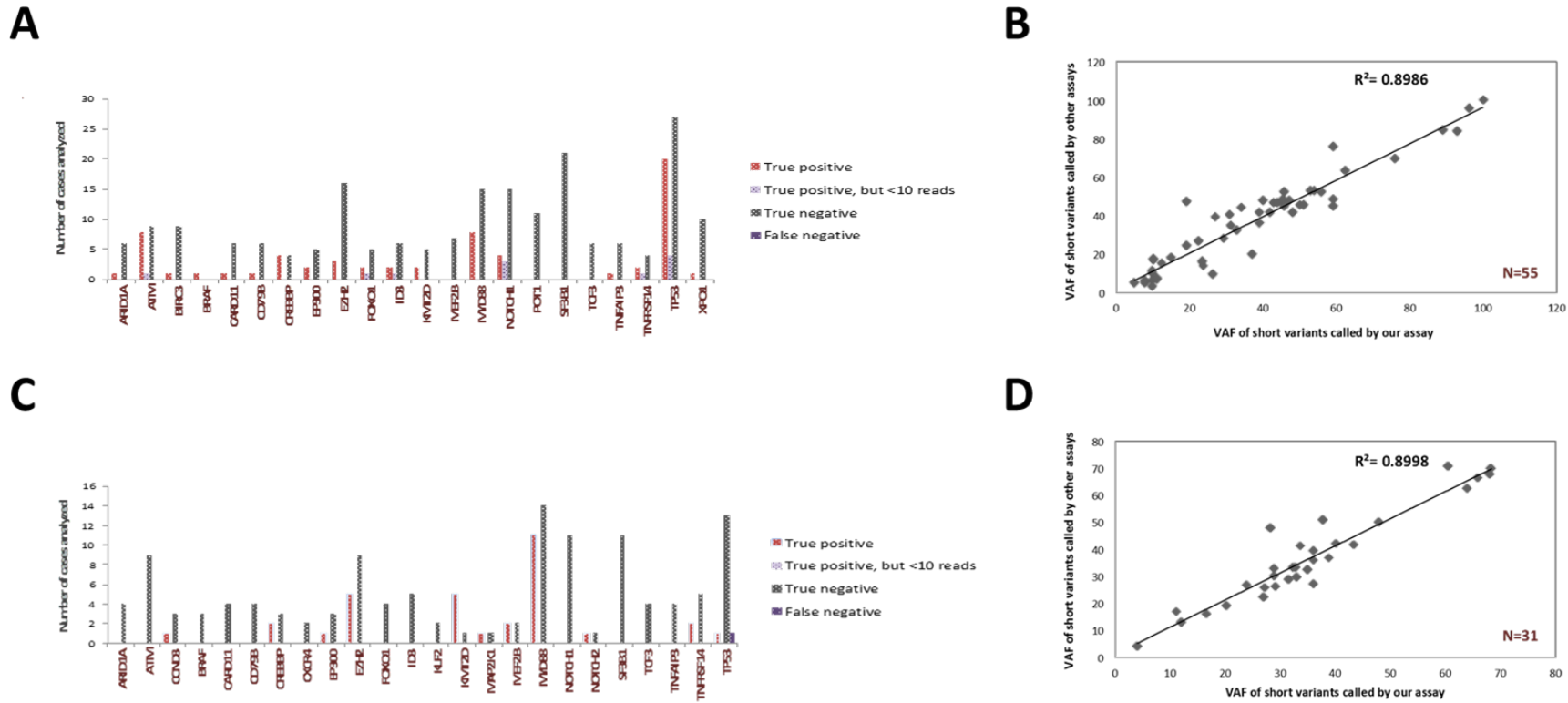
Figure 37. Alteration calling accuracy and concordance with reference platforms. The number of true or false positive and negative calls found by our assay based on the oncordance with reference/validation assays on 22 genes for PV1 (A) and 24 genes for PV2 (C); Correlation of short variant VAF called by our assay vs reference/validation assays in 55 substitutions and indels PV1 (B) and 31 for PV2 (D), respectively.

On the other hand, regarding the two cases sequenced by both panel versions, the same variants (N=9) were detected at similar VAF by both designs (R^2 =0.8984), in addition to a variant in the *CCND3* gene that was also identified by PV2, and which was not covered in the first design. All the variants were also validated by another NGS panel. Although reproducibility was confirmed, the median depth of coverage across the variants doubled in the second design (141x for PV1 vs 328x PV2). Indeed, despite not being statistically significant, PV2 VAF showed a greater correlation with those defined by the validation techniques (PV1: R^2 =0.9038; PV2: R^2 =0.9325) (Table 51).

Table 51. Detailed of the variants detected in the two patients sequenced by both panel versions (PV1 and PV2)

	Gene	Transcript	HGVS.c	HGVS.p	PV1		PV2		Validation techniques
					VAF	Coverage	VAF	Coverage	VAF
Patient 1	MEF2B	ENST00000162023.5	c.70C>T	R24W	25.5	75	27	212	36.09
	KMT2D	ENST00000301067.7	c.1378_1379delCC	P460Tfs*3	26.1	25	41.1	191	33.64
	KMT2D	ENST00000301067.7	c.4488_4491delACAC	H1497Vfs*8	25.7	209	32.3	385	34.98
	EZH2	ENST00000320356.2	c.1937A>G	Y646C	27.1	229	25.7	507	27.15
	CCND3	ENST00000372991.4	c.778C>T	Q260*	-	-	33	269	32.46
	TNFRSF14	ENST00000355716.4	c.136G>A	E46K	70.2	79	70	258	68.36
	TNFRSF14	ENST00000355716.4	c.104A>G	Y35C	68.3	84	67.5	255	68.06
Patient 2	CREBBP	ENST00000262367.5	c.4366G>T	G1456*	19	141	26.2	328	29.19
	CREBBP	ENST00000262367.5	c.4507T>G	Y1503D	27.2	199	22.2	340	26.87
	EZH2	ENST00000320356.2	c.1937A>G	Y646C	33.9	149	36.6	423	38.94

Taking into consideration this initial validation of the panel, all the variants with a mean coverage >110x, ≥10 supporting variant reads (unless validated), and VAF ≥ 6.4% for PV1 and ≥ 4.2% for PV2, were further analyzed. In all, a total of 462 variant calls were detected by both designs (259 in PV1 and 203 in PV2), with a median depth of coverage across the variants detected of ~170x (PV1: 169x, range: 21-377x; PV2: 175x, range: 18-1228x), and with >90% of the variants with a depth of coverage ≥50x (90.7% vs 93.6% in PV1 and PV2, respectively) (Figure 38).
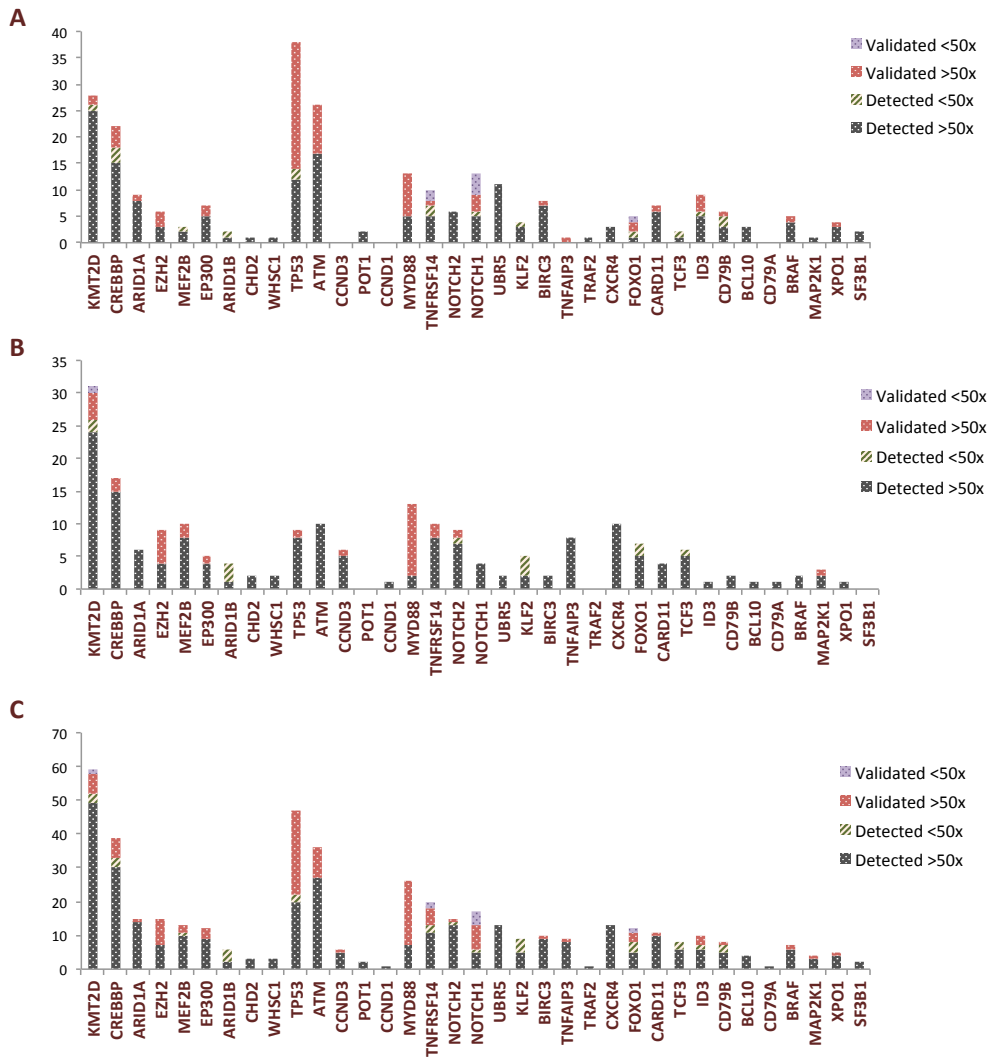
Figure 38. Total number of SNV and indel variants detected in PV1 (A), PV2 (B), and in both (C). The count of positive calls are highlighted by different colors: red (alteration validated by reference test and presented a depth of coverage ≥50x), light purple (alteration validated by reference test and presented a depth of coverage <50x), grey (alteration not validated and presented a depth of coverage ≥50x), and white and green (alteration not validated and presented a depth of coverage <50x).

Globally, the obtained results validate that both panel designs could successfully detect SNV and indels with high sensitivity and specificity, with a lower limit of VAF detection of at least 4-6%. We established thresholds of an average coverage of >110x and ≥10 supporting variant reads to consider a variant, which are consistent with those suggested by other authors who establish a threshold of 50x average

coverage to analyze a sample, in addition to a ≥10 reads with the alternative allele to analyze a variant (Hung et al., 2018), or suggest a VAF ≥5% to report it (Palomo et al., 2020).

On the other hand, the average depth of coverage in the variants detected for both designs was ~170x. Although this depth of coverage is not as high as in other published studies with targeted sequencing panels, it has been reported to be sufficient to make a reliable SNV and indels call (Cheng et al., 2015; He et al., 2016). Furthermore, despite the fact that great variability was observed in the coverage of the alterations detected in both panels, other authors have shown that a mean on-target read depth of 17-37x is required to identify 90% of the heterozygous SNV in the target regions (Meynert et al., 2013; Sims et al., 2014). Of note, 96.5% of the variants detected in our assay exceeded this range, and none of the reported variants had a depth of coverage <17x. However, when applying the NGS panel in the clinical setting, the depth of coverage obtained in some samples could be insufficient and it would be convenient to validate or re-sequence variants with low depth of coverage (<50x) before being reported (Cottrell et al., 2014; Jennings et al., 2017). In this regard, 14.6% of the known variants in PV1 would be omitted for having <10 supporting variant reads, in addition to the 48 variants that were excluded from the present analysis (25 for PV1 and 23 for PV2) and that were potentially real pathogenic variants. In clinical practice, these samples should be re-sequenced before being reported which would increase the cost and the response time of the test. Notably, the reduction of the regions captured in the PV2 partially overcomes this limitation. Although no differences were detected in the median depth of coverage for both designs due to the high variability among the variants, it was doubled in those cases sequenced twice and no case would have been removed from the analysis for presenting <10 reads in the validated cohort of PV2.

Finally, the sensitivity limits of both designs were below 10%, which is the limit established by many authors to report the data with confidence despite being limited when reporting subclonal variants or monitoring minimal residual disease (Cottrell et al., 2014; Hung et al., 2018). Nonetheless, the limit of variant frequency detection for each design was defined based on the lowest validated VAF carrying

≥10 supporting variant reads, therefore, it would be possible that the panel could call variants with lower allele frequency than the established. Thus, to conclude if this improvement in coverage depth is sufficient to increase the sensitivity of the panel, it would be necessary to analyze diluted samples with known variants at different frequencies to establish the analytical sensitivity of the panel. For the time being, we have demonstrated that the panel is useful to detect clonal SNV and indels, but additional analysis are needed to define its sensitivity for low-frequency variants and adjust the optimal depth of coverage for its application in the clinical practice.

2.2. Description of the variants found in each entity

A total of 461 variants were identified affecting the cellular pathways typically described in the different MBCN (figura 39). To understand the importance of these variants in the pathogenesis and their clinical impact on patients, the analysis of the variants detected by both panel versions in each entity was carried out individually.
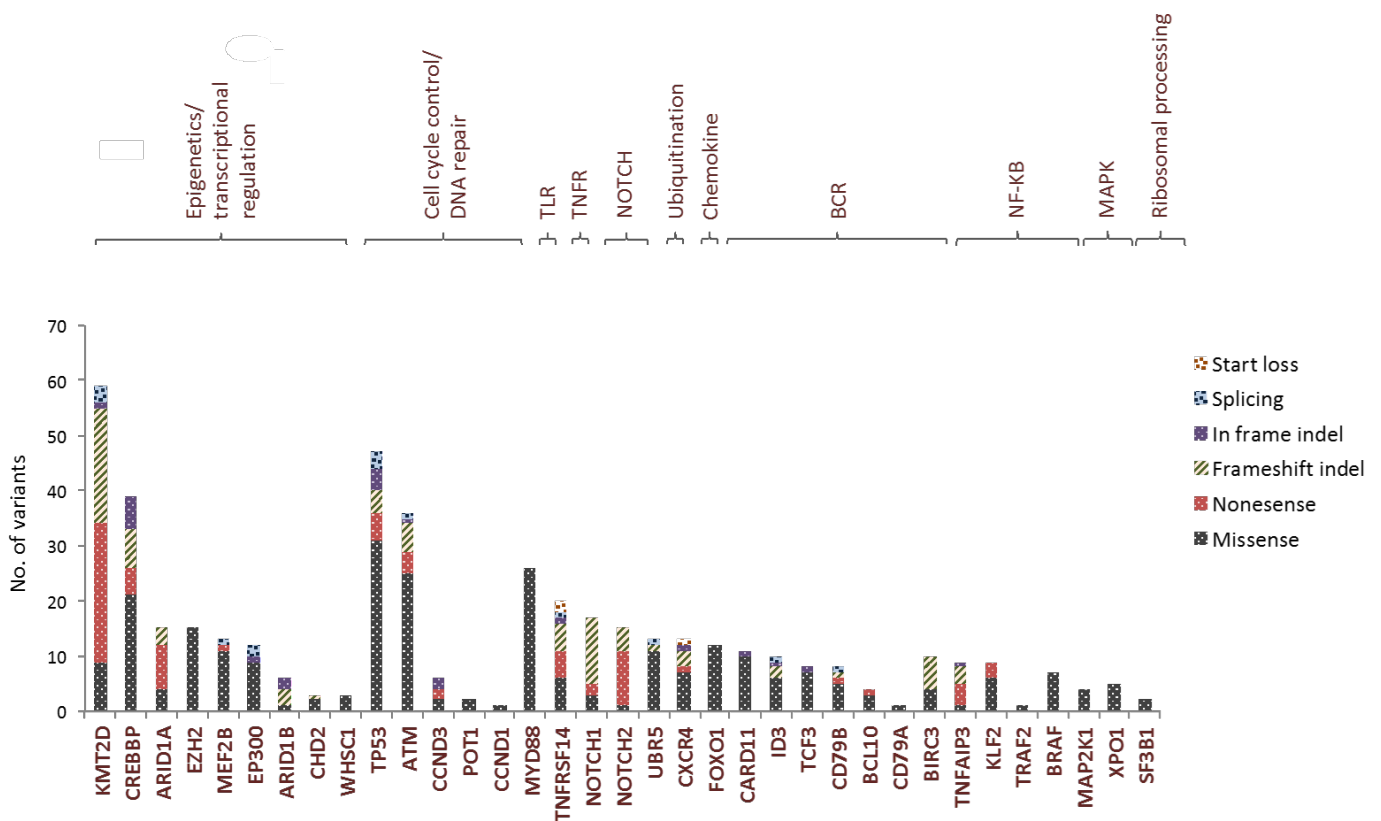
Figure 39. Summary of the SNV and small indel variants detected among the different MBCN and cellular pathways involved.

2.2.1. Mantle cell lymphoma

In all, nine out of the 12 MCL patients displayed variants. A total of 22 variants distributed in nine genes were found, with an average of two variants per patient (range: 0-5) (Figure 40).

Biologically, the chromosomal translocation t(11;14) is considered the genetic hallmark of MCL (Jares et al., 2007). However, additional genetic alterations and deregulated pathways are required to initiate and promote MCL lymphomagenesis (Pérez-Galán et al., 2010; Jares et al., 2012). These secondary genetic aberrations often affect molecular pathways that are involved in DNA damage response, cell proliferation, and cell survival, which is consistent with the results obtained in our assay. In fact, *ATM* and *TP53*, described as two of the most commonly disrupted genes in MCL, were also the most frequently altered genes in our cohort affecting 42% and 33% of the cases, respectively, and reflecting the importance of DNA repair in the pathogenesis of MCL. Furthermore, it is important to highlight that variants in *TP53* are associated with an aggressive disease course and inferior outcomes in different subsets of MCL and identify a phenotypically distinct and highly aggressive form of MCL with poor or no response to firs-line treatments (Beà et al., 2013; Eskelund et al., 2017). In addition to the *TP53* variants, alterations involving other genes with prognostic impact in MCL were also detected. Missense substitutions affecting *BIRC3* and *TRAF2* genes were found in one case each, leading to constitutive activation of the alternative NF-κβ pathway (Vogt et al., 2017). These aberrations are clinically relevant, since they are associated with resistance to the BTK inhibitor ibrutinib in *in vitro* studies (Rahal et al., 2014; Ahmed et al., 2016). In addition, a small frameshifting deletion in the PEST domain of *NOTCH1* gene, a transmembrane protein that functions as a ligand-activated transcription factor, was also detected in one case (Guruharsha et al., 2012). These *NOTCH1* variants are associated with adverse survival since NOTCH pathway activation induces proliferation and reduces apoptosis, suggesting that this signaling cascade might serve as a novel molecular target for MCL with activated NOTCH (Kridel et al., 2012). Futhermore, a patient harboring the C47R alteration in *CCND1* gene was also observed. Another variant in this amino acid (C47S) has been reported as one of the

three most frequent *CCND1* alterations in MCL together with E36K, and Y44D, and it has been observed that these variants increase protein stability and promote resistance to ibrutinib treatment (Zhang et al., 2014; Mohanty et al., 2016). However, despite the clinical impact provided by the variants described above, none of these genes has been incorporated into clinical practice for mutational screening. In addition to these variants, a small frameshifting deletion in the HECT domain of *UBR5* gene was also detected in one patient. These variants are MCL-specific and this gene is considered to play a crucial role in the pathogenesis in MCL as it participates in DNA damage response, cell cycle control in addition to E3 ligase function (Swenson et al., 2020). Finally, consistent with other studies, lesions in chromatin modifiers such as nonsense variants in *KMT2D*, and missense substitutions in *MEF2B* were also detected in our cohort in 25% (3/12) and 17% (2/12) of MCL cases, respectively (Beà et al., 2013; Zhang et al., 2014).
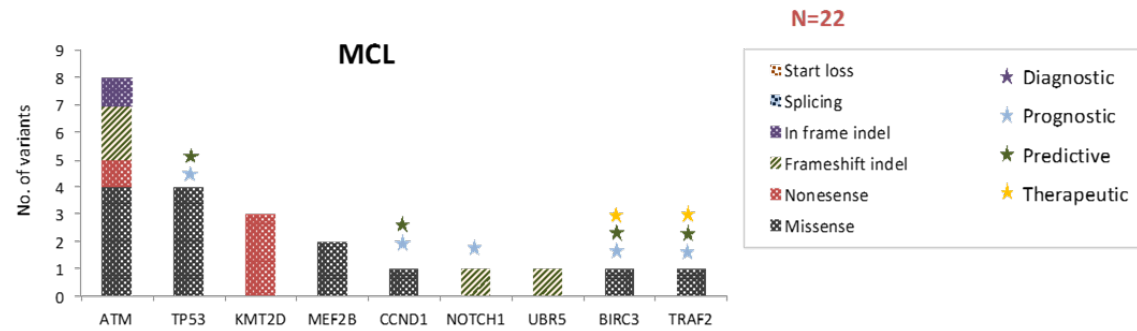
Figure 40. SNV and small indel variants detected in the MCL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.

2.2.2. Chronic lymphocytic leukemia

Globally, 30 of the 37 assessed CLL patients displayed genetic lesions, with an average of 2 variants per patient (range: 0-5). A total of 22 different variants were identified involving 15 genes (Figure 41).

It has been described that the typical genome of unselected CLL carries ~2000 molecular lesions, of which, however, only ~20 are nonsynonymous variants and only ~5 are gross structural abnormalities (Landau et al., 2015; Puente et al., 2015), which is consistent with the mutational burden detected among our patients. Besides, more than 40 recurrently altered driver genes have been identified in CLL (Landau et al., 2015; Puente et al., 2015). These recurrent lesions are not homogeneously spread across the CLL genome, but, rather, affect genes that can be integrated into a small set of pathways. Consistent with the data published in the literature, the most frequently altered genes in our assay were *TP53* (21/66, 31.8%), *ATM* (12/66, 18.2%), *NOTCH1* (8/66, 12.1%), and *BIRC3* (6/66, 9.1%). However, the frequency of these genes in our study was much higher than that reported in the literature (especially for *TP53, BIRC3*, and *ATM*), which could be explained due to the fact that our cohort was enriched in patients with a complex karyotype (22/37). With regard to the lesions in *NOTCH1*, among the eight variants detected, two affected different nucleotides within exon 34 (Q2394* and P2411Hfs*11), and were distinct from the hotspot dinucleotide deletion targeting the function of the C-terminal PEST domain (c.7541_7542delCT). Diverse NGS studies in CLL have revealed genetic lesions that affect different nucleotides in exon 34 (including frameshift and truncating variants) and which occur with lower frequency than the canonical delCT variant (12.5%) (Nadeu et al., 2016; D'Agaro et al., 2018).

In addition, other lesions have been identified that affect MAPK (*BRAF*), chromatin modification (*KMT2D, ARID1A, CREBBP,* and *CHD2*), transcription (*TCF3),* and ribosomal processing (*XPO1, SF3B1*) pathways. Moreover, three patients with missense variants in genes involved in the BCR signaling pathway (*CARD11*, and *FOXO1*) were detected. However, genes encoding for this pathway are usually not targeted by somatic variants in CLL, since BCR pathway activation in CLL usually results from contacts between tumor cells and antigens, which is influenced, among

others, by the somatic hypermutation (SHM) load of the rearranged immunoglobulin heavy-chain variable (*IGHV*) genes (Vardi et al., 2014). On the other hand, two missense variants outside the HECT domain of the *UBR5* gene, not previously reported in the literature, were also found in two patients (S2011P, and P1496S). *UBR5* is considered to play a crucial role in lymphomagenesis as it participates in DNA damage response, cell cycle control in addition to E3 ligase function, and is frequently altered in some B-cell neoplasm (Henderson et al., 2006; Benavides et al., 2013). However, is seldom disrupted in CLL (Puente et al., 2018).

Regarding the clinical association of the detected variants, several genetic lesions have been associated with clinical results in CLL. In addition to *TP53* and *ATM* variants, which are both known to confer poor prognosis, several high-throughput NGS studies have revealed recurrent variants within *NOTCH1*, *SF3B1*, and *BIRC3*, that were reported to be associated with poor clinical outcome, with higher frequencies in relapsing/treatment-refractory CLL and in Richter syndrome (Rossi et al., 2011; Schnaiter et al., 2013; Jeromin et al., 2014; Rosenquist et al., 2016). Furthermore, in a multi-center study conducted within ERIC, sequencing of *TP53*, *NOTCH1*, *SF3B1*, *BIRC3* and *MYD88* was performed in a large patient series, revealing that *TP53* and *SF3B1* variants, but not *NOTCH1* variants, remained as independent prognostic markers of shorter time to first treatment in multivariate analysis (Baliakas et al., 2015).

However, genetic lesions in *TP53* are currently the only with documented impact on therapy selection and patient management. Hence, the assessment of *TP53* status is essential for clinical decision-making and is currently the only gene analyzed in clinical practice despite the prognostic impact that other genes such as *ATM, BIRC3, NOTCH1* and *SF3B1* have shown (Hallek et al., 2008; Pospisilova et al., 2012; Rosenquist et al., 2016).
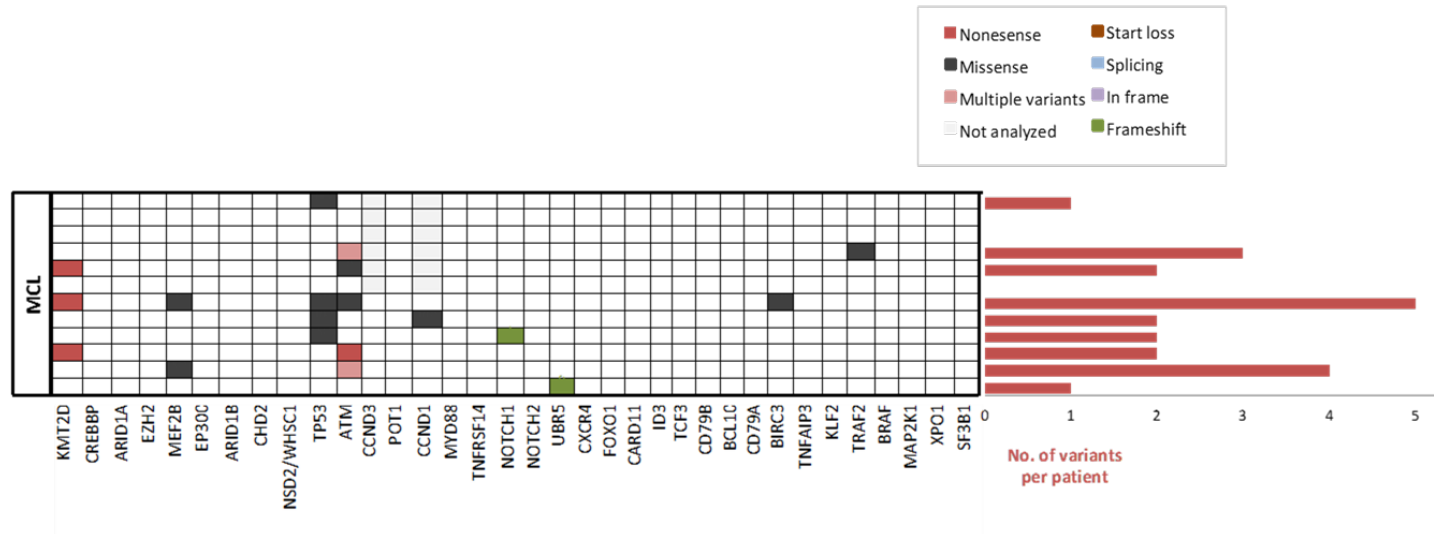
Figure 41. SNV and small indel variants detected in the CLL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.

197

2.2.3. Burkitt lymphoma

The five patients diagnosed with BL harbored a total of 20 variants among 12 genes, the average being four variants per patient (range: 0-8). Notably, in one case no variants were identified in the genes analyzed by the panel (Figure 42).

Burkitt lymphoma is a very aggressive mature B-cell lymphoma with a high degree of complexity in its coding genome (Grande et al., 2019; Panea et al., 2019), which is in line with the mutational burden detected herein. It is characterized by deregulation of the *MYC* gene through its translocation to one of the IG loci. However, recent studies have demonstrated the existence of regulatory pathways that cooperate with *MYC* in the lymphomagenesis, highlighting the importance of the BCR pathway (Sander et al., 2012). Indeed, variants that affect the transcription factor *TCF3* or interrupt its negative regulator *ID3* are highly recurrent and are considered diagnostic biomarkers in BL (Schmitz et al., 2012; Rosenquist et al., 2016). Our results are concordant with these data; although no patient presented aberrations in the *TCF3* gene, lesions in *ID3* were detected in 40% (2/5) of BL cases. Notably, all variants except one (c.300+1G>A) were located in the helix–loop–helix functional domain of the protein, what is consistent with data published (Love et al., 2012; Rohde et al., 2017).

On the other hand, we found some variants that have been more frequently associated with DLBCL, highlightening the close relationship between both entities (Bellan et al., 2009; Swerdlow et al., 2014). In this regard, we detected variants in the coiled-coil domain of *CARD11*, described in ABC-DLBCL and, in *FOXO1 gene,* associated with relapse and refractoriness in DLBCL (Trinh et al., 2013; Morin et al., 2016). In addition, one case showed a lesion affecting the *EZH2* Y641 histone methyltransferase residue, which is specific for the GCB-DLBCL subtype, and aberrations in other genes involved in epigenetic and transcriptional regulation, namely *CREBBP, MEF2B, KMT2D* and *EP300*. Of note *EP300* variant (c.3671+1G>T) was also observed in a patient with DLBCL. In addition, we detected variants that affected pathways such as cell cycle control (*TP53, ATM* and *POT1*) and ubiquitination (*UBR5*), which could be increasing genomic instability in these cases (Chang, 2013; Ramsay et al., 2013). Nevertheless, from a practical point of view none

of these genes are currently tested in the clinical management of BL patients (Rosenquist et al., 2016).
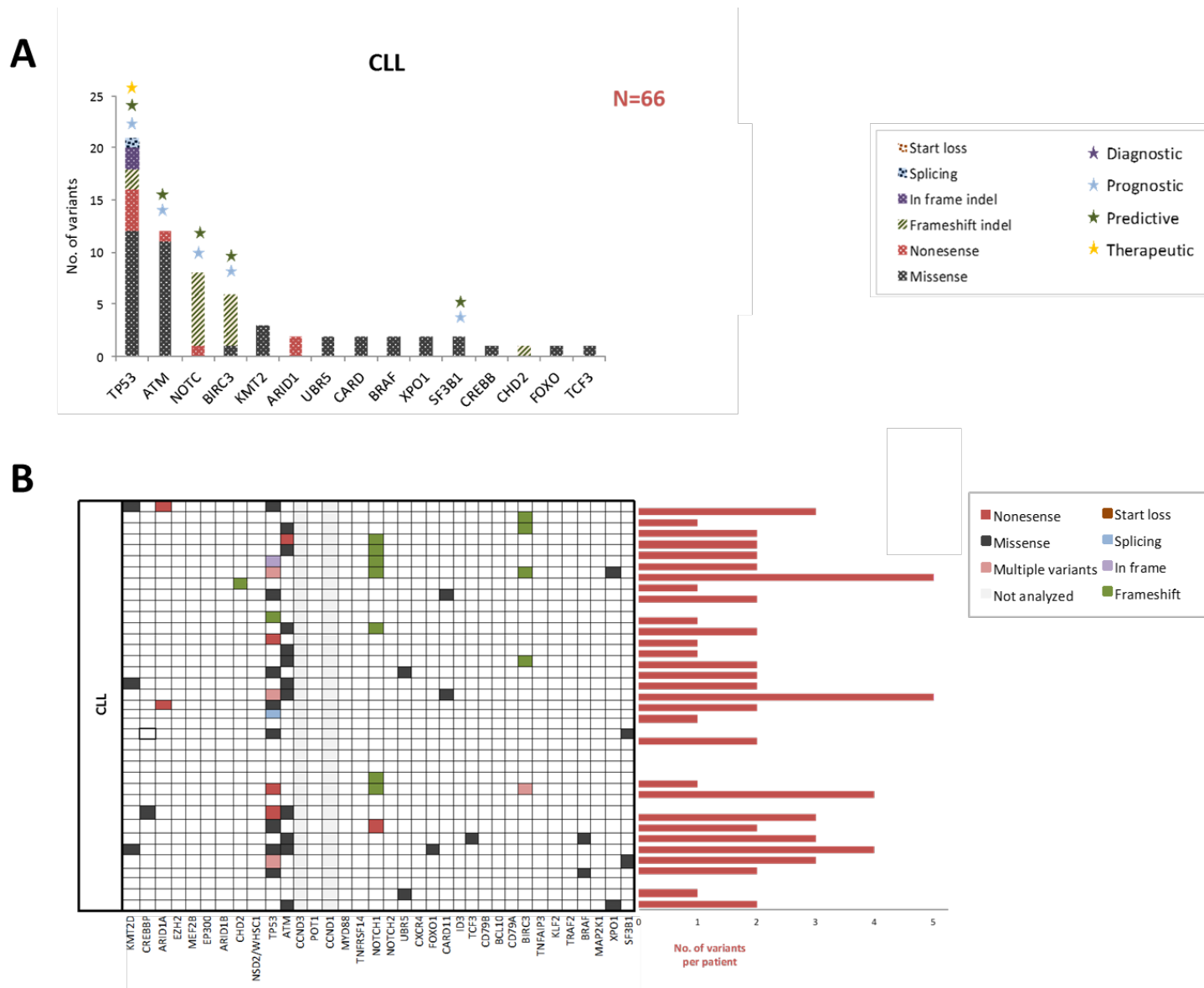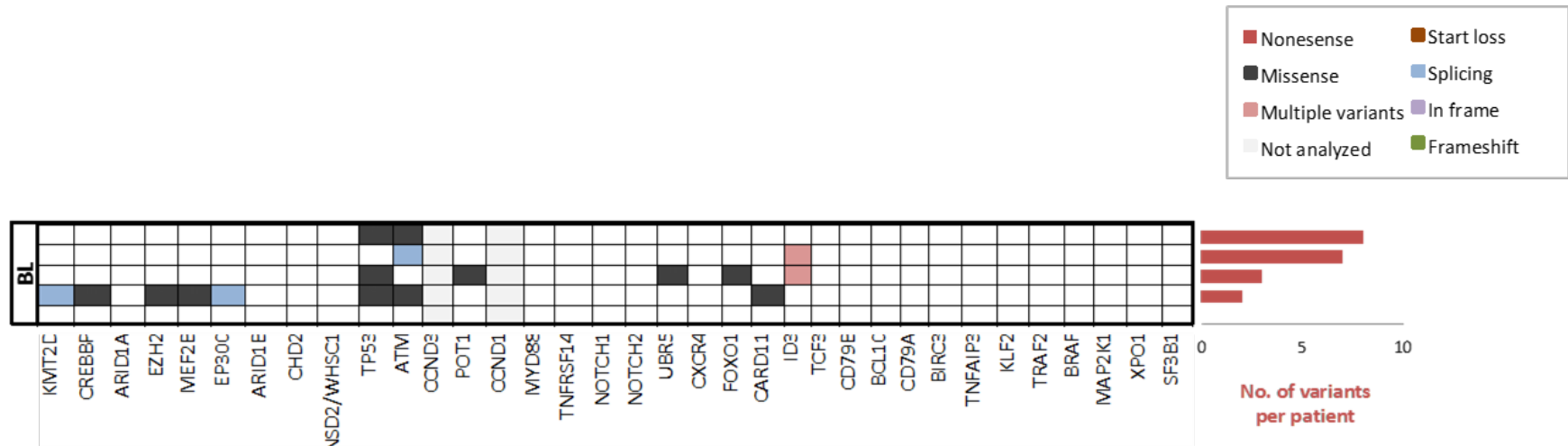
Figure 42. SNV and small indel variants detected in the BL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.
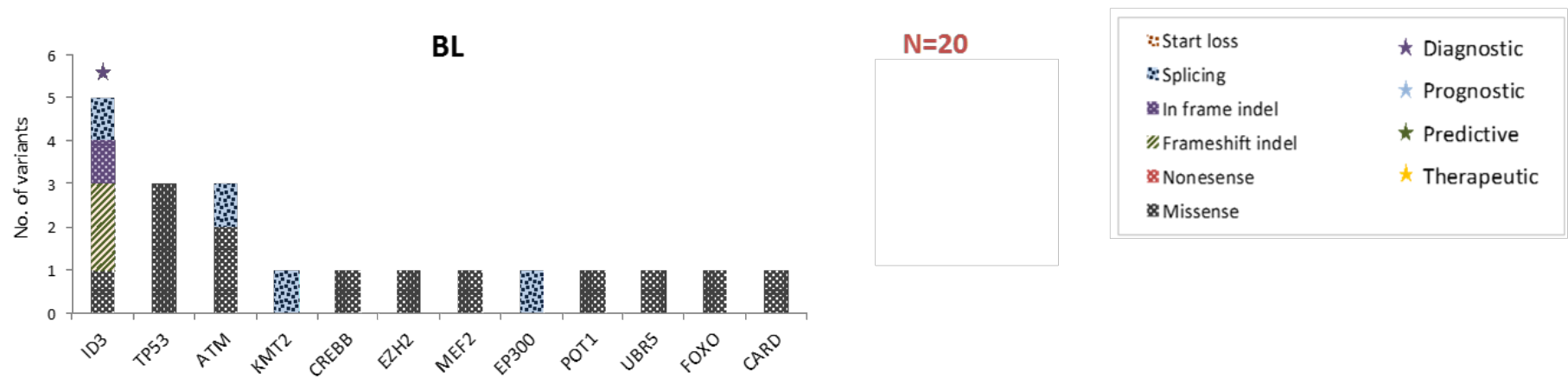
2.2.4. Diffuse large B-cell lymphoma

A total of 190 variants were found in the 55 patients diagnosed with DLBCL included in the assay. All but one displayed variants with an average of 3 variants per patient (range: 0-12) (Figure 43 and Figure 44).

The coding genome of DLBCL have shown a high degree of complexity compared with other B-cell malignancies, harboring on average between 30 and more than 100 lesions per case, with a great variability across patients (Morin et al., 2011; Pasqualucci et al., 2011; Lohr et al., 2012). These data were also reflected in our results, in which a variable number of variants in a large number of genes and pathways were observed among patients. Globally, genomic lesions disrupting the key cellular pathways involved in DLBCL were detected, which include alterations in epigenetic modifiers (such as *KMT2D, CREBBP, EZH2, MEF2B, EP300, ARID1B, ARID1A, WHSC1* and *CHD2*), genetic lesions leading to constitutive NF-κβ activity (such as variants in *TNFAIP3*, and *KLF2*) and chronic active BCR signaling (such as alterations in *FOXO1, CARD11, CD79B, ID3, BCL10*, and *CD79A*) (Pasqualucci et al., 2013). Most of the observed variants were related to these pathways and it is remarkable that 45/190 (24%) of the detected variants affected genes with prognostic and therapeutic value in DLBCL. Among them, *TP53, MYD88*, *CD79A, CD79B* and *EZH2* variants had been associated with poor outcomes (Lam et al., 2008.; Ngo et al., 2011; Pasqualucci, 2013; Rovira et al., 2016). Besides, *MYD88* variants have been shown to promote NF-κβ and JAK–STAT3 signaling in activated B-cell-like DLBCL, and patients harboring L265P variation may benefit from therapies targeting IRAK4 alone or in combination with agents targeting the BCR, NF-κβ, or JAK–STAT3 pathways (Davis et al., 2010; Milhollen et al., 2010).

Furthermore, variants in genes involved in other pathways were also detected. For instance, *CXCR4* is a key player in the dissemination of DLBCL, and variants in this gene have been associated with disease progression and poor survival, and may also be an indicator of metastasis. These variants constitute a prognostic and therapeutic marker since inhibition of the *CXCR4* expression could reduce metastasis in DLBCL (Du et al., 2019).

In addition, aberrations affecting the PEST domain of the *NOTCH1* and *NOTCH2* genes are associated with a poor prognosis in other mature B-cell neoplasms. Moreover, variants affecting the *NOTCH1* gene have been associated with a poor prognostic DLBCL subtype (Liu & Barta, 2019).

Nevertheless, at the present there is no consensus on which biological prognostic variants should be routinely assessed, and therefore, these data have not yet been incorporated into the clinical management of patients with DLBCL, nor analyzed in clinical practice (Perry et al., 2012; Swerdlow et al., 2016).
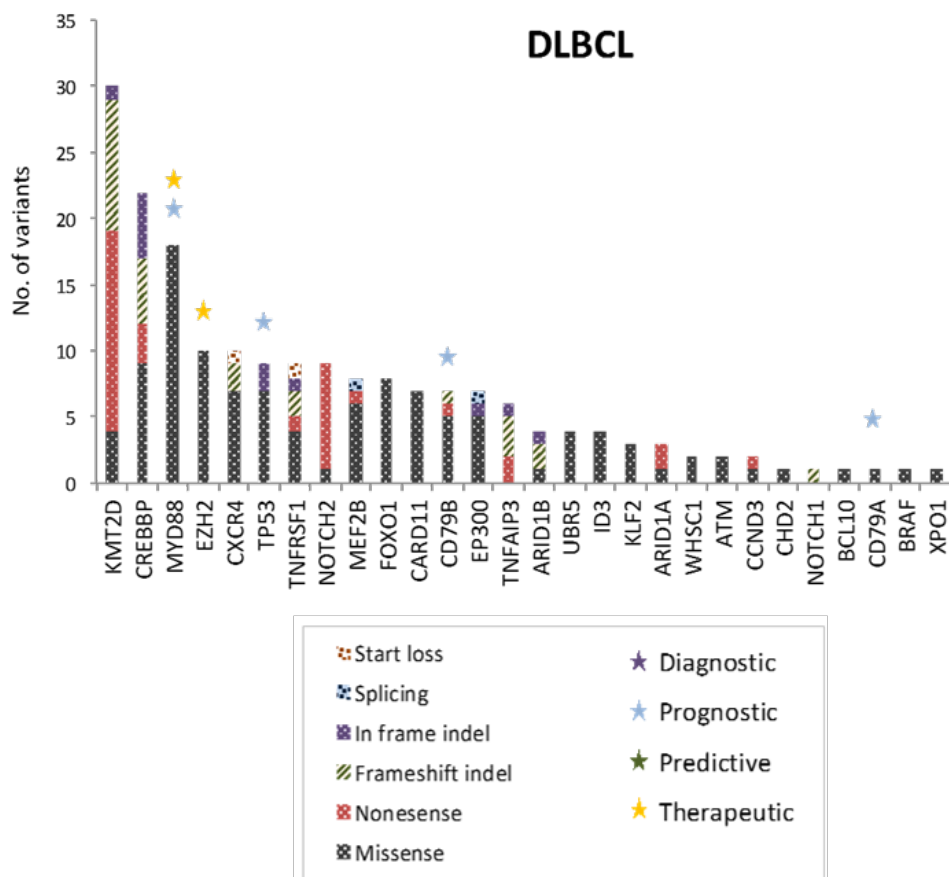
Figure 43. Distribution of SNV and small indel variants detected in the DLBCL patients among genes, and their predicted clinical impact.
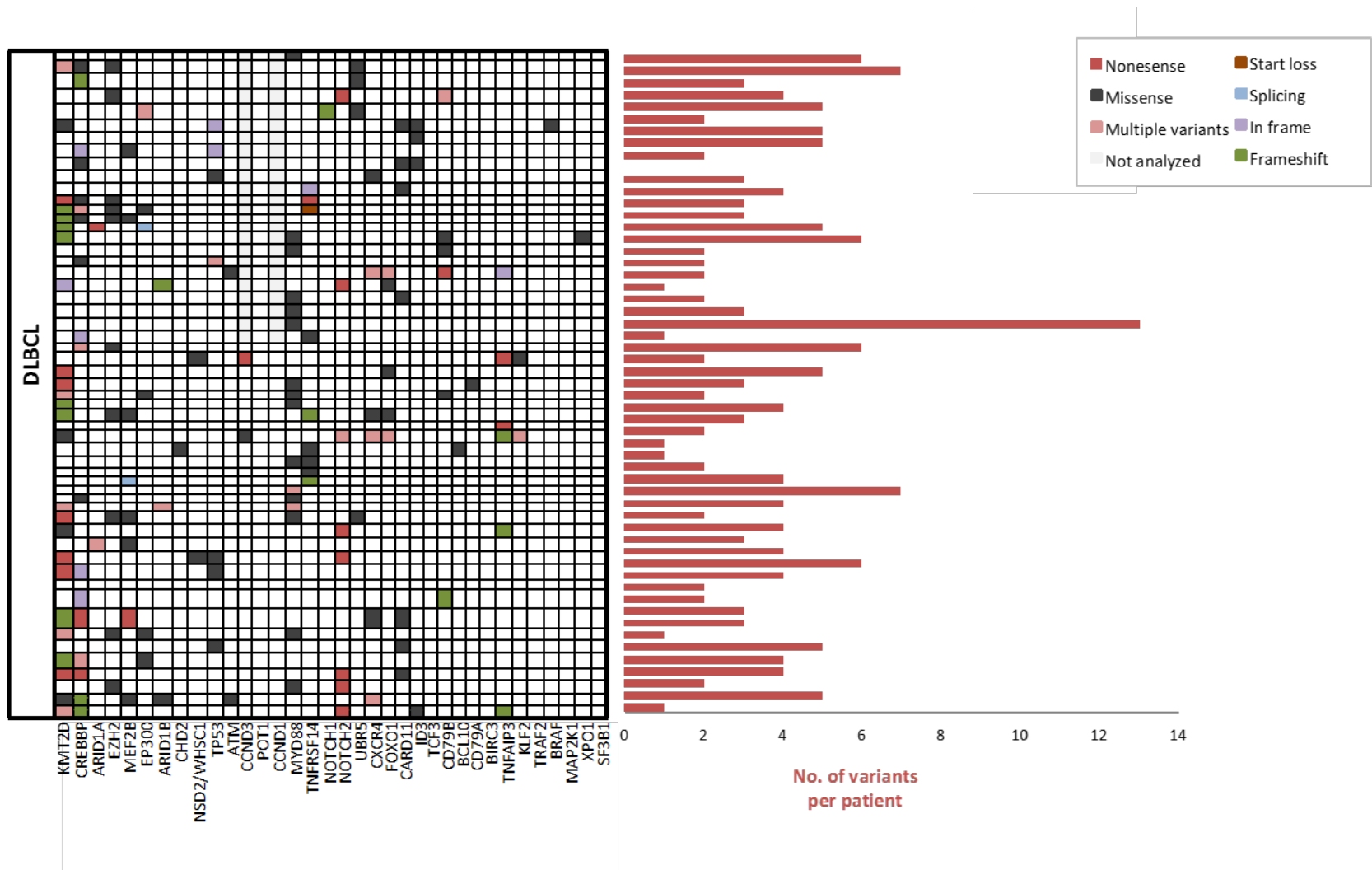
Figure 44. Distribution of SNV and small indel variants detected in the DLBCL patients among patients. Each patient is represented in a row and assessed the genes in columns.

2.2.5. Follicular lymphoma

After the analysis of the 14 FL sequenced patients, a total of 61 variants affecting 15 genes were identified. All patients showed variants, with an average of 4 variants per case (range: 1-7). These variants are detailed in Figure 45.

Although follicular lymphoma is an indolent mature B-cell neoplasm, the average number of variants detected per patient was relatively high. There is accumulating evidence that other genetic hits than t(14;18) are required for complete transformation to FL (Vaux et al, 1988; Kridel et al, 2012). Disruption of histone-modifying enzymes by genetic lesions has been recognized as a central hallmark of FL, arising in nearly all patients (Morin et al, 2011; Pasqualucci et al, 2011). Consistent with these data, 67% of the variants detected in our cohort affected genes involved in epigenetic and transcriptional regulation, and these alterations were present in almost all the patients (13/14 showing aberrations in *KMT2D, CREBBP, ARID1A/B, EZH2, MEF2B, EP300*, or *CHD2*). Interestingly, there was a high prevalence of concurrent lesions in more than one epigenetic modifier (observed in ~70% of cases), suggesting a potential cooperation among enzymatic activities to affect the expression of key regulators of B-cell homeostasis. However, it is noteworthy that 75% of these alterations affected the *KMT2D* and *CREBBP* genes, which are early oncogenic events and are believed to be present in the so-called common progenitor cell population, which contains tumor-initiating cells and acts as a disease reservoir in relapse and histological transformation (Green et al, 2015; Horton et al, 2017). In addition to their relevance in the lymphomagenesis, *EZH2, ARID1A, MEF2B, EP300, FOXO1, CREBBP* and *CARD11* have been reported to impact the prognosis of FL patients and have been included in the Follicular Lymphoma International Prognostic Index (FLIPI) (Pastore et al, 2015). Furthermore, the mutational status of *EZH2* and *CREBBP* could determine the choice of therapy in FL patients (Jiang et al, 2017; Morschhauser et al, 2017).

On the other hand, lesions affecting genes involved in other pathways have also been identified, such as *TNFRSF14*, which is frequently disrupted in this pathology, and whose variants are also associated with a worse prognosis. In addition, a variant in the *CD79B* gene was detected in one patient who could benefit from the BTK inhibitor (Jerkeman et al, 2017). Furthermore, lesions in genes involved in cell cycle control (*TP53, ATM* and *CCND3*) were also detected. Loss of proliferation control is a hallmark of cancer and is frequently seen in aggressive B-cell malignancies like MCL and DLBCL, being a rare event in FL, where it is mainly related to disease transformation (Morin et al, 2011; Okosun et al, 2014; Pasqualucci et al, 2014). In fact, the patient with aberrant *TP53* presented a transformation to DLBCL during follow-up. For the rest of the patients, no follow-up was performed, so it is not possible to determine whether there was progression.

However, it is surprising that no gene is currently analyzed in clinical practice, although a large number of biomarkers associated with both prognosis and treatment have been reported and could be of great help in clinical management of these patients.

Figure 45. SNV and small indel variants detected in the FL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.

2.2.6. Splenic marginal zone lymphoma

A total of 80 variants were identified in 36 of the 38 SMZL patients analyzed. These variants involved 23 genes, with an average of 2 variants per patient (range: 0-6) (Figure 46).

The genetic hallmark of SMZL is predominantly associated with alterations affecting the signaling pathways that regulate marginal zone differentiation, including NOTCH, NF-κβ, and BCR pathways (Arcaini et al., 2016). In line with these data, we detected variants affecting genes involved in these pathways including *NOTCH1* and *NOTCH2* (NOTCH pathway), *KLF2, TNFAIP3,* and *BIRC3* (NF-κβ pathway), and *TCF3, BCL10, FOXO1* and *CARD11* (BCR pathway). In addition to the pathogenic implications, it has been reported that variants in *NOTCH2* and *KLF2*, in addition to variants in *TP53*, which is the most frequently detected gene in our study (21% of cases, 8/38), confer a poor prognosis (Parry et al., 2015; Piva et al., 2015). Moreover, a variant in *TNFAIP3* gene was also detected in one patient, which is associated with an increased risk of high-grade transformation. Conversely, variants in *MYD88* have been shown to confer a good prognosis and are associated with longer overall survival (Parry et al., 2015; Campos-Martín et al., 2017). Interestingly, two of the seven cases with an aberrant *MYD88* in our study presented concomitant aberrations in *TP53*. Despite the prognostic value of these variants in SMZL, they do not impact on the therapeutic decision and, therefore, are not currently analyzed in clinical practice. On the other hand, in addition to their prognostic value in SMZL, alterations in PEST domain of *NOTCH2* have been suggested as diagnostic markers since they have been reported as almost specific for SMZL (Kiel et al., 2012; Rossi et al., 2012). However, variants located in the PEST domain of *NOTCH2* were also detected in 14% of patients with DLBCL of our cohort, one of them showing a variant in the same nucleotide position as a SMZL patient (c.6853C>T and c.6853delC, respectively). Although not being frequently altered in DLBCL, it has been described that those cases unclassified are enriched in *NOTCH2* lesions compared with ACB and CGB subtypes (Schmitz et al., 2018).

Apart from genes implicated in marginal zone development, a wide variety of alterations in genes involved in other pathways were identified, such as chromatin

remodeling and transcriptional regulation (*KMT2D, ARID1A, EP300, EZH2*, and *ARID1B*), cell cycle control and DNA repair (*ATM* and *CCND3*), ubiquitination (*UBR5*), chemokine signaling (*CXCR4*), MAPK signaling (*BRAF* and *MAP2K1*) and ribosomal processing (*XPO1*). These data suggest a great heterogeneity in the SMZL genome, which could explain the prognostic heterogeneity shown in the patients. Indeed, despite being an indolent lymphoma, 30% of patients have worse outcome, including up to 10% of patients that undergo transformation to DLBCL (Conconi et al., 2015). Although the molecular pathogenesis of many lymphoma entities has been elucidated in detail, relatively little is known about the genetic lesions associated with SMZL, which limits the present understanding of its pathogenesis and hampers the possibility of diagnosis and classification based on genetics, one of the mainstay criteria adopted by the WHO Classification for the diagnosis of B-cell lymphoma (Swerdlow et al., 2016).
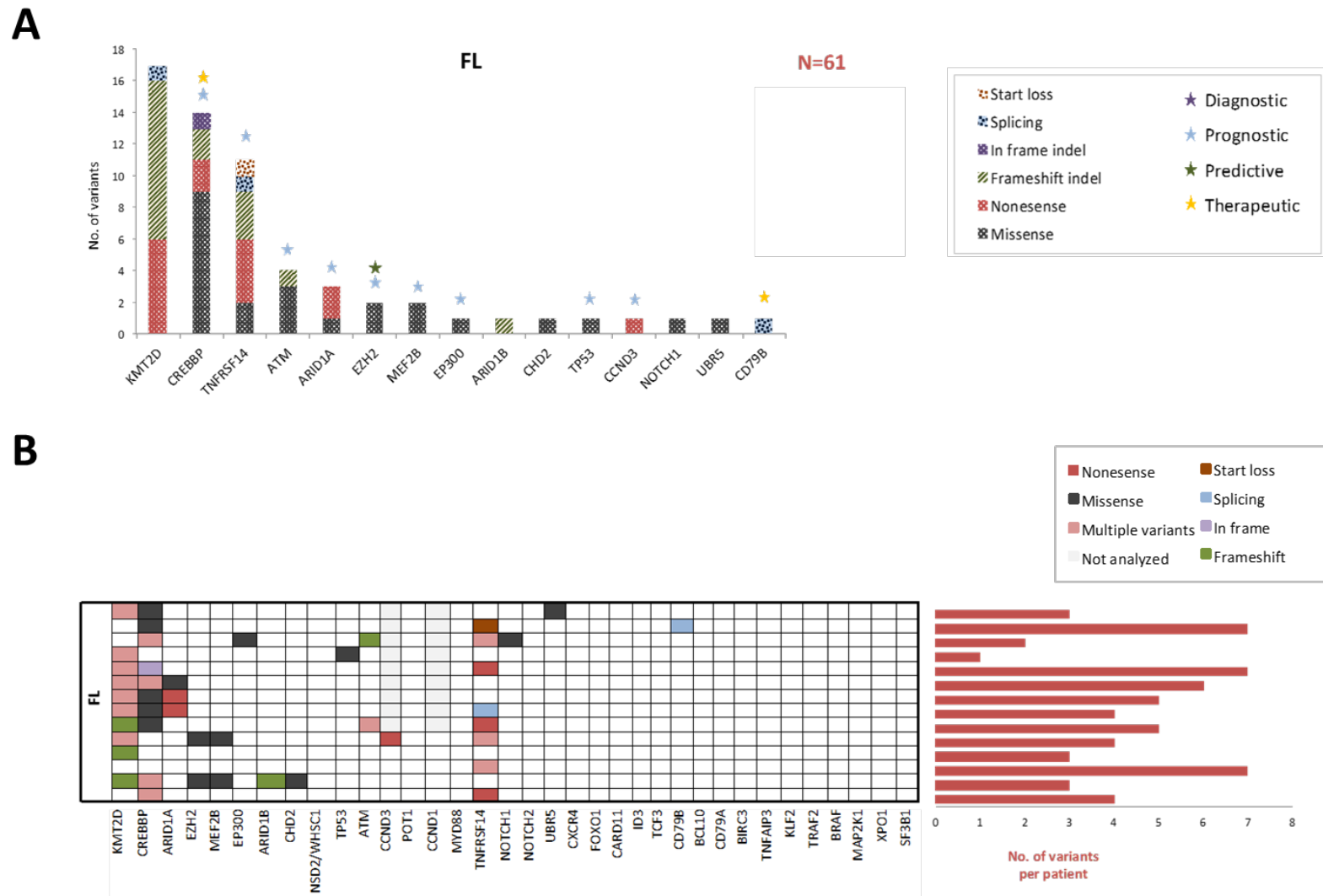
Figure 46. SNV and small indel variants detected in the SMZL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.
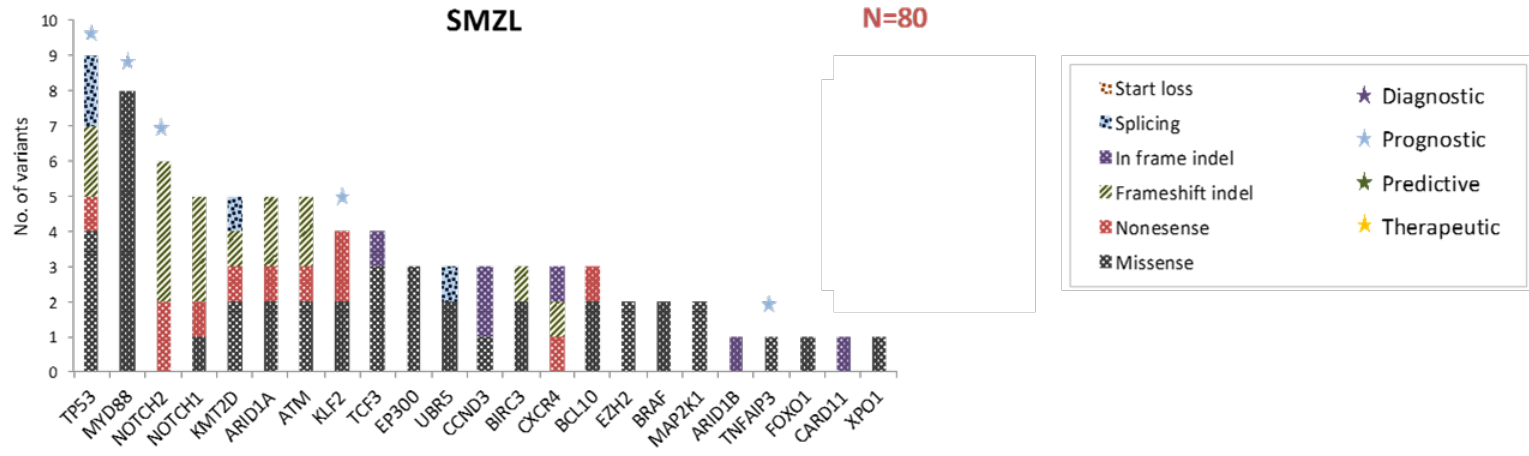
2.2.7. Other pathologies

MALT lymphomas represent a heterogeneous group with a large number of different genetic alterations depending on the site of origin (Streubel et al., 2004). Although could not analyze differences in the mutational profile of the patients according to the origin of the lymphoma, since the analysis of only four patients was not large enough to make a comparative analysis, this heterogeneity was visible among patients. While three of the patients presented one or no variants, the fourth patient harbored eight alterations in the genes analyzed by the panel (Figure 47). As expected, most of these genetic lesions affected genes related to the NF-κβ signal pathway, causing constitutive activation of this pathway, which is in line with that described in the literature (Kuper-Hommel & Van Krieken, 2012). Interestingly, there was a case with a single lesion affecting the *POT1* gene, recently described in MALTL and whose inactivation has been shown to accelerate tumorigenesis (Pinzaru et al., 2016; Cascione et al., 2019). However, in clinical practice this information is missing, since no variants detected in MALTL has shown to have a diagnostic, prognostic, or predictive impact and therefore, the mutational screening is not performed in any gene for the clinical management of MALTL patients.

In contrast, regarding Hairy cell leukemia, the *BRAF* V600E alteration was identified as a diagnostic marker. However, while this variant is detected in 90% of cHCL cases, it is absent in patients with vHCL. Therefore, it can also be used to differentiate these two entities that show differences in their clinical course and could be clinically relevant (Tiacci et al., 2011). Furthermore, although cHCL and vHCL show differences in their morphology and immunophenotype, which may guide their differential diagnosis, a subgroup of cHCL has been described morphologically and immunophenotypically very similar to *BRAF* V600E positive cHCL, but does not display this variant. This subgroup is associated with lesions in *MAP2K1*, in the same way as vHCL, and IGHV4-34 usage, which has shown to be a poor prognostic factor in HCL. In our assay, a total of 11 variants affecting eight genes were detected in the four sequenced HCL patients, three with cHCL and one with vHCL (Figure 48). As expected, we did not detected lesions in *BRAF* in vHCL, but neither were they identified in one of the patients diagnosed with cHCL in whom mutational screening

had not been performed in clinical practice. Indeed, this patient harbored a variant in *MAP2K1*, in addition to lesions in *WHSC1, ARID1A*, and *XPO1* genes. Although the IGHV data were not available, these results suggest that this patient could correspond to the IGHV4-34 cHCL subgroup. Furthermore, consistent with the data published, another variant in *MAP2K1* gene was detected in the patient diagnosed with vHCL. Both, IGHV4-34 and vHCL are independently associated with poorer overall survival and lack of response to treatment (Arons et al., 2011). For this reason, performing mutational screening of both *BRAF* and *MAP2K1* genes is essential in clinical practice and should always be carried out, since, although the morphology and immunophenotype seem to help in the differential diagnosis of cHCL and vHCL, they do not help to differentiate the IGHV4-34 cHCL subgroup.

Finally, regarding the case diagnosed with splenic diffuse red pulp lymphoma, although it did not show lesions in genes described as frequently altered in this entity, three variants were detected in genes involved in pathways described in this pathology, such as cell cycle regulation (*ATM*) and BCR signaling (*FOXO1* and *TCF3*) pathways (Curiel-Olmo et al., 2017; Jaramillo Oquendo et al., 2019). Furthermore, the absence of alterations in *NOTCH2, BRAF* and *MAP2K1* genes aided in the differential diagnosis with SMZL, cHCL and vHCL (Curiel-Olmo et al., 2017).
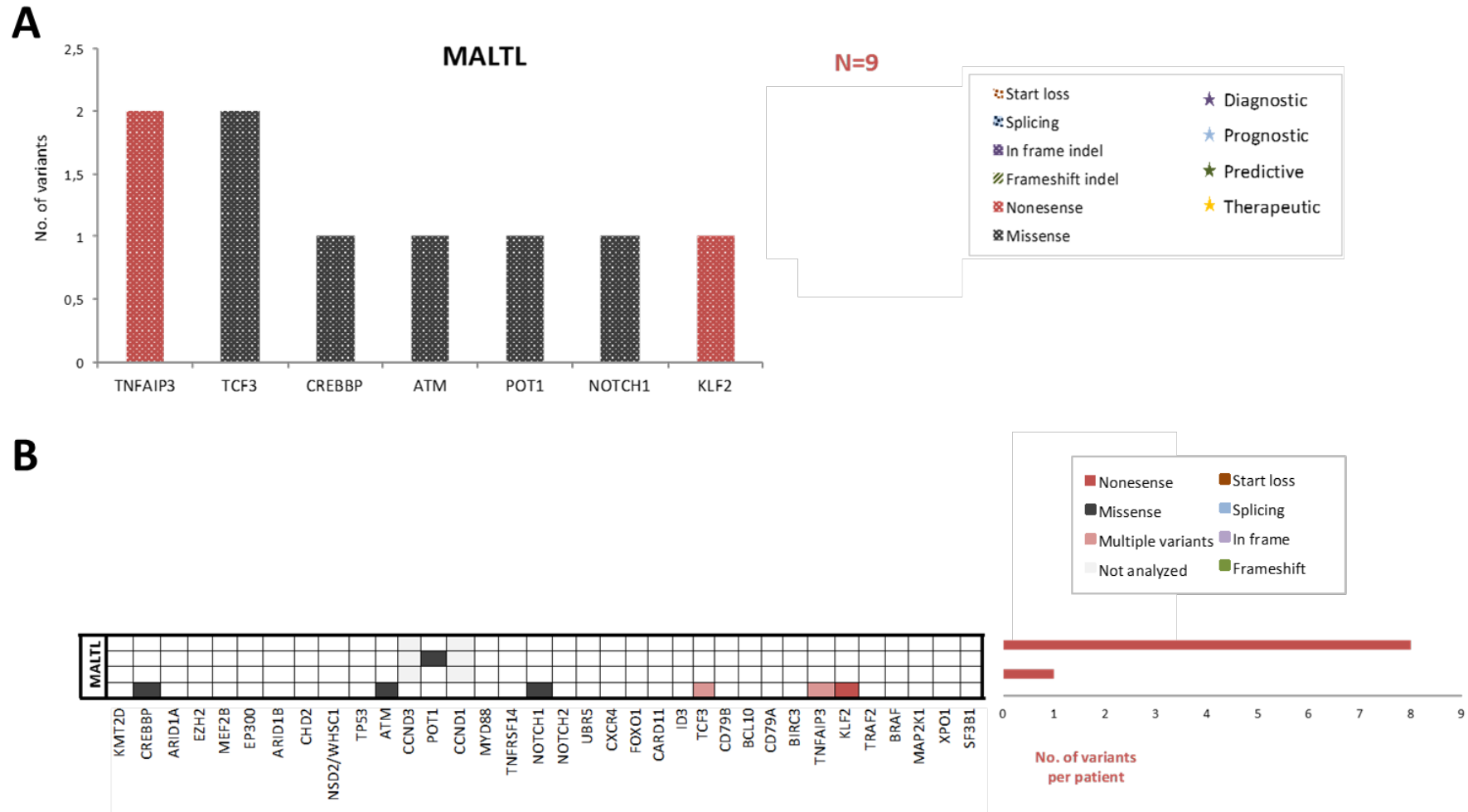
Figure 47. SNV and small indel variants detected in the MALTL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.
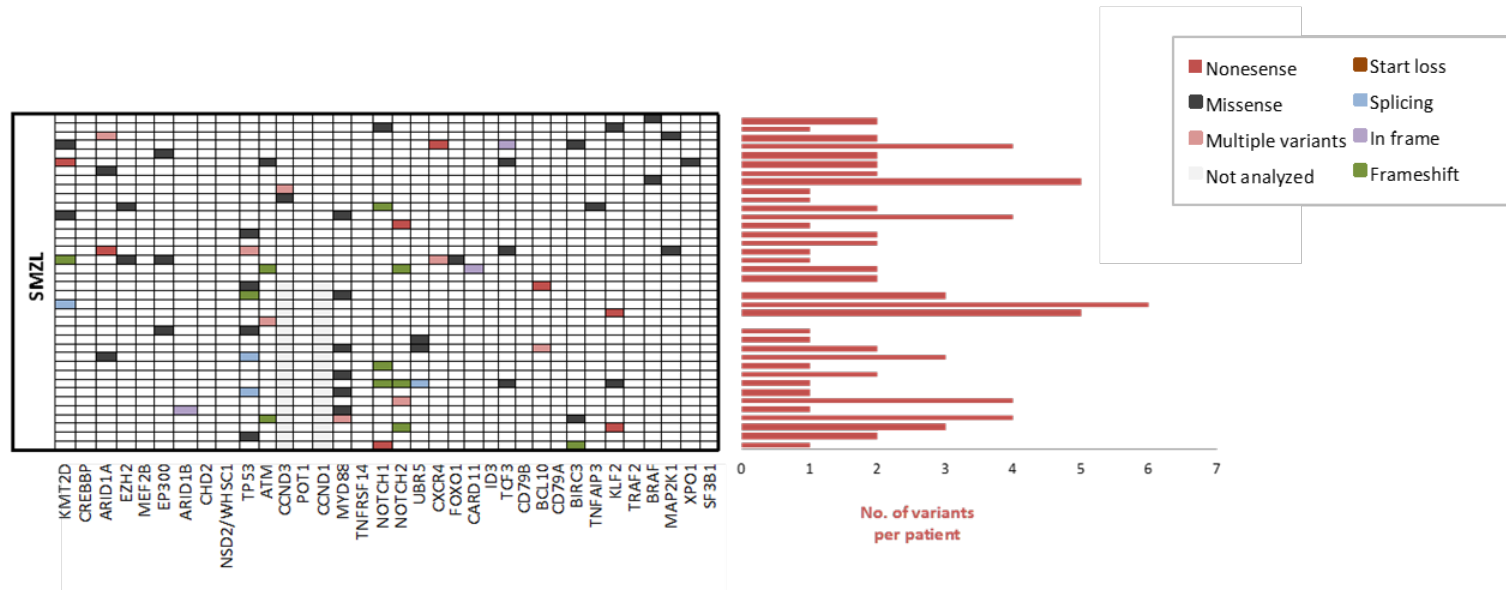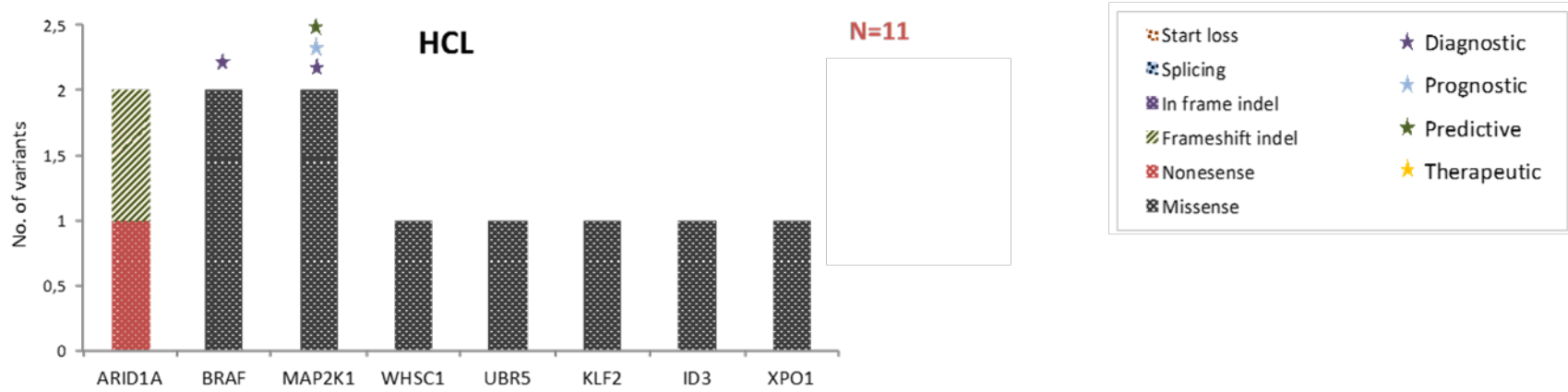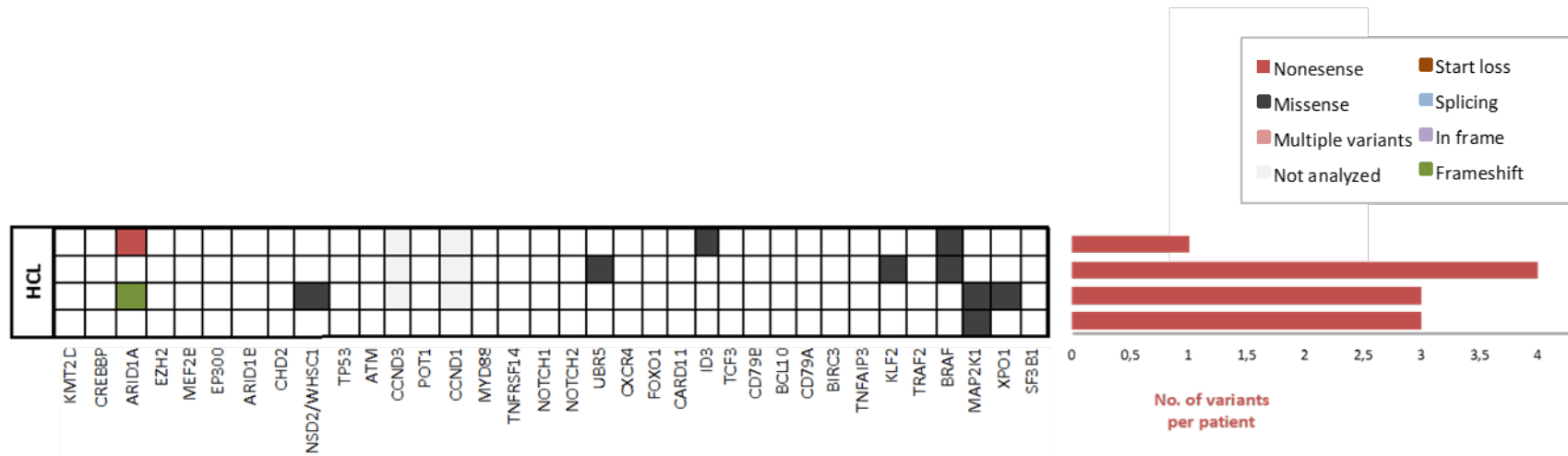
Figure 48. SNV and small indel variants detected in the HCL patients. (A) Distribution of the detected variants among genes, and their predicted clinical impact; (B) Detailed distribution of variants among patients, each patient is represented in a row and assessed the genes in columns.

2.3. Interpretation of variants: Consequences of using a non-hotspot panel

Transforming the vast amount of data generated by NGS into meaningful clinical information is a complex and often challenging process. The larger the area of the genome that is sequenced, the greater the probability of finding rare or novel variants that need to be interpreted. While novel or rare variants that cause a frameshift or introduce a stop codon will probably be pathogenic if that mechanism has been specifically described in that gene, missense variants present more difficulties when interpreting. The developed panel does not focus exclusively on the hotspot variants described in the literature, but rather analyzes all the exonic regions of the genes included in the design, making the variant analysis and interpretation more complicated.

In the present study, variants were classified based on the five-level classification system proposed by the ACMG guidelines for interpretation and classification of germline alterations, and only variants classified as pathogenic, likely pathogenic or of uncertain significance, were included in the analysis (Richards et al., 2015). These guidelines assign strength of evidence for various criteria regarding a particular variant and rules for combining all the criteria to classify that variant into five pathogenic categories. However, differences in variant interpretations have been shown among practitioners, which may be due to differences in the approach used to search, curate, select, and weigh the evidence. Furthermore, there is not the same published evidence, in terms of quality and availability, for all the variants (Yohe et al., 2015; Amendola et al., 2016). These limitation was also reflected in our assay, in which, while there was abundant evidence for certain variants or genes (such as T*P53, MYD88, EZH2, NOTCH1*, *NOTCH2*, and KMT2D), only limited or conflicting information was available for others (*FOXO1, ATM, TCF3, UBR5* and *CARD11*, for instance), making its classification more subjective (Figure 49). Moreover, in addition to the hotspot variants described in some genes (such as *ARID1A* for HCL, *XPO1, MYD88, BRAF, CARD11* and *EZH2*), 23 additional variants were identified, of which 15 could be classified as likely pathogenic as they were nonesense or frameshift variants (n=2), or, because, despite being missense variants, they were identified in genes or pathologies for which more published data were

available (n=13). However, the rest of the detected variants (n=8) remained as variants of uncertain significance, as they were missense variants and there was insufficient published evidence available for their classification. Therefore, overall, only ~60% of the 461 variants included in the analysis could be classified as pathogenic or likely pathogenic because they presented high or moderate evidence of pathogenicity, while the remaining ~40% corresponded to variants of uncertain significance that could not be classified as pathogenic or benign due to insufficient information currently available (Figure 49).
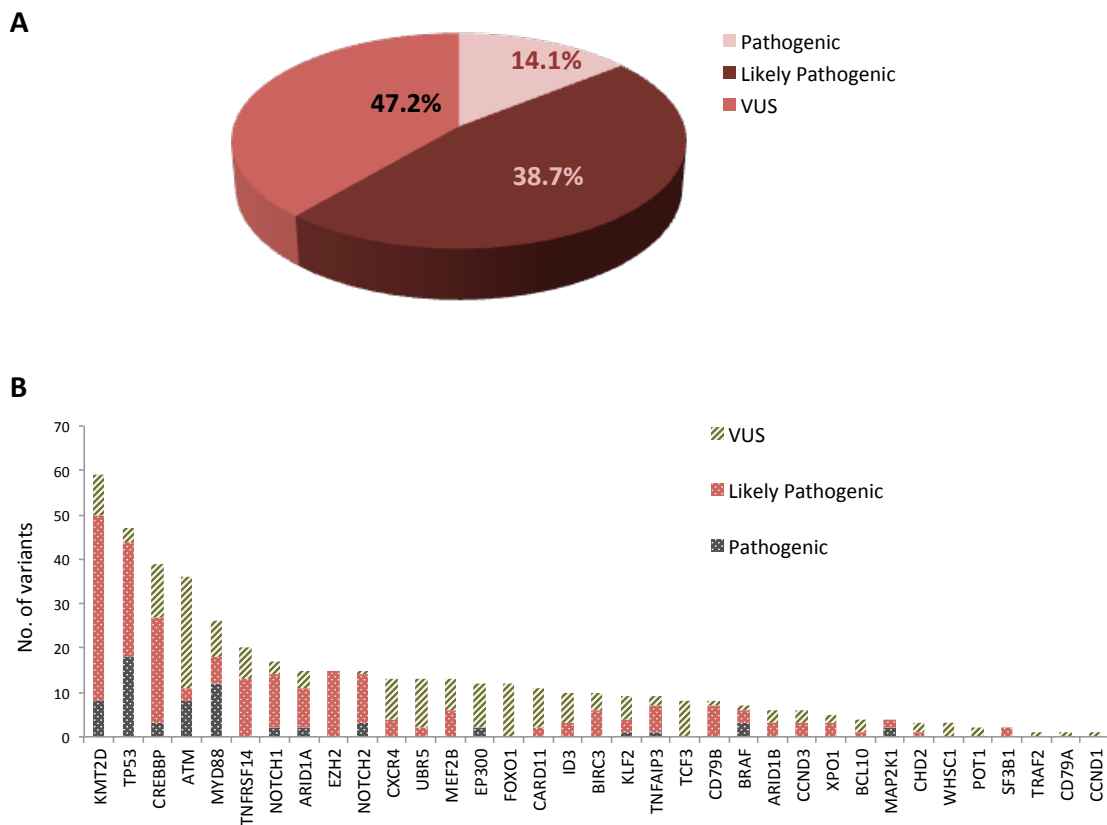
Figure 49. Classification of SNV and small indel variants detected in our assay. (A) Total frequency of each type of variant: pathogenic, likely pathogenic and variant of uncertain significance (VUS); (B) Number of pathogenic, likely pathogenic and VUS variants detected in each gen

Of note, due to a lack of evidence supporting the clinical utility of these variants, reporting of VUS is not considered a necessary component of tumor sequencing analysis (Strom, 2016). Whether and how to report VUS is highly laboratory and test dependent. However, due to their high frequency, and the fact that the lack of evidence to support their pathogenicity does not mean that they cannot be pathogenic variants, not incorporating them in the analysis together with the other variants with proven pathogenicity could mask the genetic complexity of the tumor.

On the other hand, in 2017 guidelines that provide a systematic approach for the classification, interpretation and reporting of somatic variants were published (Li et al., 2017). These guidelines emphasize on clinical significance and actionability of variants, and therefore, improve the clinical interpretation of sequencing reports that are based on therapeutic, diagnostic, and prognostic information. However, this approach is less useful for classifying somatic variants that may contribute to oncogenesis but whose actionability has not already been tested in clinical trials. Furthermore, the actionability of a variant may change over time, requiring the need for accurate variant designation supported by current evidence. For these reasons, we decided not to use these guidelines to classify the variants detected in our assay.

Finally, since the normal tissue of each patient was not analyzed, as an indirect method to avoid reporting benign germline variants, those alterations present in >1% allele frequency in the population databases were eliminated. However, since somatic variants can equal germline heterozygous and homozygous variant frequencies, respectively, variants with VAF at almost 50% or 100% should be considered potentially germline when non-tumoral tissue is analyzed.

3. Rearrangements analysis

3.1. Optimization of the rearrangements analysis algorithm

Samples were analyzed using three different callers (Lumpy, Smoove, and DELLY) in addition to the analysis carried out by the DREAMgenics company. Data obtained with these four methods were compared, individually, with the results previously obtained by routine techniques, FISH mainly (PV1= 57 patients, 124 rearrangement calls; PV2= 51 patients, 103 rearrangement calls), to refine the filtering of the data generated and assess the sensitivity and specificity of each caller in detecting these abnormalities. Herein, the optimal analysis algorithm for the detection of rearrangements will be established, while the analysis and characterization of the identified rearrangements of each target gene will be described in the following points.

Regarding PV1, the first analysis was carried out with the Lumpy tool. Although results generated by this caller displayed relatively high sensitivity (48/53 known translocations detected, 90.6%), it was striking that it detected rearrangements in most of the target genes for all patients, regardless of their diagnosis. In this regard, this caller generates an extremely high proportion of false positive results, with a specificity of only 22.5% and showing positive results in almost 50% (153/308) of the rearrangements not assessed in routine procedures, which were expected to be enriched in non-altered cases (Table 52). Moreover, most of the rearrangements presented two or more partners and several of them were found to be repeated at high rates between different target genes and patients. When focusing in these recurrently detected read-pairs, we identified seven highly repeated rearrangement partners, involving 884 rearrangements in our raw database, which had been identified among the top 16 rearrangement hot spot regions described in a study carried out by *Xu et al* in which they analyzed the genome of 1,481 healthy subjects from the 1000 Genomes Project (Xu et al., 2014). Thus, those rearrangements with hot spot regions were considered as non-pathogenic and were filtered from the initial output in order to reduce the false positive detection rate (Table 53). However, the output generated by Lumpy was still very complex, with a large

number of nonspecific rearrangements and an unacceptable specificity of 53.5%. For this reason, the use of Lumpy as a single caller for the detection of the rearrangements was discarded and its analysis was not performed on PV2.

Table 52. Rearrangements results obtained by the Lumpy caller.

Target	Routine		Not filtered	Filtered			
			Positive	Positive	Negative	Sensitivity	Specificity
BCL2	Positive	20	15	14	6	70,0%	71,4%
	Negative	14	5	4	10		
	Not tested	74	3	1	73		
BCL6	Positive	10	10	10	0	90,9%	45,2%
	Negative	32	31	18	14		
	Not tested	66	48	15	51		
MYC	Positive	16	16	13	3	81,3%	54,2%
	Negative	24	19	11	13		
	Not tested	68	45	19	49		
CCND1	Positive	6	6	6	0	100,0%	50,0%
	Negative	2	1	1	1		
	Not tested	100	57	13	87		

Routine techniques included FISH and CBA

Table 53. Annotation of the hot regions of translocation detected by Lumpy and which were described among the 16 main hot regions of translocation in healthy subjects from the 1000 Genomes Project.

Hot Region	Chromosome	Start	End	Size (pb)	Gene	Repeats	Number of cases detected in healthy	Number of cases detected in our target genes
1	12	66451361	66451530	170		SINE, Simple	8627	12
3	2	33141301	33141770	470	LINC00486	Simple	2833	648
9	6	382041	382470	430			1371	106
11	4	1708921	1709060	140	SLBP	SINE	1206	4
12	1	62390831	62390970	140	INADL	SINE, Simple	1193	1
14	16	33428141	33428570	430		LINE	1116	106
16	7	107410631	107410760	130	SLC26A3	SINE, Simple	1067	7

(Adapted from Xu et al., 2014)

The second step was the analysis using the Smoove tool. Due to the filters that this program applies to the output generated by Lumpy, as expected, the specificity was improved, increasing up to 100% for all the target alterations. Similarly, no false positives were observed neither when DELLY was used nor in the analysis carried out

by DREAMgenics. As a consequence, sensitivity of all three methods is compromised except for *CCND1*, for which 100% of known cases were identified (Table 54). On the other hand, despite the fact that the specificity achieved by the three methods was 100%, when analyzing the raw data non-specific rearrangements were detected in the three methods, as was the case with Lumpy. However, in contrast to Lumpy, these rearrangements were easily identified as non-specific or non-pathogenic rearrangements, and were removed from the analysis. Filtered results included, on one hand, those with frequent rearrangement partners, which appeared repeated between the different patients and target genes, and some of which have been described by *Xu et al* to be among the top 16 translocation hot regions in human genomes (Xu et al., 2014) (Table 55). On the other hand, some rearrangements which showed low quality values were also filtered. It is noteworthy that since a few real rearrangements previously identified by FISH were detected with low quality, low quality rearrangements could not be filtered directly and visualization of IGV was needed to refine the rearrangements that would be considered. In this regard, when we observed filtered rearrangements in IGV, both were characterized by being highly nonspecific in the target region, which was shown with a large number of reads mapping different rearrangement pairs (Figure 50). Globally, specificity was maintained and similar sensitivity (around 60%) was achieved for all the methods after the filtering process. The highest global sensitivity was obtained when DELLY was used (69.8%) (Table 54).

With regard to PV2, the rearrangements obtained were analyzed and compared with data previously obtained by conventional techniques (n=51 patients, 103 rearrangement calls). Non-specific rearrangements detected in the raw data were filtered in the same way as in PV1. As can be seen in Table 56, on two occasions these non-specific rearrangements coincided with those detected in PV1. On the other hand, 10 non-specific rearrangements were detected that were not repeated between the different callers, target genes or panel versions. However, collecting all non-specific rearrangements in a database could aid in their identification and filtering. Although, as previously detailed, these cases are easily discounted when analyzed in IGV.

Similarly to PV1, a specificity of 100% was reached with all methods for all target genes except for *MYC*, for which all three callers detected a gene rearrangement with *IGH,* in a *MYC*-negative case by FISH. However, manual IGV analysis, in addition to the high quality of results obtained by the three tools, suggests that it could be a "real" translocation, and it should be validated before discarding it. Regarding sensitivity, comparative of results obtained by each caller were very similar to those observed in PV1, although an improvement was observed in those target regions for which the design was optimized, as will be detailed in the following sections. While all three methods managed to detect 100% of the *CCND1* rearrangements, sensitivity ranged from 50 to 85% for the other target genes (DELLY being the one with the highest overall sensitivity, 83.9%). However, as in PV1, DREAMgenics continued to be the method that showed the highest sensitivity for *BCL*2 (85.71% vs 78.57% for DELLY, and 64.29% for Smoove) (Table 54).

On the other hand, in both PV1 and PV2, all methods provided the genomic coordinates of the rearrangement breakpoints. In many cases, different callers detected several breakpoints for the same rearrangement, which differed by a few base pairs. This variability is inherent to the technique, therefore in these cases only one of the identified break points was taken into account. However, in some cases, breakpoints that differed by 5-10Kb were identified for the same rearrangement. Due to this huge difference in the position of the breakpoints detected, in these cases it was considered that they were two distinct breakpoints, whose difference in their genomic position could be a consequence of other structural variations such as deletions or duplications of genomic material.

Altogether, DELLY proved to be the best rearrangement analysis tool in both versions of the panel. However, although Smoove did not detect any case that had not already been detected by DELLY, the joint analysis of these two tools provided greater reliability, both in the detection of the rearrangement and in the identification of the rearrangement partner. On the other hand, although DELLY sensitivity was quite high for *BCL2*, DREAMgenics presented a higher sensitivity for this target gene in both panel versions and was also considered in the final characterization of rearrangements.

Table 54. The results obtained by the three callers used to detect the rearrangements (Smoove, Delly, and DREAMgenics) for PV1 and PV2.

PV1

Target	Routine		Smoove Not filtered Positive	Smoove Filtered Positive	Smoove Filtered Negative	Smoove Sensitivity	Smoove Specificity	DELLY Not filtered Positive	DELLY Filtered Positive	DELLY Filtered Negative	DELLY Sensitivity	DELLY Specificity	DREAMgenics Not filtered Positive	DREAMgenics Filtered Positive	DREAMgenics Sensitivity	DREAMgenics Specificity
BCL2	Positive	20	11	11	9			13	13	7			14/20	14/20		
	Negative	14	0	0	14	55,0%	100,0%	0	0	14	65,0%	100,0%	0/12	0/12	70,0%	100,0%
	Not tested	74	0	0	74			0	0	74			0/14	0/14		
BCL6	Positive	10	9	8	2			10	10	0			9/10	9/10		
	Negative	32	3	0	32	80,0%	100,0%	8	0	32	100,0%	100,0%	0/22	0/22	90,0%	100,0%
	Not tested	66	5	1	65			19	1	65			1/15	1/15		
MYC	Positive	16	8	7	9			14	8	8			7/16	7/16		
	Negative	24	2	0	24	43,8%	100,0%	13	0	24	50,0%	100,0%	0/21	0/21	43,8%	100,0%
	Not tested	68	0	0	68			28	0	68			0/9	0/9		
CCND1	Positive	6	6	6	0			6	6	0			6/6	6/6		
	Negative	2	0	0	2	100,0%	100,0%	2	0	2	100,0%	100,0%	0/1	0/1	100,0%	100,0%
	Not tested	100	0	0	100			68	0	100			7/39	0/39		

PV2

Target	Routine		Smoove Not filtered Positive	Smoove Filtered Positive	Smoove Filtered Negative	Smoove Sensitivity	Smoove Specificity	DELLY Not filtered Positive	DELLY Filtered Positive	DELLY Filtered Negative	DELLY Sensitivity	DELLY Specificity	DREAMgenics Not filtered Positive	DREAMgenics Filtered Positive	DREAMgenics Sensitivity	DREAMgenics Specificity
BCL2	Positive	20	9	9	5			12	11	3			12/14	12/14		
	Negative	14	1	0	19	64,3%	100,0%	11	0	19	75,6%	100,0%	5/19	0/19	85,7%	100,0%
	Not tested	74	0	0	34			19	1	33			2/11	0/11		
BCL6	Positive	10	5	5	2			7	6	1			5/7	5/7		
	Negative	32	0	0	27	71,4%	100,0%	23	0	27	85,7%	100,0%	0/19	0/19	71,4%	100,0%
	Not tested	66	1	1	32			30	3	30			3/18	3/18		
MYC	Positive	16	2	2	2			4	3	1			3/4	2/4		
	Negative	24	1	1	22	50,0%	100,0%	23	1	22	75,0%	95,7%	1/21	1/21	50,0%	95,2%
	Not tested	68	0	0	40			40	0	40			1/19	0/19		
CCND1	Positive	6	6	6	0			6	6	0			6/6	6/6		
	Negative	2	0	0	3	100,0%	100,0%	3	0	3	100,0%	100,0%	0/1	0/1	100,0%	100,0%
	Not tested	100	0	0	58			57	0	58			3/37	0/37		

Routine techniques included FISH and CBA

Table 55. Non-specific rearrangements detected in PV1 by the different methods.

Target gene	Caller	Genomic coordinates of the partner gene	No. of cases with the same rearrangement partner	Defined among top 16 rearrangement hot regions by *Xu et al* (Xu et al., 2014)
BCL6	DELLY	2:33141310 (LINC00486)	19	Yes
		12:66451372-66451373	9	Yes
		1:42351357	1	No
	SMOOVE	11:21929951	8	No
MYC	DELLY	2:33141310 (LINC00486)	14	Yes
		12:66451372-66451373	15	Yes
		3:25802410	26	No
		18:35460556	10	No
		11:85068747-85068748	2	No
		15:55218276-55218277	1	No
		18:19318826-19318827	1	No
		16:79447108-79447109	2	No
	SMOOVE	18:35460556	2	No
		3:25802410	1	No
CCND1	DELLY	2:33141318 (LINC00486)	40	Yes
	*DREAMgenics	14:106330070 (IGH)	7	No

* These rearrangements were detected in non-target pathologies, and always with low allelic freq (< 0.1) and qual 0.

Table 56. Non-specific rearrangements detected in PV2 by the different methods.

Target gene	Caller	Genomic coordenates of the partner gene	No. of cases with the same translocation partner	Defined among top 16 translocation hot regions by *Xu et al* (Xu et al., 2014)
BCL2	DELLY	2:33141674 (LINC00486)	35	Yes
		5:51080866	1	No
		6:9226860	1	No
		10:92051146	1	No
		12:59762521	1	No
		X:118845824	4	No
	DREAMgenics	X:118845824	7	No
		1:154878460 (KDSR)	2	No
	SMOOVE	X:118845824	2	No
BCL6	DELLY	2:33141301	56	Yes
		6:93097206	3	No
		13:82429523	1	No
		1:42351357	1	No
		13:22344755	1	No
MYC	DELLY	12:47478547	1	No
		11:85068747-85068748	16	No
		2:33141301	66	Yes
		2:223731499	1	No
		6:24600340	1	No
		5:88060637	1	No
		X:45587867	2	No
	DREAMgenics	12:47478547	1	No
CCND1	DELLY	2:33141654	66	Yes
		7:105741885	2	Yes
	DREAMgenics*	14:106329461	3	No

* These translocations were detected in pathologies for which they were not targeted, and always with low allelic frequency (< 0.1) and quality parameter at 0
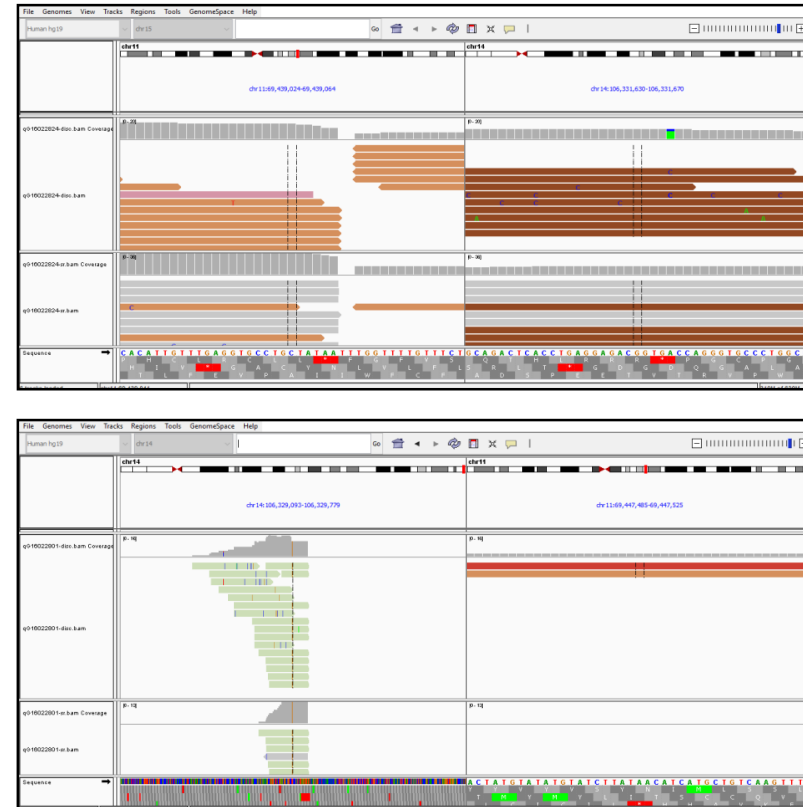
Figure 50. Examples of the visualization of different translocation partners in IGV. (A) Visualization of non-specific translocation partners, each chromosome partner is represented in a different colour; (B) Visualization of a real *CCND1*-IGH rearrangement (picture above) and a false one (picture below) due to the low quality of the reads mapping the rearrangement partner, and the scarcity of reads in the target region.

.

3.2. Analysis and characterization of rearrangement target genes included in the panel

Since no positive case was identified by conventional techniques for the rearrangement targets other than the *CCND1, BCL2, BCL6*, and *MYC* genes, included in PV1, and no additional case was detected by NGS, this section will focus exclusively in analyzing the rearrangements detected in these four target genes.

3.2.1. *CCND1*

With regard to *CCND1*, the design developed in both versions of the panel was optimal since 100% of the previously identified rearrangements were detected (n=6 cases for PV1, and n=6 cases for PV2). As expected, this gene was rearranged in all MCL patients, as the *CCND1* rearrangement is the hallmark of MCL. No further rearrangements were identified in patients diagnosed with other B-cell malignances.

When focusing on the detected breakpoints, although it has been described that most of them (30–50%) map to the ~200bp major translocation cluster (MTC) near *CCND1* (~120 Kb at 3' from the gene), all breakpoints sequenced in our study were located outside the MTC (Fernàndez et al., 2005; Jares & Campo, 2008). Furthermore, consistent with other publications, our analysis revealed a remarkable diversity in the breakpoints, which were found scattered in a large genomic region of 109 kb (from 174 bp to 109 kb 3 'of CCND1) (Greisman et al., 2012) (Figure 51).

Besides, the rearrangement partner identified in all cases was the IGH gene and, in the same way as in *CCND1*, our analysis revealed a great variability in the distribution of breakpoints in this locus (Figure 51). Moreover, as previously described by other authors, ~60% of the cases (7/12) displayed two distinct breakpoints at the IGH locus, which in all cases consisted of a D_H and a J_H coding end breakpoints (Greisman et al., 2012). In addition, in one case, two breakpoints were also observed at the *CCND1* locus, which were located at 80 kb and 70 kb 3' from *CCND1*, respectively. As shown in other studies, these multiple breakpoints could correspond to the sequence from both derivative chromosomes involved in the translocation. The different sequences identified could be a consequence of the loss of material associated with some chromosomal breaks and recombinations that generated the

rearrangement (Greisman et al., 2012). Of the cases with a single breakpoint at IGH, three cases displayed a D_H coding end breakpoint, one case had a breakpoint in the J_H region, and one case had a V_H breakpoint. Furthermore, the latter presented an additional translocation of *CCND1* with chromosome 1p13.2, which was detected by all three softwares (DELLY, Smoove and DREAMgenics) (Table 57).
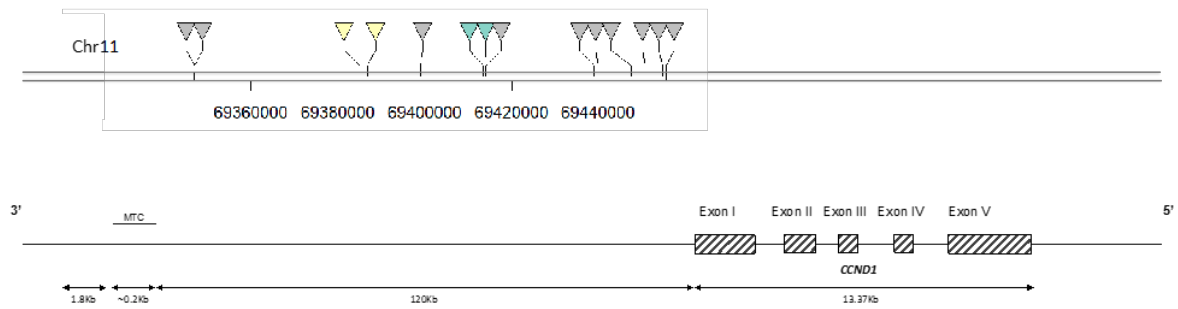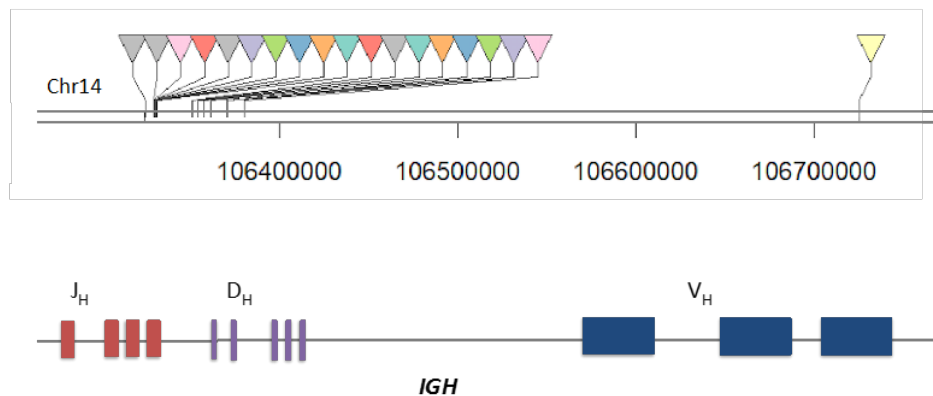
Figure 51. Scheme of the breakpoints identified involving the *CCND1*-IGH rearrangements. (A) *CCND1* gene and (B) its partner gene IGH. In gray the cases with a single breakpoint or rearrangement identified. The breakpoints corresponding to the same case with multiple breakpoints or rearrangements are identified with the same color.

Table 57. Detail of the rearrangements detected involving the *CCND1* gene.

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene
CCND1-1	11:69454808	1.065 kb 3' from *CCND1*	14:106324592	IGHD
CCND1-2	11:69439040	16.832 kb 3' from *CCND1*	14:106357561 14:106331650	IGHD6-19 IGHJ1
CCND1-3	11:69413465	42.408 kb 3' from *CCND1*	14:106725199	IGHV3-23
	11:69413822	42.051 kb 3' from *CCND1*	1:112660208	1p13.2
CCND1-4	11:69413985	41.888 kb 3' from *CCND1*	14:106370544 14:106330070	IGHD3-29 IGHJ5
CCND1-5	11:69447492	8.381 kb 3' from *CCND1*	14:106350737 14:106329467	IGHD4-23 IGHJ6
CCND1-6	11:69455699	0.174 kb 3' from *CCND1*	14:106329480	IGHD
CCND1-7	11:69346891	108.982 kb 3' from *CCND1*	14:106370541 14:106330845	IGHD3-9 IGHJ3
CCND1-8	11:69346872	109.001 kb 3' from *CCND1*	14:106361495 14:106330845	IGHD3-16 IGHJ4
CCND1-9	11:69386636 11:69384951	69.237 kb 3' from *CCND1* 79.922 kb 3' from *CCND1*	14:106370542 14:106330472	IGHD3-9 IGHJ4
CCND1-10	11:69450769	5.105 kb 3' from *CCND1*	14:106380221 14:106329464	IGHD3-3 IGHJ6
CCND1-11	11:69398929	56.944 kb 3' from *CCND1*	14:106329461	IGHJ6
CCND1-12	11:69439851	16.022 kb 3' from *CCND1*	14:106354408	IGHD

Although *CCND1* rearrangements with chromosome 1p13 have not been described in the literature, complex rearrangements involving the *IGH* and *CCND1* loci and other chromosomes have been observed in a subset of primary tumors and MCL cell lines (Salaverria et al., 2008). Moreover, secondary alterations of chromosome 1 are frequently found in t(11;14)(q13;q32)-positive B-cell non-Hodgkin lymphoma (NHL) subtypes, including MCL, for which, in most cell lines the most frequently altered and deletion-prone loci were 1p31 and 1p21. Besides, MCL cell lines with alterations involving the cytoband 1p13 have also been described (Rudolph et al., 2006).

On the other hand, in this patient with two *CCND1* rearrangements found, the FISH results using LSI IGH/CCND1 Dual Color, Dual Fusion Translocation Probe (Abbot Molecular, Des Plaines, Illinois, USA) showed a classic dual fusion rearrangement pattern (2F1O1G). In addition, although a cytogenetics culture was set-up in this case, not enough metaphases were obtained to get an informative result. However, the possibility that it could be a small rearrangement, not observable by FISH due to its resolution limit cannot be discarded, which would explain the classic two-fusion pattern observed. Moreover, when visualizing these *CCND1* rearrangements in IGV, the existence of a rearrangement with 1p31 was confirmed, and its evidence was even more robust than for the *CCND1-IGH* rearrangement (Figure 52). Furthermore, the hypothesis that it could be a region homologous with the IGH gene was discarded using the BLAST tool (NCBI). However, it would be advisable to confirm these results with other techniques.
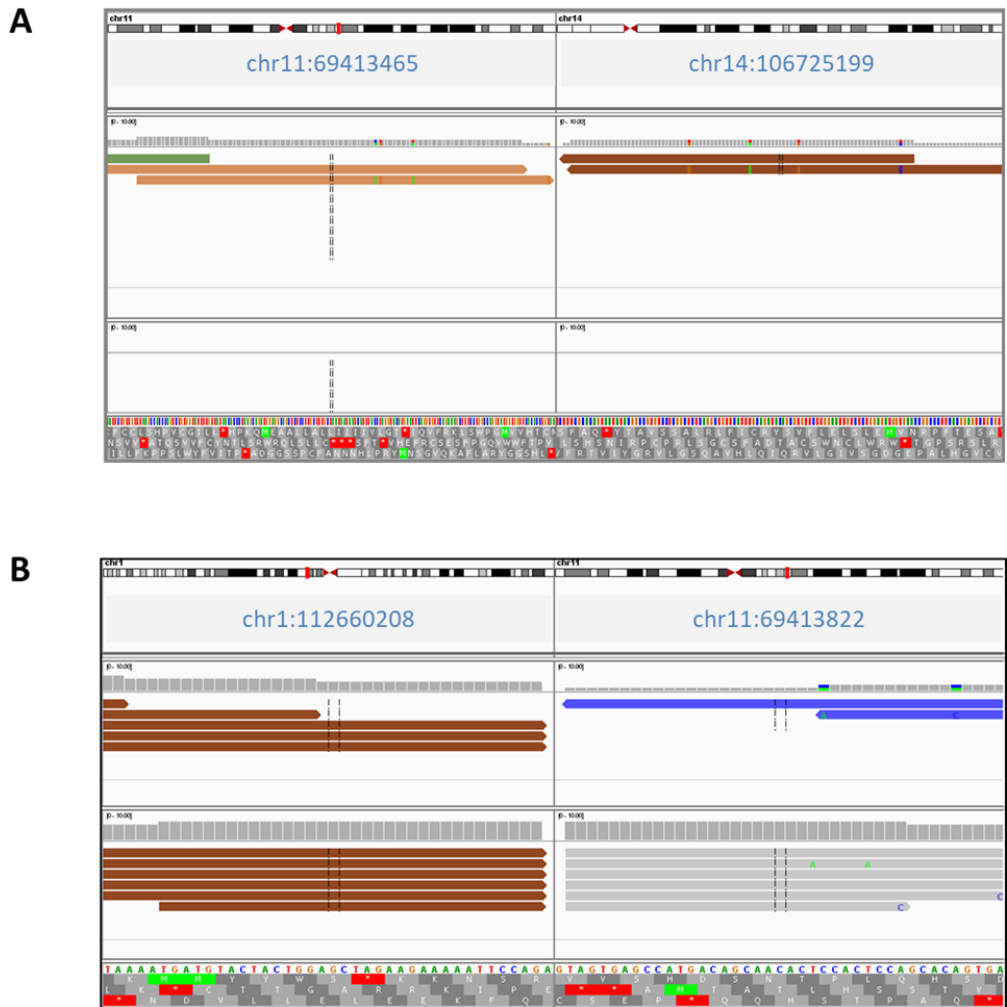
Figure 52. IGV image of the two rearrangements detected in a patient with MCL. (A) Rearrangement detected between *CCND1,* on chromosome 11 (left), and IGH, on chromosome 14 (right). The reads that are identified in light brown covering chromosome 11 have their partner read mapping chromosome 14. The reads identified in dark brown covering chromosome 14 have their partner read mapping chromosome 11. This is an image indicative of a chromosomal rearrangement between both chromosomes; (B) Rearrangement detected between chromosome 1 (left), and *CCND1,* on chromosome 11 (right). The reads that are identified in dark brown covering chromosome 1 have their partner read mapping chromosome 11. The reads that are identified in blue covering chromosome 11 have their partner read mapping chromosome 1. This is an image indicative of a chromosomal rearrangement between both chromosomes.

3.2.2. *BCL2*

Detection of rearrangements in *BCL2* was improved after optimizing the *BCL2* design in PV2 (sensitivity increased from 70% to 86%). In line with previous knowledge, among the 31 *BCL2* rearrangements detected in 27 patients, most breakpoints were located within either the major breakpoint region (MBR) or the minor cluster region (mcr) (64.5% and 9.7%, respectively). However, a large number of breakpoints (25.8%) dispersed between MBR and mcr were also observed (Buchonnet et al., 2000; Godon et al., 2003). Although these breakpoints were correctly covered in the PV1, in four out of the 14 positive cases sequenced in PV2 (2.8%) the breakpoint was located at the end of the mcr, which represents the limit of the design in PV1, and at 3' mcr, which was not covered in PV1 (Figure 53, Table 59). Two of these cases corresponded to *BCL2* rearrangements missed by PV1 that were re-sequenced with PV2. Thus, probably the remaining four cases not detected in PV1, from which no further material was available, would be detected with the optimized design for *BCL2*.

However, two of the 14 cases with known *BCL2* translocation by FISH could not be identified in PV2. Although the PV2 design covers all the regions that have been described involved in *BCL2* rearrangements, the possibility that in these cases the rearrangement occurs in a region not covered by the panel cannot be discarded (Akasaka et al., 1998; Godon et al., 2003; Espinet et al., 2008). On the other hand, although these two cases were detected by FISH, we did not know the percentage of altered cells, so these discordances could also be due to a low tumor infiltration. Furthermore, although the same biopsy on which the FISH study was performed was used for NGS, the possibility that the sample fragment used for NGS displayed less tumor infiltration than the fragment used for FISH cannot be discarded, although the immunohistochemical studies performed on the sequenced samples identified tumor cellularity in all cases.

Apart from the cases previously identified by FISH, an additional *BCL2-IGH* translocation was detected in a patient diagnosed with SMZL in which *BCL2* had not been tested in the routine procedure. Despite not being common, a patient with SMZL carrying this alteration has been previously described (Baseggio et al., 2012).

On the other hand, focusing on the rearrangement partners, all cases presented rearrangements of *BCL2* with the *IGH* gene, and as it was observed for *CCND1-IGH* translocation, our analysis revealed some variability in the distribution of breakpoints at this *locus*. However, despite this variability, a large number of coincident breakpoints were identified among patients, being the JH6 junction segment the most frequent, which is consistent with the literature (Buchonnet et al., 2000) (Figure 53, Table 59). In contrast to what has been described by other authors, the breakpoints were not only located at J_H region, but also along D_H and V_H regions (Godon et al., 2003). Moreover, coinciding with what was observed in *CCND1*, 59% of the cases (16/27) displayed two breakpoints at the IGH locus, which in all cases consisted of a D_H and a J_H coding end breakpoint, except in a case where the second breakpoint was on chromosome 14, but outside the IGH region. Furthermore, in four cases, two breakpoints were also observed at the *BCL2* locus. These multiple breakpoints appear to correspond to the sequence from both derivative chromosomes involved in the translocation, as previously observed in *CCND1*. However, two of these cases showed a large difference in the coordinates of the breakpoints involved in the translocation (>200kb vs 6kb for the remaining) and when visualizing these rearrangements at the IGV, a rearrangement pattern was observed that agrees with the hypothesis of a double rearrangement in the same *BCL2* allele (Figure 54), which have been previously described by other authors (Seité et al., 1993; Vaandrager et al., 2000).

Finally, despite the fact that the *BCL2-IGH* rearrangement is a diagnostic marker in FL, many of the patients with *BCL2-IGH* translocation were diagnosed of DLBCL (14/27; 52%). In all, 78% (11/14) of these DLBCL patients corresponded to cases that came from an initial FL, which later transformed into a DLBCL, or cases with DH or TH lymphoma, which would explain this high incidence of *BCL2* rearrangements in this pathology. The *BCL2-IGH* rearrangement was also identified in two patients diagnosed with BL and SMZL. The remaining cases corresponded to patients diagnosed with FL (11/27; 41%), as expected. Notably, our analysis did not reveal any difference in breakpoint distribution among patients diagnosed with different lymphomas. Indeed, survival studies have not shown any correlation between the

breakpoint location and the clinical outcome in patients with FL (Buchonnet et al, 2002).
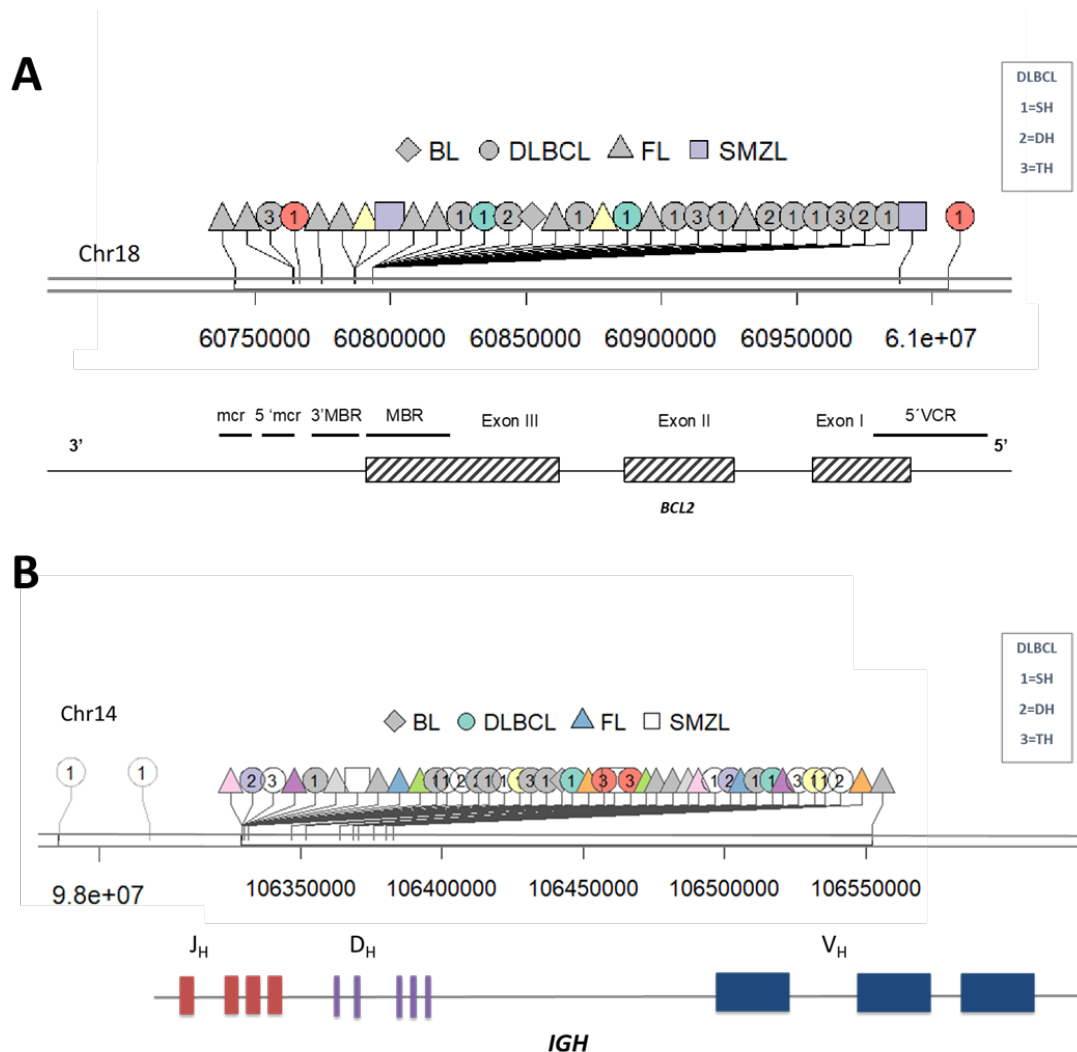
Figure 53. Scheme of the breakpoints identified involving the *BCL2*-IGH rearrangements. (A) *BCL2* gene and (B) its partner gene IGH. In gray the cases with a single breakpoint or rearrangement are identified. The breakpoints corresponding to the same case with multiple breakpoints or rearrangements are identified with the same color. The shape of the breakpoints shows the diagnosis of the patients. Patients diagnosed with DLBCL are identified with 1= Single hit lymphoma (SH); 2= Double hit lymphoma (DH); and 3= Triple hit lymphoma (TH).

Table 58. Detail of the rearrangements detected involving the *BCL2* gene

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the target gene	Patient diagnosis
BCL2-1	18:60793552	Exon 3 of *BCL2* (MBR)	14:106370542 14:106330074	IGHD3-9 IGHJ5	DLBCL
BCL2-2	18:60793458	Exon 3 of *BCL2* (MBR)	14:106330071	IGHD2-2	DLBCL
BCL2-3	18:60793458 18:60793496	Exon 3 of *BCL2* (MBR) 3.302 kb 3' *BCL2* (~3' MBR)	14:106330069 14:106382688	IGHJ5 IGHD2-2	DLBCL
BCL2-4	18:60793550	Exon 3 of *BCL2* (MBR)	14:106370364 14:106329446	IGHD3-10 IGHJ6	DLBCL (DH)
BCL2-5	18:60793458	Exon 3 of *BCL2* (MBR)	14:106329465	IGHJ5	DLBCL (DH)
BCL2-6	18:60793504	Exon 3 of *BCL2* (MBR)	14:106351764 14:106331463	IGHD3-22 IGHJ1	DLBCL (TH)
BCL2-7	18:60793458	Exon 3 of *BCL2* (MBR)	14:106330071	IGHJ5	BL
BCL2-8	18:60793496	Exon 3 of BCL2 (MBR)	14:106370539 14:106329459	IGHD3-9 IGHJ6	FL
BCL2-9	18:60793552	Exon 3 of BCL2 (MBR)	14:106330070	IGHJ5	DLBCL (TH)
BCL2-10	18:60793495 18:60787272	Exon 3 of BCL2 (MBR) 3.307 kb 3' *BCL2* (~3' MBR)	14:106330075 14:106382691	IGHJ5 IGHD2-2	FL
BCL2-11	18:60793447	Exon 3 of BCL2 (MBR)	14:106363817	IGHD2-15	FL
BCL2-12	18:60793461	Exon 3 of BCL2 (MBR)	14:106351901 14:106329462	IGHD3-22 IGHJ6	FL
BCL2-13	18:60786790	3.789 kb 3' *BCL2* (3' MBR)	14:106363819 14:106329174	IGHD2-15 IGHJ6	FL

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene	Patient diagnosis
BCL2-14	18:60774504	16.075 kb 3' *BCL2* (5' mcr)	14:106363818 14:106329451	IGHD2-15 IGHJ6	FL
BCL2-15	18:60742638	47.941 kb 3' *BCL2* (3' mcr)	14:106375769 14:106329450	IGHD1-7 IGHJ6	FL
BCL2-16	18:60787278 18:60987949	3.301 kb 3' *BCL2* (3' MBR) 1.336 kb 5' *BCL2* (5´VCR)	14:106346897 14:106329451	IGHD1-26 IGHJ6	SMZL
BCL2-17	18:60764459	26.120 kb 3' *BCL2* (mcr)	14:106380225 14:106329449	IGHD3-3 IGHJ6	DLBCL (TH)
BCL2-18	18:60793550	Exon 3 of *BCL2* (MBR)	14:106329462	IGHJ6	DLBCL
BCL2-19	18:60793600	Exon 3 of *BCL2* (MBR)	14:106329465	IGHJ6	DLBCL
BCL2-20	18:60793496	Exon 3 of *BCL2* (MBR)	14:106370539	IGHD3-9	DLBCL
BCL2-21	18:60793549	Exon 3 of *BCL2* (MBR)	14:106552288	IGHD3-9	FL
BCL2-22	18:60763910	26.673 kb 3' *BCL2* (mcr)	14:106329458	IGHJ6	FL
BCL2-23	18:60793548	Exon 3 of *BCL2* (MBR)	14:106382688 14:106329462	IGHD2-2 IGHJ6	DLBCL
BCL2-24	18:60793462	Exon 3 of *BCL2* (MBR)	14:106368509 14:106330064	IGHD5-12 IGHJ5	DLBCL
BCL2-25	18:60793447	Exon 3 of *BCL2* (MBR)	14:106363817	IGHD2-15	FL
BCL2-26	18:60766640 18:61005948	23.939 kb 3' *BCL2* (mcr) 19.335 kb 5' *BCL2* (~5' VCR)	14:106329450 14:97496686 14:98643070	IGHJ6 14q32.2 14q32.2	DLBCL
BCL2-27	18:60793552	Exon 3 of *BCL2* (MBR)	14:106382688 14:106329462	IGHD2-2 IGHJ6	DLBCL (DH)

A Chr18:60987949
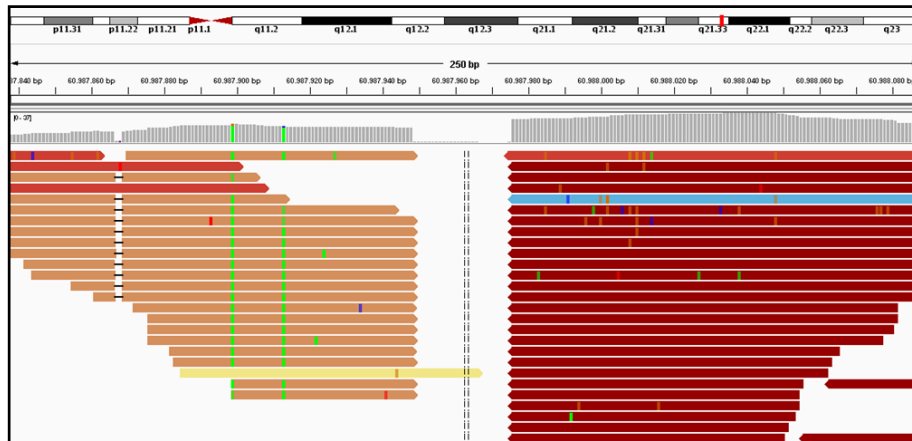
B Chr18:60787278
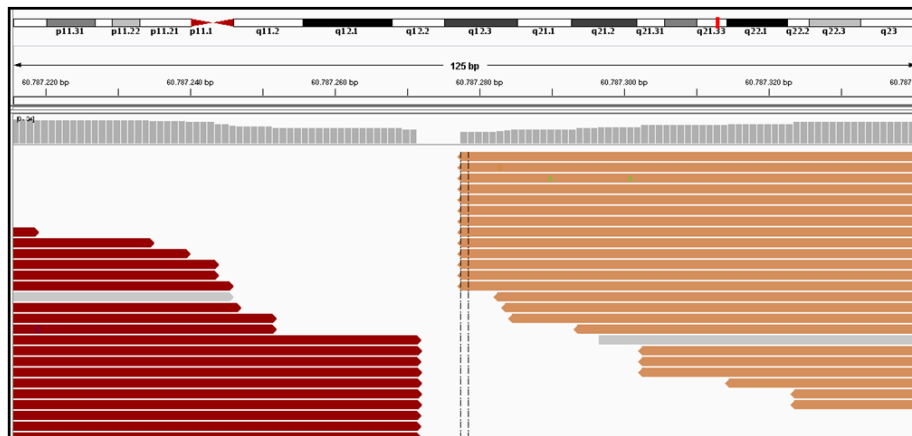
Figure 54. IGV image of the two break points identified in a patient with double *BCL2*-IGH rearrangement. (A) Genomic coordinates from one of the breakpoints on chromosome 18; (B) Genomic coordinates from the second breakpoint on chromosome 18. In both figures the reads identified in light brown covering chromosome 18 have their paired-read mapping chromosome 14. This image is indicative of a double chromosome rearrangement between *BCL2* and IGH.

3.2.3. *BCL6*

With regard to *BCL6*, the design developed for both versions of the panel turned out to be optimal since approximately 94% of the previously identified rearrangements were detected (n= 10/10 cases for PV1, and n= 6/7 cases for PV2). Furthermore, the same *BCL6* rearrangement was successfully detected in a patient sequenced in both panels. In contrast, one known *BCL6* translocation previously described by FISH could not be detected in a patient diagnosed with MALT lymphoma. Although the panel design covers a much larger region than the two rearrangement clusters described in the literature (Iqbal et al. 2007; Jarosova et al. 2016), the possibility that the rearrangement occurs in a region not covered by the panel cannot be discarded. However, the results obtained from the FISH performed on the same biopsy (but in FFPE tissue, unlike the sequenced sample that was OCT embedded fresh tissue sample) showed a very low percentage of altered cells (10-15%), with a classic rearrangement pattern (1F1O1G).These data support our hypothesis that this case could not be detected mainly due to a low tumor cell infiltration in the sequenced sample.

Rearrangements affecting the *BCL6* gene occur predominantly in the major breakpoint region (MBR) encompassing the first noncoding exon and a portion of the first intron of *BCL6*. In addition, an alternative breakpoint cluster (ABR) located 245–285 kb 5' *BCL6* has also been described (Iqbal et al., 2007; Jarosova et al., 2016). Although ABR breakpoints have been associated with a follicular growth pattern, few cases diagnosed with DLBCL with breakpoints localized in this cluster have been identified (Iqbal et al., 2007). In line with these data, our analysis revealed that 57% of the breakpoints detected were located in the MBR, of which, 75% corresponded to IG-*BCL6* rearrangements. In addition, three breakpoints located near ABR cluster were described in two patients with DLBCL. On the other hand, our analysis also revealed that around 30% of the breakpoints were not within any of the previously described clusters, but were distributed up to 30kb 5' of *BCL6*. Specific breakpoint sequence motifs have been described in *BCL6* rearrangements, and they have been reported to be differentially used in IG–*BCL6* and non-IG–*BCL6* rearrangements (Lu et al., 2013). In our study, the analysis of the nucleotide sequences involved in the different breakpoints has not been carried out. However, a different distribution pattern of the breakpoints of non-IG-*BCL6*

rearrangements has been observed compared to IG-*BCL6*, with a greater tendency for the latter to accumulate in the MBR region (nine IG-*BCL6* in MBR vs two outside MBR; three non-IG-*BCL6* in MBR vs five outside MBR).

In addition, two cases were identified with a double breakpoint in *BCL6*, as well as in the partner gene (Table 59; Figure 55). Due to the proximity (~6Kb) between the two breakpoints identified in each of the two genes involved in the rearrangement, and the translocation pattern observed in IGV, these multiple breakpoints appear to correspond to the sequence from both derivative chromosomes involved in the rearrangement, as it was observed in *CCND1* and *BCL2.* However, unlike these, *BCL6* seems to display more stable rearrangements, since there is a lower incidence of cases with different breakpoints in their derivative chromosomes as a consequence of the loss of material associated with the rearrangement.
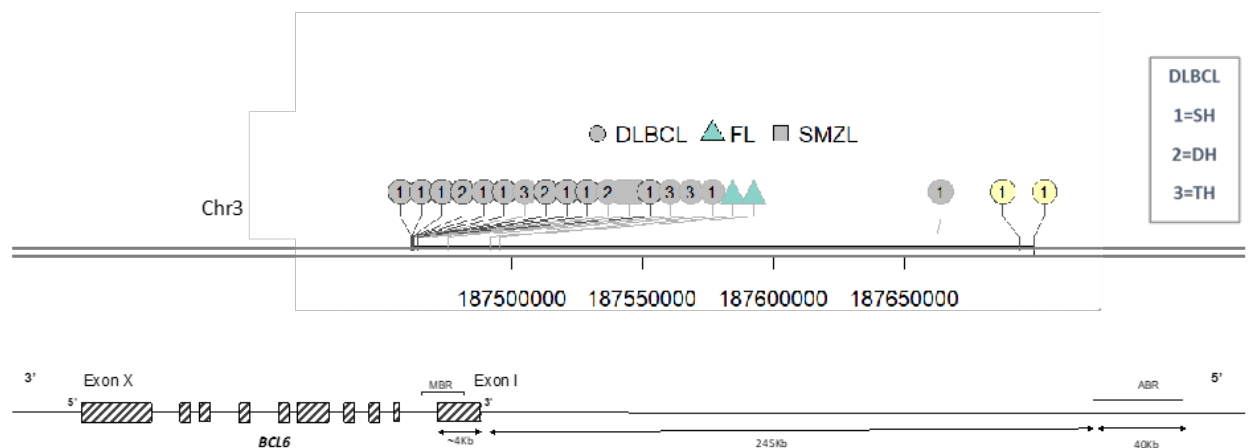
Figure 55. Scheme of the breakpoints identified involving the *BCL6* gene. In gray the cases with a single breakpoint or rearrangement are identified. The breakpoints corresponding to the same case with multiple breakpoints or rearrangements are identified with the same color. The shape of the breakpoints shows the diagnosis of the patients. Patients diagnosed with DLBCL are identified with 1= Single hit lymphoma (SH); 2= Double hit lymphoma (DH); and 3= Triple hit lymphoma

Regarding translocation partners, *BCL6* has been commonly described rearranged with diverse partners, including one of three immunoglobulin genes (IGH, IGK, IGL) loci or diverse non-IG chromosomal loci (Ohno, 2006; Nedomova et al., 2013; Jarosova et al., 2016). Consistent with previous publications, 58% of the rearrangements in our study involved IG genes (n=8 with IGH, and n=3 with IGL), while the remaining 42% (n=8) comprised non-IG regions.

Regarding non-IG partners, to date, more than 20 genomic regions have been described, involving almost all chromosomes. In our study, a great variability has been observed in these partners, since none of the detected regions has been identified repeatedly in two or more patients. In addition, a translocation involving the 3q25 cytoband has been observed in two patients (3q25.1 in the first, and 3q25.33 in the second patient), suggesting an inversion. This translocation has been previously reported in the literature (Jarasova et al., 2016). However, as far as we know, many of the partner genes identified in our study have not been previously reported by other authors (Jarasova et al., 2016) (Table 59; Figure 56).

On the other hand, clinical and biological differences between cases with IG-*BCL6* and non-IG-*BCL6* rearrangements have been observed in DLBCL patients (Akasaka et al., 2000; Ueda et al., 2002). Although there is no consensus on the effect of *BCL6* translocation on the clinical outcome, studies on IG-*BCL6* rearrangements commonly point that high *BCL6* expression at mRNA and/or protein levels is a favourable prognostic indicator of DLBCL through the repressory control of antiapoptotic *BCL2* and other known *BCL6* pro-survival target genes (Lossos et al., 2001; Ueda et al., 2002; Ohno, 2006). On the contrary, in cases with *BCL6* rearrangements with non-IG loci, *BCL6* mRNA levels were unexpectedly low (Ueda et al., 2002; Lossos et al., 2004; Ohno, 2006). Moreover, the survival rate of DLBCL patients with non-IG partners has been shown to be inferior to that of those with IG-*BCL6* translocation, suggesting that non-IG-*BCL6* fusion is a poor prognostic indicator of DLBCL (Akasaka et al., 2000; Barrans et al., 2002). In this line, we found a higher frequency of rearrangements with non-IG genes in cases with double-hit and triple-hit lymphomas compared with other DLBCL (4/6 vs 2/11, P= 0.1094).

Due to the worse prognosis observed in DLBCL patients with non-IG-*BCL6* rearrangements, compared to cases with IG-*BCL6* rearrangements, a worse prognosis could also be expected in DH/TH lymphoma patients with non- IG-*BCL6* rearrangements. However, to our knowledge, no study has been published that analyze the *BCL6* rearrangement partners and their clinical impact on this subset of aggressive lymphomas.

Nevertheless, despite the clinical importance of the identification of the translocation partner, the strategy used in clinical practice does not allow obtaining this information in the vast majority of cases. In fact, in our assay, only two of the 18 cases with *BCL6* rearrangements previously detected by cytogenetic techniques had the translocation partner identified. In both cases the rearrangement partner was detected by karyotype (6p21 in one case and 14q32 in the other), and in both cases it coincided with the one identified by NGS (Figure 56).
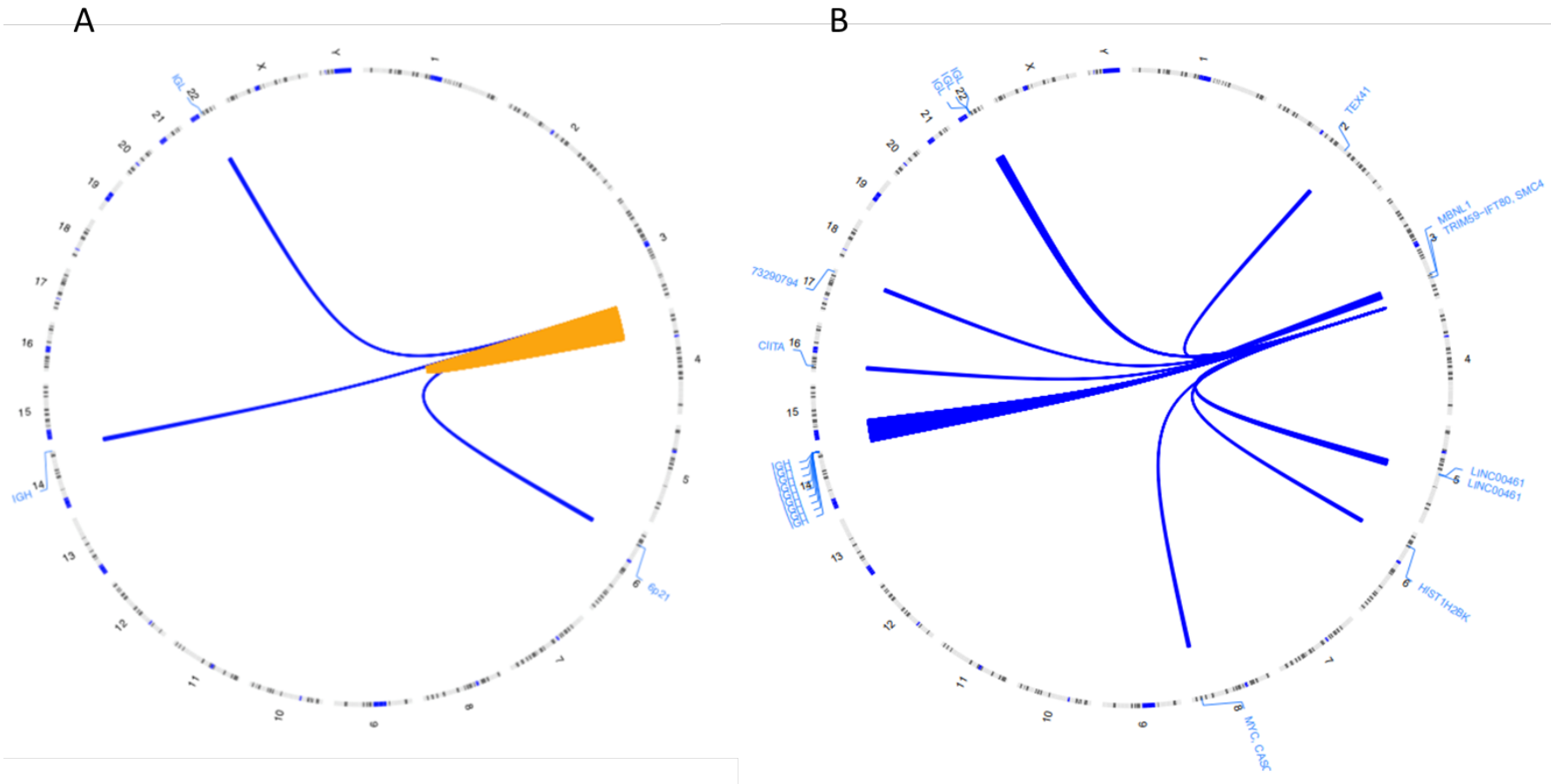
Figure 56. Circosplot showing the rearrangements identified involving the *BCL6* gene by conventional techniques and by NGS. (A) Rearrangements detected by conventional techniques, CBA and/or FISH and; (B) with the designed panel. Chromosomes are represented on the outermost circumference, blue lines indicate the rearrangements between two chromosomes. Cases for which the rearrangement pair is unknown are shown in brown.

Table 59. Detail of the rearrangements detected involving the *BCL6* gene.

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene	Partner gene function	Patient diagnosis	Reported in the literature
BCL6-1	3:187492172 3:187495695	28.659 kb 5' *BCL6* 32.182 kb 5' *BCL6*	5:89358574 5:87948324	5q14.3 5q14.3 (LINC00461)	Post-transcriptional regulation Enhanced proliferation and inhibited apoptosis	FL	No
BCL6-2	3:187463987	0.474 kb 5´*BCL6*	22:23223324	22q11.22 (*IGLV3-1*)	Immunoglobulin lambda locus	DLBCL	Yes
BCL6-3	3:187462672	Intron 1 *BCL6* (MBR)	22:22804907	22q11.22 (*IGLV7-35*)	Immunoglobulin lambda locus	DLBCL (DH)	Yes
BCL6-4	3:187462334	Intron 1 *BCL6* (MBR)	22:23223256	22q11.22 (*IGLV3-1*)	Immunoglobulin lambda locus	DLBCL	Yes
BCL6-5	3:187462992	Intron 1 *BCL6* (MBR)	6:27107400	6p22.1 (*HIST1H2BK*)	Core component of nucleosome	DLBCL (DH)	No
BCL6-6	3:187475788	12.275 kb 5' *BCL6*	2:145908001	2q22.3 (*TEX41*)	Long Intergenic Non-Protein Coding RNA 953	DLBCL	No
BCL6-7	3:187466248	2.735 kb 5' *BCL6*	8:128792690	8q24.21 (*MYC*)	Controls cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis	DLBCL (TH)	Yes
BCL6-8	3:187462520	Intron 1 *BCL6* (MBR)	16:10971638	16p13.13 (*CIITA*)	Transcription of genes of the MHC I and II	DLBCL (TH)	Yes
BCL6-9	3:187469593	6.080 kb 5' *BCL6*	3:151999952	3q25.1 (*MBNL1*)	Regulates alternative splicing	DLBCL (TH)	Yes

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene	Partner gene function	Patient diagnosis	Reported in the literature
BCL6-10	3:187461954	Intron 1 *BCL6* (MBR)	14:106240461	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL (DH)	Yes
BCL6-11	3:187462994	Intron 1 *BCL6* (MBR)	3:160118247	3q25.33 (*TRIM59-IFT80; SMC4*)	*Lnc-RNA*; Chromosome condensation and sister chromatid cohesion	SMZL	No
BCL6-12	3:187462864	Intron 1 *BCL6* (MBR)	14:106325488	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-13	3:187461547	Intron 1 *BCL6* (MBR)	14:106324788	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-14	3:187462903	Intron 1 *BCL6* (MBR)	14:106325843	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-15	3:187461906	Intron 1 *BCL6* (MBR)	14:106327132	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-16	3:187699781	236.268 kb 5' *BCL6* (~ABR)	14:106213838	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
	3:187693923	230.411 kb 5' *BCL6* (~ABR)	14:106330372	14q32.33 (I*GH*)	Immunoglobulin heavy locus		
BCL6-17	3:187462295	Intron 1 *BCL6* (MBR)	14:106239950	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-18	3:187461543	Intron 1 *BCL6* (MBR)	14:106326517	14q32.33 (I*GH*)	Immunoglobulin heavy locus	DLBCL	Yes
BCL6-19	3:187662482	198.988 kb 5' *BCL6* (~ABR)	17:73290797	17q25.1		DLBCL	No

3.2.4. *MYC*

Detection of rearrangements involving *MYC* at 8q24.1 in MBCN is important for diagnostic and prognostic purposes. However, its detection is often hampered by the wide variation in breakpoint location within the *MYC* region, which can affect up to 1 Mb from *MYC* on each side (Einerson et al., 2006; Bertrand et al., 2007). Due to a bioinformatics error, the PV1 only covered the *MYC* gene region in addition to 100 kb on each side of the gene. However, and despite this design limitation, 50% (8/16) of the sequenced cases with previously known *MYC* rearrangements were detected. In order to increase the sensitivity for this target, larger captured regions for *MYC* were designed in PV2. Due to the large size of the genomic region involved in *MYC* rearrangement, encompassing more than 2 Mb, only those genomic regions most frequently involved in the translocation were included to minimize the size of PV2 design (Haralambieva et al., 2004; Einerson et al., 2006). Following this strategy, 75% (n=3/4 cases sequenced) of known rearranged cases were detected, including one case missed in PV1 which was re-sequenced. Despite the improvement in sensitivity observed in PV2, the number of cases included to analyze the ability of the panel to detect *MYC* rearrangements was quite small. For this reason, although our results are promising, additional cases should be analyzed to confirm if the design developed in PV2 is optimal for detecting *MYC* rearrangements. Besides, a *MYC*-IGH translocation was identified in PV2 that had not been previously detected with the IGH/MYC dual-color dual-fusion probe (Abbot Molecular). Performing either *IGH/MYC* dual-fusion or *MYC* break-apart probes studies in isolation can provide false-negative results because of the complex nature of the *MYC* rearrangement (Peterson et al., 2019). Unfortunately, no additional tumor sample was available to perform FISH with the *MYC* break-apart probe. Besides, it could also be a rearrangement involving a small genomic region, not detectable by FISH. Since all the callers identified this rearrangement, and also with high quality, and the visualization of the sequences in IGV support the idea that it is a "real" rearrangement.

Regarding the translocation partners, although IG genes have been reported to be the most frequent, non-IG genes have also been described in 35% to 53% of *MYC*

rearranged DLBCL cases (Karube & Campo, 2015). According to these data, in our study 83% (n=5/6) of the *MYC* rearrangements detected in DLBCL involved the IGH gene, and exceptionally, only in one case (n=1/6, 17%) with a TH lymphoma *MYC* was identified rearranged with the *BCL6* gene. In the context of double/triple-hit, it is of special interest to know whether the *BCL6-MYC* rearrangement leads to the simultaneous activation of *BCL6* and *MYC* genes. In a previous study carried out by *Ryan et al*, using the combination of chromatin immunoprecipitation and next-generation sequencing to map acetylated enhancer elements, they observed that while the *BCL6-MYC* translocation led to strong *MYC* activation by the interaction between the *MYC* promoter and the *BCL6* enhancer elements, *BCL6* was not activated (Ryan et al., 2015). Therefore, the authors suggested that *MYC-BCL6* did not represent a double-hit, but was equivalent to a single-hit *MYC* activation rearrangement, and proposed the term "pseudo double-hit" for this particular translocation. According to this evidence, it is possible that our patient is a pseudo triple-hit. However, further studies will need to be performed to evaluate the activation of *BCL6* and *MYC* genes due to this rearrangement.

Regarding the breakpoints, a difference was observed in their distribution in relation to the rearrangement partner. While the breakpoint in the only non-IG-*MYC* case was identified outside the gene (~39kb 3' from *MYC*), all the breakpoints corresponding to the IG-*MYC* cases were located within the gene. These data agree with that described by other authors and display a similar pattern to what has been reported in Burkitt lymphoma, suggesting a common mechanism and developmental stage at which these rearrangements arise (Chong et al., 2018; Nguyen et al., 2017). Furthermore, it is noteworthy that while in all cases with DH/TH lymphomas the breakpoint was located in the first intron of *MYC*, in the only case diagnosed with DLBCL (not DH/TH) the breakpoint was identified in the first exon of the gene. However, *Chong et al* analyzed 1048 biopsies from patients diagnosed of de novo DLBCL and tFL with DLBCL morphology and observed that breakpoints of the MYC-SH (single hit) or MYC-DH/TH, when rearranged with IG genes, were indistinctly distributed in the "genic cluster" (a genomic region spanning from 1.5 kb upstream of the transcription start site to the end of MYC intron 1) (Chong et al., 2018).

Regarding the prognostic value of *MYC* rearrangements in DLBCL, although it has been confirmed that they have a strong negative prognostic impact on the survival of patients with DLBCL, it has been seen that this impact is evident mainly in the first two years after diagnosis, both in MYC-SH (single hit) and MYC-DH/TH. Furthermore, although a negative impact is observed on MYC-SH, only in the case of MYC-DH/TH the decrease in PFS and OS is significant. Besides, it has recently been pointed out that the type of *MYC* rearrangement partner (IG vs non-IG) has a prognostic impact in DLBCL. A study carried out by *Rosenwald et al* demostrates that DLBCL patients with non-IG-*MYC* rearrangement do not differ in outcomes with wt MYC DLBCL patients (Rosenwald et al., 2019). And in the same way is observed in DLBCL patients with DH/TH. These data show that the exclusive knowledge of the *MYC* status (rearranged or not rearranged) is not sufficient to establish the prognosis in these patients. Therefore, strategies that allow the identification of the rearrangement partner (IG vs non-IG) should be implemented in the clinical practice.

In BL, the *MYC*-IGH translocation is the most common genetic lesion, observed in approximately 80% of cases. In addition, variant *MYC* rearrangements with IGK or IGL genes have been identified in approximately 10% of cases (Nguyen et al., 2017). In our study, *MYC* rearrangements involved the *IGH* gene in all BL patients (n=3/3). Regarding the breakpoints, according to what have been described in the literature, we identified two cases with class I breakpoints (located in the first intron and first exon of *MYC*) and one case with a class III breakpoint (located ~64Kb 5' from *MYC*). In addition, a patient has been identified with a complex *MYC*-IGH rearrangement, located at the same breakpoint on chromosome 8q24, which also involves the *ITGAL* gene, locus at 16p11.2. *ITGAL* has been described to be involved in a variety of immune phenomena including leukocyte-endothelial cell interaction, cytotoxic T-cell mediated killing, and antibody dependent killing by granulocytes and monocytes. In addition, it also contributes to natural killer cell cytotoxicity (Barber et al., 2004). Complex rearrangements involving *MYC*-IGH genes in BL patients have been previously described by other authors (Zimonjic et al., 2001; Liu et al., 2007). However, in our case this pattern was not previously identified by routine techniques, either by karyotype (in which a simple rearrangement with chromosome

14q32 was detected) or by FISH, for which IGH/MYC dual-color dual-fusion probe (Abbot Molecular) was used, and which coincided with what was observed by karyotype (classic dual fusion rearrangement pattern with 2F1O1G). Since all the callers detected this rearrangement with high quality parameters, and after confirming it with IGV, these data could indicate the possibility of a complex rearrangement involving a small region of chromosome 16, below the resolution limit of FISH, which would not allow to see a complex pattern of FISH. *MYC* rearrangement is the genetic hallmark of BL, and although it is not exclusive to this pathology, it represents a diagnostic marker in this disease. Despite the fact that it does not provide prognostic information in this entity, unlike DLBCL, the possible clinicopathological implication of complex rearrangements in BL has not yet been established.

In this line, a patient diagnosed with MCL harboring a complex rearrangement of the *MYC* gene has been identified. MCL with *MYC* abnormalities, referred by some as "double-hit" MCL, are rare. However, other authors have previously reported MCL patients carrying *MYC* rearrangement (Setoodeh et al., 2013; Hu et al., 2016). In "double-hit" MCL with *MYC* rearrangement, about half cases show translocation with the IGH or IGL gene loci, and the other half show non-IG translocation with *MYC*. Furthermore, our case showed a rearrangement of the *MYC* gene with two non-IG partners, 4q13.2 and 11q14.1, with a common breakpoint outside the gene region (at ~73Kb 5' from *MYC*). As it was observed in the previous case, this complex rearrangement was not identified by FISH, for which *MYC* break-apart probe was used and which resulted in a classic rearrangement pattern (1F1O1G). However, both rearrangements were detected by the three callers, and the rearrangement was confirmed by IGV. These data indicate that it could be a rearrangement that is involving a smaller genomic region in one of its partner genes, and that for this reason a complex rearrangement pattern is not observed by FISH. "Double-hit" MCL represent a relatively unique group of MCL with highly aggressive clinical behavior, and it has been associated with advanced-stage disease, a high proliferation rate, p53 expression, aggressive morphologic features, complex karyotypes, and a poor prognosis (Setoodeh et al., 2013; Hu et al., 2016).

Furthermore, the complex rearrangement of *MYC*, involving two or more rearrangement pairs, could be expected to provide an additional aggressive character to this entity. However, as far as we know, no studies have been carried out that have analyzed this event in MCL. In addition, *MYC* rearrangement was detected in a patient diagnosed with FL. This rearrangement was located outside the gene region (~13Kb 3' from *MYC*), and involved the *ZBTB5* gene, locus at 9p13.2. *ZBTB5* has been described as a potent transcription repressor of cell cycle arrest gene p21 and as a potential proto-oncogene that stimulates cell proliferation (Koh et al., 2009). *MYC* rearrangement is a rare event in low-grade B-cell lymphomas and, while it is typically associated with an inferior prognosis and poor response to therapy in DLBCL, its clinical significance in FL is less clear and should be interpreted in the context of overall cytogenetic changes (Nguyen et al., 2017).

Finally, no differences have been observed in the breakpoints of the derivative chromosomes, which indicate that the *MYC* rearrangements occur fairly stable and do not lead to the loss of material associated with the rearrangements. All the rearrangements identified and the breakpoints described are shown in Figures 57 and 58, and Table 60.
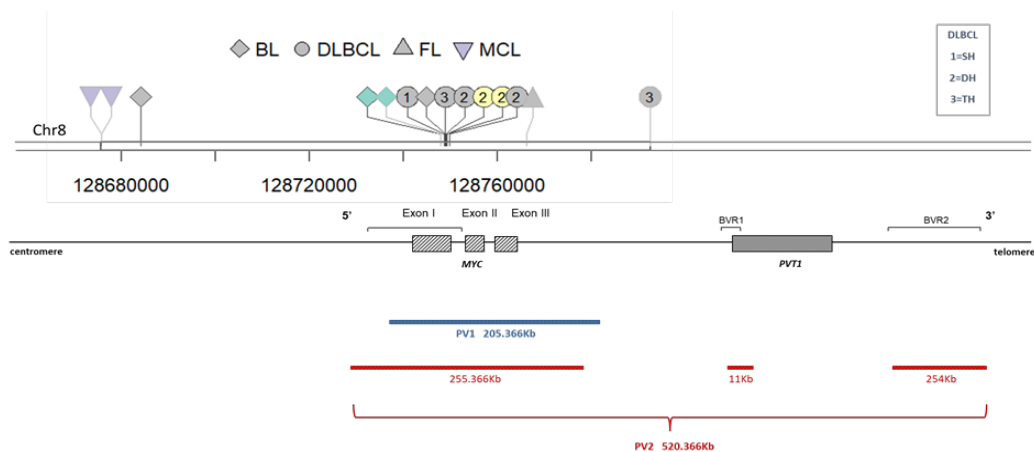
Figure 57. Scheme of the breakpoints identified involving the *MYC* gene. In gray the cases with a single breakpoint or rearrangement are identified. The breakpoints corresponding to the same case with multiple breakpoints or rearrangements are identified with the same color. The shape of the breakpoints shows the diagnosis of the patients. Patients diagnosed with DLBCL are identified with 1= Single hit lymphoma (SH); 2= Double hit lymphoma (DH); and 3= Triple hit lymphoma (TH).
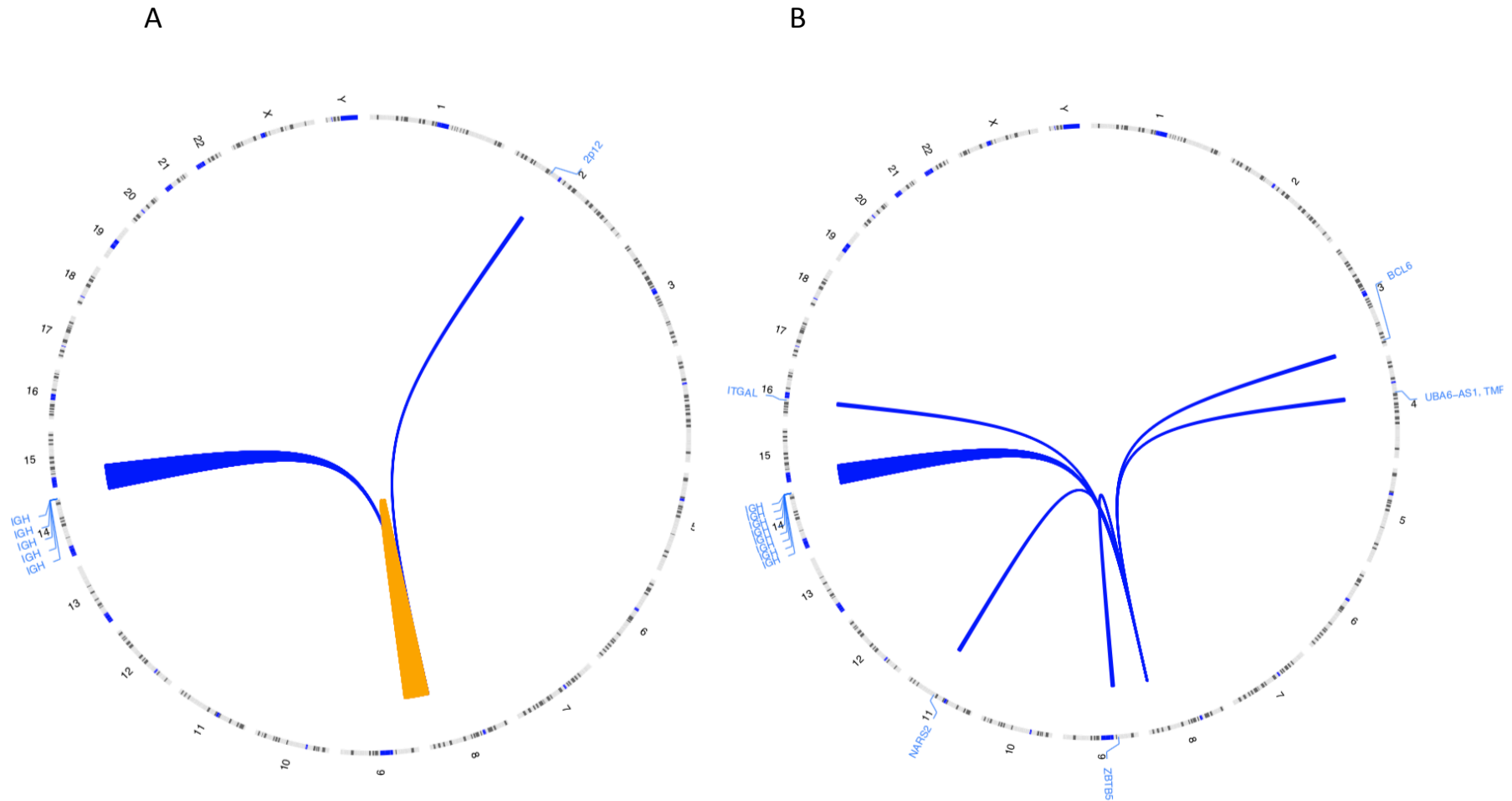
Figure 58. Circosplot showing the rearrangements identified involving the *MYC* gene by conventional techniques and by NGS. (A) Rearrangements detected by conventional techniques, CBA and/or FISH and; (B) with the designed panel. Chromosomes are represented on the outermost circumference, blue lines indicate the rearrangements between two chromosomes. Cases for which the rearrangement pair is unknown are shown in brown.

Table 60. Detail of the rearrangements detected involving the *MYC* gene.

Patient ID	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene	Partner gene function	Patient diagnosis
MYC-1	8:128749265	*MYC* intron 1	14:106176343 14:106056032	14q32.33 (*IGH*) 14q32.33 (*IGH*)	Immunoglobulin heavy locus Immunoglobulin heavy locus	DLBCL (DH)
MYC-2	8:128684200	64.114 kb 5' *MYC*	14:106326617	14q32.33 (*IGH*)	Immunoglobulin heavy locus	BL
MYC-3	8:128748069 8:128748073	*MYC* exon 1 *MYC* exon 1	14:106241010 16:30501529	14q32.33 (*IGH*) 16p11.2 (*ITGAL*)	Immunoglobulin heavy locus Involved in a variety of immune phenomena: leukocyte-endothelial cell interaction, natural killer cell cytotoxicity, etc.	BL
MYC-4	8:128748926	*MYC* intron 1	14:106330384	14q32.33 (*IGH*)	Immunoglobulin heavy locus	BL
MYC-5	8:128792625	38.945 kb 3' *MYC*	3:187466290	3q27.3 (*BCL6*)	Transcriptional repressor	DLBCL (TH)
MYC-6	8:128749047	*MYC* intron 1	14:106069549	14q32.33 (*IGH*)	Immunoglobulin heavy locus	DLBCL (TH)
MYC-7	8:128749455	*MYC* intron 1	14:106209850 14:106238270	14q32.33 (IGH) 14q32.33 (IGH)	Immunoglobulin heavy locus Immunoglobulin heavy locus	DLBCL (DH)
MYC-8	8:128766371	12.692 kb 3' *MYC*	9:37443790	9p13.2 (*ZBTB5*)	Potential proto-oncogene stimulating cell proliferation	FL
MYC-9	8:128748722	*MYC* exon 1	14:106095160	14q32.33 (*IGH*)	Immunoglobulin heavy locus	DLBCL
MYC-10	8:128675658 8:128675667	72.658 kb 5' *MYC* 72.648 kb 5' *MYC*	4:68730281 11:78260168	4q13.2 (*UBA6-AS1; TMPRSS11D*) 11q14.1 (*NARS2*)	ncRNA; role in the host defense system on the mucous membrane Play a critical role in protein biosynthesis	MCL
MYC-11	8:128750025	*MYC* intron 1	14:106056549	14q32.33 (*IGH*)	Immunoglobulin heavy locus	DLBCL (DH)

4. Analysis of copy number alterations

4.1. Global detection of target CNA

To perform the CNA analysis, data were analyzed using the CopywriteR bioinformatics tool. In line with the other types of abnormalities assessed, the parameters used to filter the CopywriteR output were adjusted comparing with data previously obtained by conventional techniques.

Regarding PV1, the maximum concordance with available known CNA results was obtained fixing a bin of 100Kb and segment.mean values of <1.8 and >2.2 to consider lost and gained regions, respectively. In addition, those regions with the same copy number state separated by less than 100 Kb were joined to avoid segmentation of large CNA. On the other hand, since CopywriteR uses off-target reads instead of on-target reads to infer DNA copy number profiles, an additional analysis was performed in which reads from SNP regions, which were included in PV1 design distributed throughout the genome to have an overview of it, were bioinformatically removed from the raw data to find out if the SNP could be excluded from the design and therefore a smaller design could be made for PV2.

Overall, PV1 displayed a high sensitivity (70.1% to 100%) and a specificity of almost 100% for all known target alterations (97.6% to 100%). Both strategies showed similar sensitivity and specificity, being the latter slightly better in the analysis with no SNP (Table 61). Therefore, it was decided to only consider results generated from data without SNP and to eliminate the SNP from the PV2 design, which also made possible to reduce the size of the panel considerably.

Table 61. Sensitivity and specificity obtained for each target alteration with PV1 analyses.

PV1			NGS			
	Routine techniques		Without SNPs		With SNPs	
Target alteration	**N. of cases altered**	**N. of cases not altered**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
Loss 1p36	18	65	94.4% (17/18)	98.7% (64/65)	94.4% (17/18)	100% (65/65)
Gain 2p24-2p12	11	71	100% (11/11)	97.6% (69/71)	90.9% (10/11)	100% (71/71)
Gain chr 3/3q	19	66	100% (19/19)	100% (66/66)	84.2% (16/19)	100% (66/66)
Gain chr 5/5q	6	77	83.3% (5/6)	100% (77/77)	83.3% (5/6)	100% (77/77)
Loss 6q13-q26	25	60	92% (23/25)	98.6% (59/60)	88% (22/25)	100% (60/60)
Loss 7q22.1-q35	8	77	87.5% (7/8)	100% (77/77)	75% (6/8)	100% (77/77)
Loss 8p21.1-p23.1	10	73	100% (10/10)	100% (73/73)	90% (9/10)	100% (73/73)
Gain 8q24 (MYC)	14	69	100% (14/14)	100% (69/69)	85.7% (12/14)	100% (69/69)
Loss 9p21	13	71	84.6% (11/13)	98.8% (70/71)	69.2% (9/13)	100% (71/71)
Loss 11q22.1-q24	16	72	81.3% (13/16)	100% (72/72)	81.3% (13/16)	100% (72/72)
Gain chr12	20	67	90% (18/20)	98.7% (66/67)	80% (16/20)	100% (67/67)
Loss 13q14	32	58	78.1% (25/32)	98.4% (57/58)	78.1% (25/32)	100% (58/58)
Loss 17p13 (TP53)	34	57	88.2% (30/34)	98.4% (56/57)	88.2% (30/34)	100% (57/57)
Gain chr18	13	172	92.3% (12/13)	100% (72/72)	69.2% (9/13)	100% (72/72)
Total			**90%**	**99.3%**	**83.3%**	**100%**

The target alterations *L_CREBBP, L_EP300* and *L_KLF2* are not present, since by conventional techniques no case presenting these alterations was reported.

Additionally to the high detection rate achieved with the panel, alterations identified were highly concordant in terms of size and location with those previously defined by conventional techniques (Figure 59). The size of detected CNA ranged from only 200 Kb to whole chromosome gains or losses. Moreover, among the 206 CNA previously defined by genomic arrays, the median overlap between the exact coordinates established by both methods was 99.97% (range: 0-100). Besides, detailed information regarding the 24 previously known CNA missed by the panel was further analyzed to elucidate potential causes of discordance. Of note, 10 of the 11 CNA with available percentage of abnormal cells by FISH or genomic arrays showed the alteration in less than 30% of tumor cells. The only discordant deletion detected in a higher proportion of cells (50%) was a 9p deletion found by genomic arrays which probably would not have been detected by the panel due to its reduced size (66Kb) (Table 62). Although data regarding the mosaicism level of the remaining CNA were not available, these results suggest that the lowest limit of detection for CNA could be established at 20-30%, which is similar to the generally assumed for genomic microarray platforms.
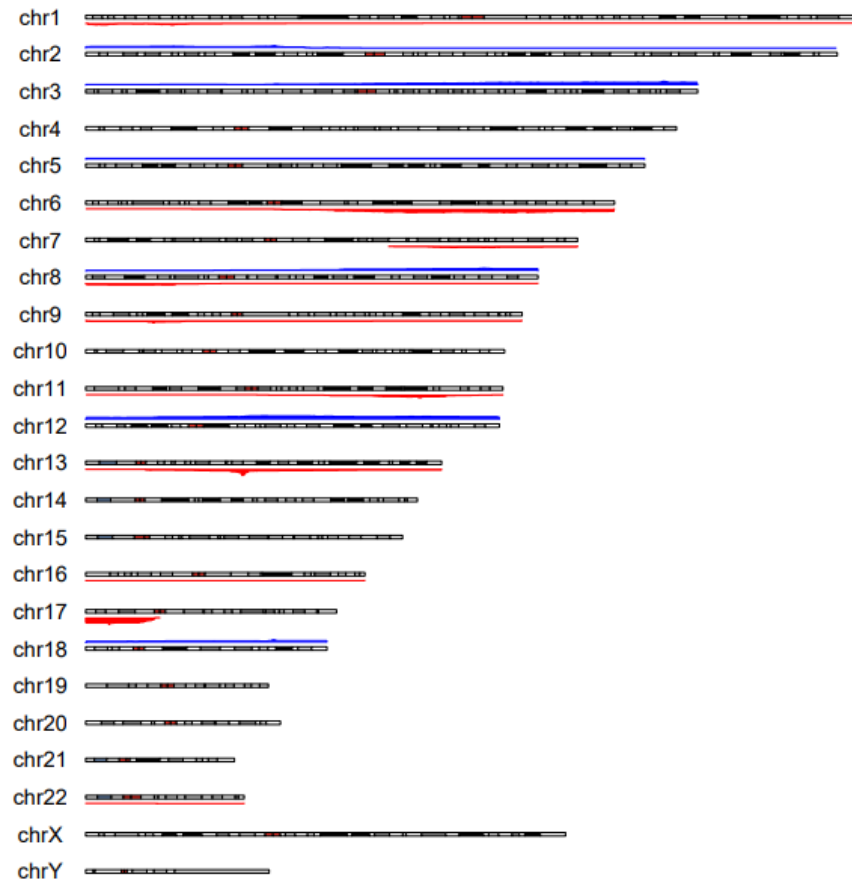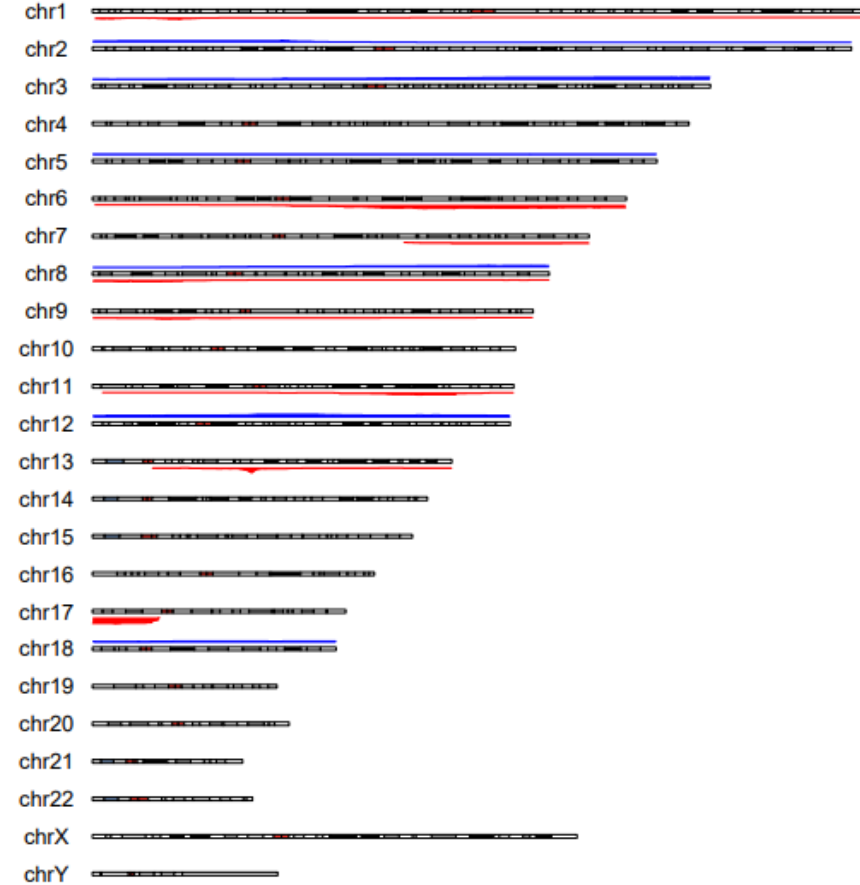
Figure 59. Karyoplots representing the CNA identified by conventional techniques and NGS in patients sequenced with PV1. (A) Alterations detected by conventional techniques (CBA, FISH and/or arrays); (B) CNA defined by CopywriteR (without SNP). CNA in each chromosome are shown in blue (gains) and red (losses) lines, and the thickness of the lines is proportional to the number of cases.

Table 62. Known altered cases not detected by PV1 for each target alteration.

Target CNA	N. of cases	Diagnostic	Technique that identified the alteration	Size (Kb)	% of cells with the alteration
Loss 1p36	1	SMZL	Genomic arrays	323	-
Gain chr 5/5q	1	SMZL	Karyotype	180,915	1/20 metaphases analyzed*
Loss 6q13-q26	2	SMZL	Genomic arrays	8,732	-
			Karyotype	40,815	6/30 metaphases analyzed*
Loss 7q22.1-q35	1	MALT	Karyotype	49,900	2/20 metaphases analyzed*
Loss 9p21	2	DLBCL	Genomic arrays	209	20%
			Genomic arrays	66	50%
Loss 11q22.1-q24	3	CLL	FISH	-	28%
			FISH	-	12%
			Genomic arrays	20,826	-
Gain chr12	2	DLBCL	Genomic arrays	33,376	20%
			Genomic arrays	111,826	-
Loss 13q14	7	CLL	Genomic arrays	31,003	22%
		CLL	FISH	-	21%
		DLBCL	Genomic arrays	115,169,878	-
		CLL	Genomic arrays	1,025	-
		CLL	Genomic arrays	890	-
				5,863	-
		MCL	FISH	-	8%
Loss 17p13 (TP53)	4	CLL	FISH	-	19%
			FISH	-	19%
			FISH	-	10%
			Genomic arrays	670	-
Gain chr18	1	DLBCL	Genomic arrays	78,077	-

In all cases studied by genomic arrays, the same DNA sample was analyzed by both techniques. * Altered metaphases from a total metaphases analyzed, as it could be biased by the *in vitro* proliferative capacity of the altered cells it has not been considered to assess the detection limit of the panel.

When the same CNA analysis was performed in PV2, a large number of false positive results were detected both in tumor and in control samples. For this reason, new parameters were evaluated for PV2. Initially, two bin sizes were tested: 500kb for all the regions and an analysis using a smaller bin of 200kb in those target CNA regions
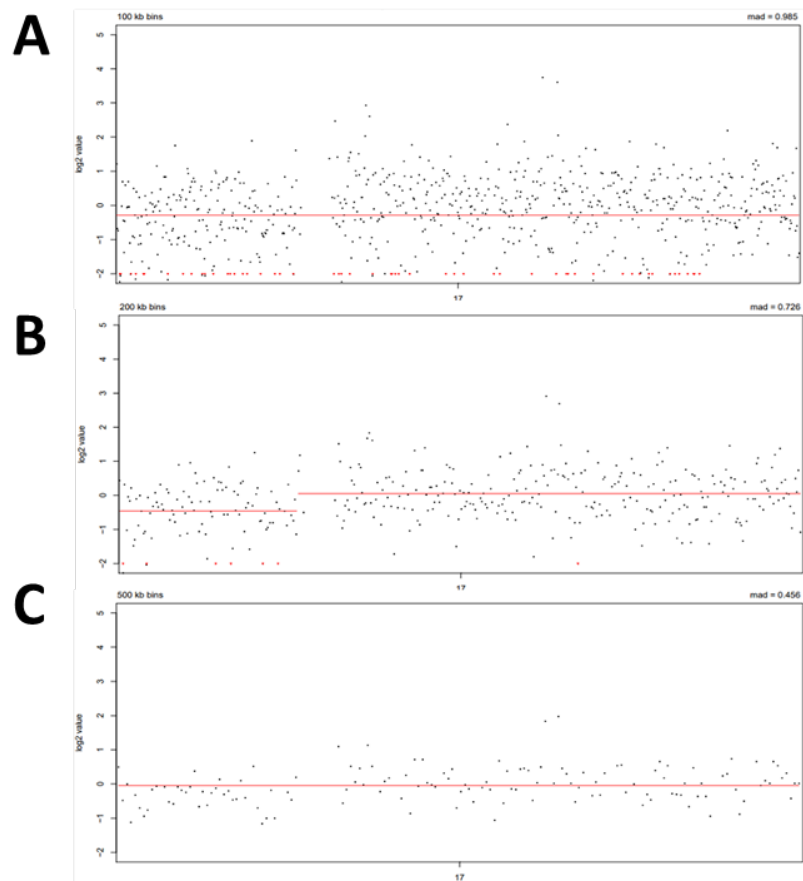
that were expected to be smaller with the aim to improve its detection (Figure 60).

Figure 60. Example of the visualization of the CopyWriteR CNA calls when different bin sizes were applied. The figure shows a region of chromosome 17 observed with different bin sizes: 100Kb (A), 200Kb (B) and 500Kb (C) for the same patient. While in A and C no CNA are found, the 200kb bin shows a deletion involving the 17p13 (*TP53*) that had previously been identified with genomic arrays.

Contrarily to the expected, although both methods displayed similar global specificity (~96%), the 200Kb bin resulted in a lower sensitivity than the 500Kb bin (63.6% vs 68.8%, respectively). Moreover, in those target regions for which sensitivity was improved with 200Kb bin, the specificity was compromised due to the detection of false positive CNA (Table 63). Besides, since some false positive and unexpected *de novo* results showed segment.mean values close to the established thresholds, further analysis by setting more restrictive segment.mean values was carried out. When considering only those CNA with segment.mean values <1.7 (losses) and >2.3 (gains), specificity increased up to ~99% (Table 63). For these reasons, it was decided to take into account those results obtained after applying a bin of 500Kb and restrictive values for segment.mean.

With regard to the detected CNA, it is important to highlight that, as in PV1, a high concordance was observed in terms of location and size with the alterations described by conventional techniques (Figure 61). On the contrary, the sensitivity obtained in PV2 was not as good as expected from the PV1 experience, especially in target regions such as losses in 1p36 (21.4%) or 9p21 (16.7%). In this regard, due to the largest bin used in PV2, it displayed a lower resolution than PV1. The smallest CNA identified was 1Mb in size and only 4/15 (27%) CNA which were known to be below 3Mb were correctly identified. However, discrepances with conventional techniques were not only restricted to small abnormalities. In addition, as 80% of the CNA detected in 30-40% of cells by conventional techniques were correctly identified, the limited sensitivity found in PV2 could not be associated with a lower detection of CNA with low mosaicism than that obtained for PV1. Thereby, 11/16 (68.8%) of missed CNA from which the percentage of abnormal cells was available were found in at least 50% of the analyzed sample (Table 64).

Table 63. Sensitivity and specificity obtained for each target alteration with PV2 analyses.

PV2			NGS					
	Routine techniques		200kb bin		500kb bin		500kb bin + SEG.MEAN filters (>2.3 for gains; <1.7 for losses)	
Target alteration	N of cases altered	N of cases not altered	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Loss 1p36	14	26	35.7% (5/14)	97.8% (25/26)	21.4% (3/14)	100% (26/26)	21.4% (3/14)	100% (26/26)
Gain 2p24-2p12	5	33	60% (3/5)	96.2% (31/33)	60% (3/5)	99.1% (32/33)	60% (3/5)	98.1% (32/33)
Gain chr 3/3q	6	36	-	-	100% (6/6)	100% (36/36)	100% (6/6)	100% (36/36)
Gain chr 5/5q	3	34	-	-	100% (3/3)	100% (34/34)	100% (3/3)	100% (34/34)
Loss 6q13-q26	17	23	70.6% (12/17)	100% (23/23)	76.5% (13/17)	100% (23/23)	76.5% (13/17)	100% (23/23)
Loss 7q22.1-q35	6	41	83.3% (5/6)	100% (41/41)	66.7% (4/6)	100% (41/41)	66.7% (4/6)	100% (41/41)
Loss 8p21.1-p23.1	4	34	75% (3/4)	98.1% (33/33)	100% (4/4)	100% (34/34)	100% (4/4)	100% (34/34)
Gain 8q24 (MYC)	6	32	50% (3/6)	100% (32/32)	66.7% (4/6)	98% (31/32)	66.7% (4/6)	98% (31/32)
Loss 9p21	6	32	66.7% (4/6)	100% (32/32)	16.7% (1/6)	100% (32/32)	16.7% (1/6)	100% (32/32)
Loss 11q22.1-q24	8	33	37.5% (3/8)	100% (33/33)	37.5% (3/8)	100% (33/33)	37.5% (3/8)	100% (33/33)
Gain chr12	12	32	-	-	83.3% (10/12)	95.7% (30/32)	83.3% (10/12)	95.7% (30/32)
Loss 13q14	6	34	50% (3/6)	98.1% (33/34)	66.7% (4/6)	96.2% (32/34)	66.7% (4/6)	96.2% (32/34)
Loss 17p13 (TP53)	11	40	90.9% (10/11)	87.5% (34/40)	72.2% (8/11)	97.9% (39/40)	72.7% (8/11)	97.9% (39/40)
Gain chr18	13	28	-	-	100% (13/13)	93.6% (26/28)	100% (13/13)	93.6% (26/28)
Total			**62.4%**	**96.1%**	**66.7%**	**96.6%**	**66.4%**	**98.6%**

The target alterations *L_CREBBP, L_EP300* and *L_KLF2* are not present, since by conventional techniques no case presenting these alterations was reported.
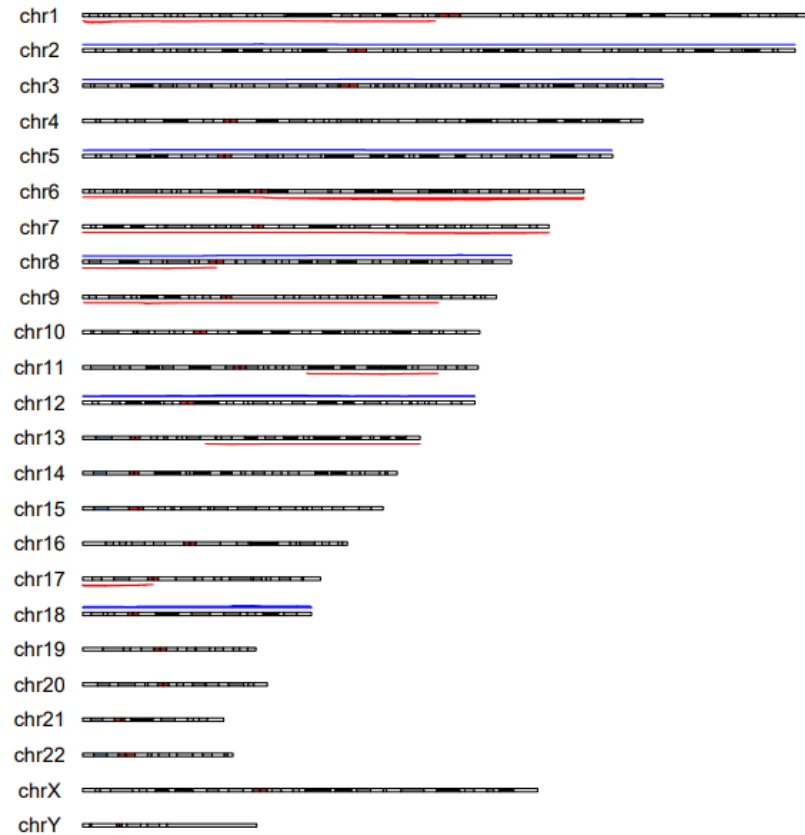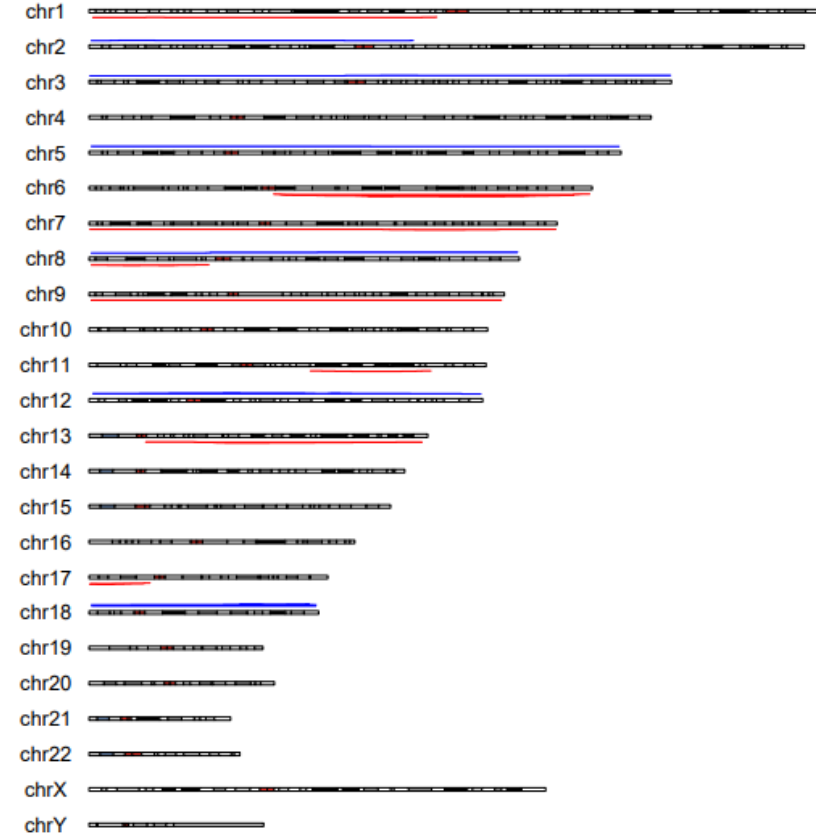
Figure 61. Karyoplots representing the CNA identified by conventional techniques and NGS in patients sequenced with PV2. (A) Alterations detected by conventional techniques (CBA, FISH and/or arrays); (B) CNA defined by CopywriteR (with 500kb bin). CNA in each chromosome are shown in blue (gains) and red (losses) lines, and the thickness of the lines is proportional to the number of cases.

Table 64. Known altered cases not detected by PV2 for each target alteration.

Target CNA	N of cases	Diagnostic	Technique that identified the alteration	Size (Kb)	% of cells with the alteration
Loss 1p36	10	DLBCL	Genomic arrays	6,085	-
		DLBCL	Genomic arrays	7,400	-
		DLBCL	Genomic arrays	11,713	-
		DLBCL	Genomic arrays	623	80%
		FL	Genomic arrays	9,073	-
		DLBCL	Genomic arrays	5,121	90%
		DLBCL	Genomic arrays	1,015	50%
		DLBCL	Genomic arrays	1,878	20%
		DLBCL	Genomic arrays	365	50%
		DLBCL	Genomic arrays	10,685	60%
Gain 2p24-2p12	2	MALT	Genomic arrays	2,923	-
		DLBCL	Genomic arrays	1,021	90%
Loss 6q13-q26	4	FL	Genomic arrays	27,580	-
		DLBCL	Karyotype	171,115	8/10 metaphases analyzed*
		DLBCL	Genomic arrays	2,265	80%
		DLBCL	Genomic arrays	38,619	20%
Loss 7q22.1-q35	2	FL	Genomic arrays	40,414	-
		SMZL	FISH	-	18%
Loss 9p21	4	DLBCL	Genomic arrays	1,516	-
		DLBCL	Genomic arrays	2,338	50%
		DLBCL	Genomic arrays	5,764	60%
		DLBCL	Genomic arrays	180	90%
Loss 11q22.1-q24	3	SMZL	Genomic arrays	1,762	-
		FL	Genomic arrays	24,149	-
		MCL	Genomic arrays	14,451	-
Gain chr12	2	SMZL	Karyotype	133,851.895	4/10 metaphases analyzed*
		SMZL	Karyotype	133,851.895	6/10 metaphases analyzed*
Loss 13q14	2	MCL	Genomic arrays	2,637	-
		DLBCL	Genomic arrays	32,894	-
Loss 17p13 (TP53)	1	DLBCL	Genomic arrays	21,039	-

In all cases studied by genomic arrays, the same DNA sample was analyzed by both techniques. * Altered metaphases from a total metaphases analyzed, as it could be biased by the *in vitro* proliferative capacity of the altered cells it has not been considered to assess the detection limit of the panel.

On the other hand, with regard to the sample included to test the *MYC* amplification detected by FISH, the gene amplification could also be confirmed by NGS (Figure 62).
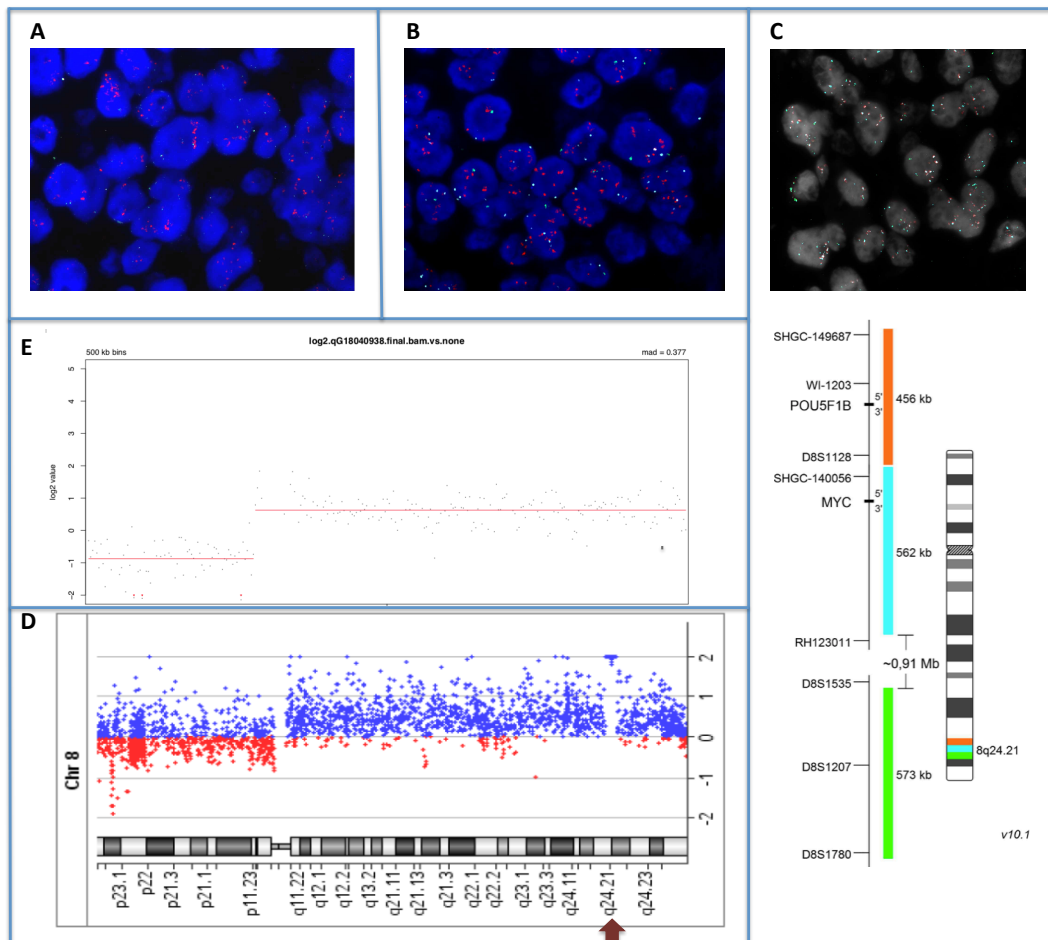
Figure 62. Detail of the *MYC* amplification corresponding to the case diagnosed of FDCS, observed using different approaches. (A) FISH with *MYC* split probe revealed amplification of the orange probe (5 'MYC), involvement of *MYC* could not be confirmed; (B) FISH with *MYC*/IGH dual fusion probe showed amplification of the *MYC* signal, but as it is a large probe, it is not possible to determine whether the observed pattern is due to amplification of a region adjacent to *MYC*; (C) FISH with Tri-Color *MYC* split probe (Metasystems, Heidelberg, Germany) confirmed the *MYC* amplification, since an amplification of the aqua and orange signals is observed; The probe design is shown at the bottom of the image (D); Genomic arrays showed gain of the entire 8q arm in addition to a very small amplification. However, the array used did not include probes covering the *MYC* gene and it could not be determined if it was is involved in the amplification; (E) CopywriteR identified a gain of the entire long arm of chromosome 8 (8q). Although in this image the *MYC* region is not zoomed in, an amplification of the gene was detected by NGS.

Globally, although PV2 CNA analysis allowed to correctly detect most of the known abnormalities with high specificity, the obtained sensitivity was heterogeneous among target regions and it would be insufficient to be implemented in a routine setting. This limitation in sensitivity could be associated with both the performance of the sequencing or the changes made in the design. Regarding the quality of the sequencing, cases in which the alteration was not detected did not present lower coverage when compared with the others. However, it is possible that the higher heterogeneity in the coverage between the samples analyzed with PV2, which the program uses as control samples to performe the baseline, indirectly affects the capability of the tool to infer the CNA in the tested sample. On the other hand, although it was not expected, the removal of some target regions and SNP from the design in PV2 might had an effect on the density and distribution of off-target sequences generated in the sequencing runs, as detailed in section "Sequencing quality analysis". Therefore, even though CopywriteR does not use the on-target reads directly, it indirectly needs the enrichment of these target regions so that higher number of off-target sequences are generated there. This would explain the reduced capability of CopywriteR to analyze CNA in PV2, while the sensitivity was not compromised when SNP removal was bioinformatically simulated in PV1.

Finally, taking all the results into account, we conclude that both the bioinformatics method used, CopywriteR, and the PV1 design are optimal for the detection of CNA in target regions. However, the inclusion of SNP along the genome in the design greatly increases the size of the panel to be used in routine diagnosis. Thus, a third version of the panel including only those SNP located in the target regions should be evaluated. It is expectable that this strategy would allow to considerably reduce the size of the panel offering a sensitivity equivalent to PV1.

4.2. CNA detection in non-target regions

Although it was not the main objective of the panel, we wanted to test if the inclusion of SNP scattered throughout the genome allowed CopywriteR tool to infer CNA also in non-target regions. Therefore, a preliminary study was carried out in a selected cohort of 37 cases diagnosed of CLL from PV1 runs from which previous genomic arrays information was available.

The comparative study showed that 122/144 (85%) alterations previously described by genomic arrays were correctly identified, with sizes ranging from 200Kb to 173Mb. However, CopywriteR also generated 216 discordant CNA. To minimize the false positive rate, CNA in sex chromosomes were discarded, and more restrictive values of segment.mean (<1.6 for losses and >2.4 for gains) were established for non-target regions. Furthermore, those alterations with a size <5Mb were also eliminated. This strategy allowed the identification of 46/61 known CNA ≥5Mb (75%) with a high concordance in the genomic coordinates defined previously, moreover discordant findings were reduced to two.

These results show the ability of the method to detect CNA also in non-target regions using the off-target sequences (Figure 63). However, the sensitivity and the resolution obtained was lower compared to that detected in target regions. The cut-off size established to avoid an unacceptable false positive rate is in line with the recommendations of genomic array guidelines, since <5Mb alterations located outside the target regions have been associated with constitutionally benign variants (Simons et al., 2012; Schoumans et al., 2016). Nonetheless, although they generally should not be reported, some small CNA such as those involved in chromothripsis phenomenon could be clinically relevant (Edelmann et al., 2012; Salaverria et al., 2015) and would be missed by the designed panel. In all, these results suggested that although the panel could offer an orientative information regarding CNA distributed along the genome, it was not equivalent to genomic arrays, therefore, in a routine setting, its analysis should be restricted to target abnormalities.

The poorer results found in non-target regions are consistent with our hypothesis that the distribution and density of the off-target sequences depends on the target regions included in the design. In this regard, since SNP were eliminated in the second version, data of CNA in non-target regions from PV2 have not been analyzed. Indeed, results obtained from PV1 support the idea of restrict the analysis to target regions and partially eliminating the SNP from a third version of the panel, maintaining only those included in large regions related to target CNA.

A **B**

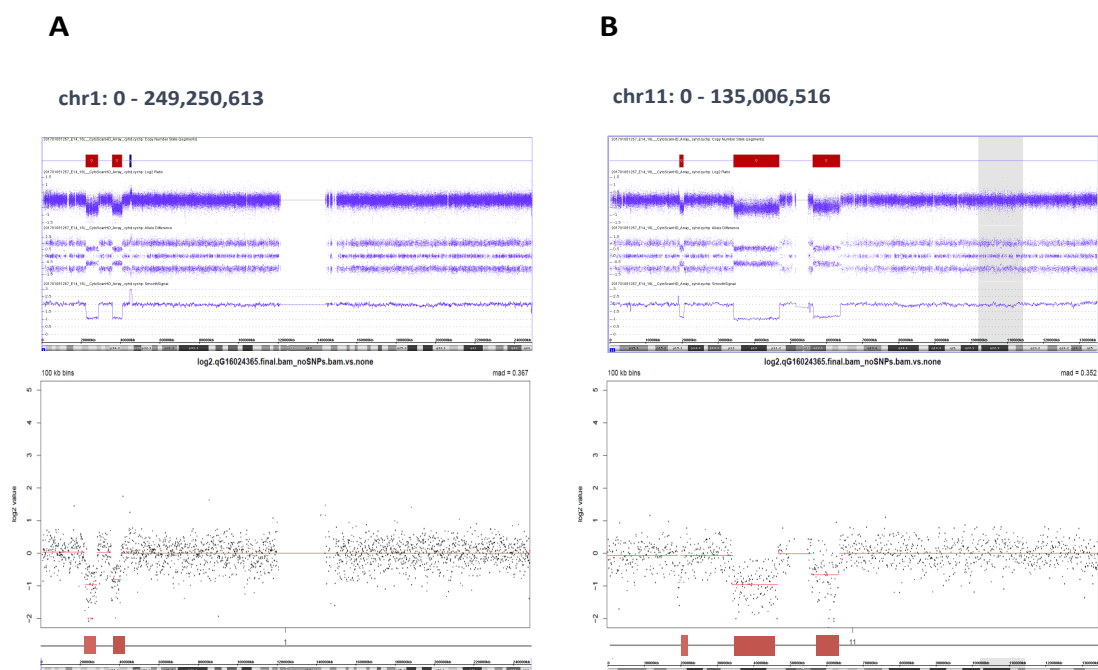**Figure 63. Example of CNA detection by CopywriteR using off-target sequences in non-target regions.** The figure shows two examples (A and B) of the high concordance, in terms of location and size, between the deletions detected by genomic arrays (top) and that detected by CopywriteR with off-target sequences (bottom) in two non-target regions [chr1: 0-249,250,613 (A); and chr11: 0-135,006,516 (B)].

4.3. Clinical implications of detecting CNA with the designed panel

In this section, we will focus mainly on describing the CNA detected in CLL and SMZL, and their clinical impact, since these pathologies are the ones in which CNA had been studied the most, prior to conducting this assay.

4.3.1. CNA detected in chronic lymphocytic leukemia

The genomic landscape of CLL is heterogeneous, lacking a specific cytogenetic abnormality (Puiggros et al., 2014). However, a variety of characteristic, well-stablished cytogenetic abnormalities are observed, including deletions of 13q14, 11q22-23, 17p13, and trisomy 12, which have a role as prognostic markers, this being the main reason for their detection (Döhner et al., 2000). Although current guidelines only recommend as "mandatory" the screening for 17p13 (*TP53*) deletions before any treatment, the study of the four alterations by FISH is a common routine practice (Hallek et al., 2018).

When comparing the capability of the panel to detect these four target CNA with the FISH technique, a sensitivity of 100% was observed for trisomy 12, while it decreased to approximately 78% for deletions (Table 65). Undetected alterations presented a tumor infiltration burden of ≤20% by FISH, this sensitivity cut-off being similar to that obtained with genomic arrays. An association between higher percentages of abnormal nuclei and more adverse outcome has been described for all the four FISH probes (Marasca et al., 2013; Puiggros et al., 2013; González-Gascón et al., 2016; Hernández et al., 2015). Nonetheless, although lower percentage of cells with del(17p) (≤25%) predict better survival [longer time to first treatment (TTFT) and OS] within *TP53*-deleted group, these patients continue to display a shorter TTFT compared with patients without del(17p) (Yuan et al., 2019). Furthermore, studies carried out with ultra-deep-NGS have observed that small subclones with *TP53* aberrations identified before treatment became the predominant population at the time of CLL relapse and anticipated the development of chemorefractory, reflecting the importance of these subclones in the later course of the disease (Rossi et al., 2014). For this reason, it would be advisable that the panel could achieve higher sensitivity to detect abnormalities at very low frequencies, which could probably be

achieved by increasing the depth of coverage of the samples. Regarding specificity, although it was very high for most abnormalities, two additional alterations (a trisomy 12, and a deletion of 13q14) not previously described by FISH and genomic arrays were detected, indicating the possibility of false positive results (Table 65). However, these results should be validated before drawing any conclusions.

Table 65. Comparison of the detection of the four target alterations in CLL by routine FISH assessment and the designed NGS panel.

| Target alteration | FISH | | NGS | | |
	N of cases altered	N of cases not altered	Sensitivity	Specificity	Alteration size (Kb)
Loss 11q22.1-q24 (ATM)	11	26	72.7% (8/11)	100% (26/26)	14,100 - 55,900
Gain chr12	5	32	100% (5/5)	97% (31/32)	133,200,000-133,500,000
Loss 13q14	21	16	76.2% (16/21)	94% (15/16)	200 - 14,100
Loss 17p13 (TP53)	18	19	77.8% (14/18)	100% (19/19)	1,000 – 21,100
Total			**78.6%**	**97.8%**	

Besides, the panel also provided information about the specific region involved in the alteration, which coincided with great accuracy with that previously described by genomic arrays. This information could be informative for alterations such as del(13q14) and del(11q22-q23). Although del(13q14) as sole abnormality has been reported as good prognosis marker, differences have been observed in the course of the disease associated with the size of the deleted region. In this sense, CLL cells with larger deletions that span the *RB1* gene (Type 2 deletion, which encompasses both the DLEU2/MIR15A/MIR16B locus and *RB1*, and the Type 3 deletion, which encompasses the same as Type 2 but with a size greater than 10 Mb) have shown a distinct molecular behaviour and have been associated with a poorer clinical course (Ouillette et al., 2008; 2011; Mosca et al., 2010). In our study, 19% (n=3/16) of the identified deletions corresponded to type 2 (n=2) and type 3 deletions (n=1).

Similarly, del(11q22-q23), which mainly involves deletion of *ATM,* sometimes can involve the neighboring *BIRC3* gene, which may or may not be deleted along with *ATM* (Rossi et al., 2013). *BIRC3* loss represents a poor prognosis marker independent of other factors (Asslaber et al., 2019). In our study, in 87.5% (7/8) of the cases identified with del(11q22-q23) the deletion affected both the *ATM* and *BIRC3* genes. Even though the proven impact in the outcome of these deletion patterns, in clinical practice this information is not considered for the management of CLL patients.

On the other hand, genomic complexity is a relevant prognostic and predictive marker in CLL. A complex karyotype (CK) is defined by the presence of ≥3 chromosomal aberrations (structural and/or numerical) identified by using chromosome-banding analysis. Besides, genomic complexity has also been detected using genomic arrays. In this sense, two recent publications of the European Research Initiative on CLL (ERIC) group have underscored that patients with high complexity (≥5 abnormalities by CBA or ≥5 CNA by GM) are those showing the poorest outcomes (Baliakas et al., 2019; Leeksma et al., 2020). Indeed, genomic complexity represents an independent poor-prognostic marker, also constitutes a novel predictive marker of refractoriness. This has been demonstrated in patients treated not only with standard chemoimmunotherapy regimes (Baliakas et al., 2015; Herling et al., 2016; Puiggros et al., 2017; Baliakas et al., 2019) but also in the initial clinical trials with the novel mechanism-based therapeutic agents such as ibrutinib or venetoclax (Thompson et al., 2015; Chanan-Khan et al., 2016; Anderson et al., 2017), regardless of the presence of *TP53* aberrations.

For this reason, an enriched cohort of CLL patients with CK (22/37) was included to analyze the ability of the panel to detect genomic complexity in these patients. The obtained results were similar to those previously observed with CBA or genomic arrays (Figure 64). The concordance was greater for those alterations in regions considered "target" in the panel. However, the sensitivity obtained in non-target regions was lower, as detailed in the previous section. These results indicated that although the panel could offer an orientative information regarding CNA distributed along the genome, it was not equivalent to genomic arrays or CBA. Therefore, in a routine setting, it would not be useful to assess global genomic complexity, limiting

its analysis to target anomalies. However, as the panel interrogates genomic regions other than those considered target in CLL, it provides additional information to that obtained with FISH.

In contrast to conventional techniques (CBA and FISH), the panel provides information regarding mutational status of several genes in the same assay (Figure 64). This is especially important for *TP53*, as it has been observed that patients with both del(17p) and *TP53* mutations present the shortest TTFT and overall survival (OS) (Yuan et al., 2019). In the same line, CLL patients with trisomy 12, which constitute a heterogeneous group with an intermediate prognosis, when display with *NOTCH1* aberrations, are associated with a more aggressive disease and worse clinical outcome (Autore et al., 2018; Rosati et al., 2018).
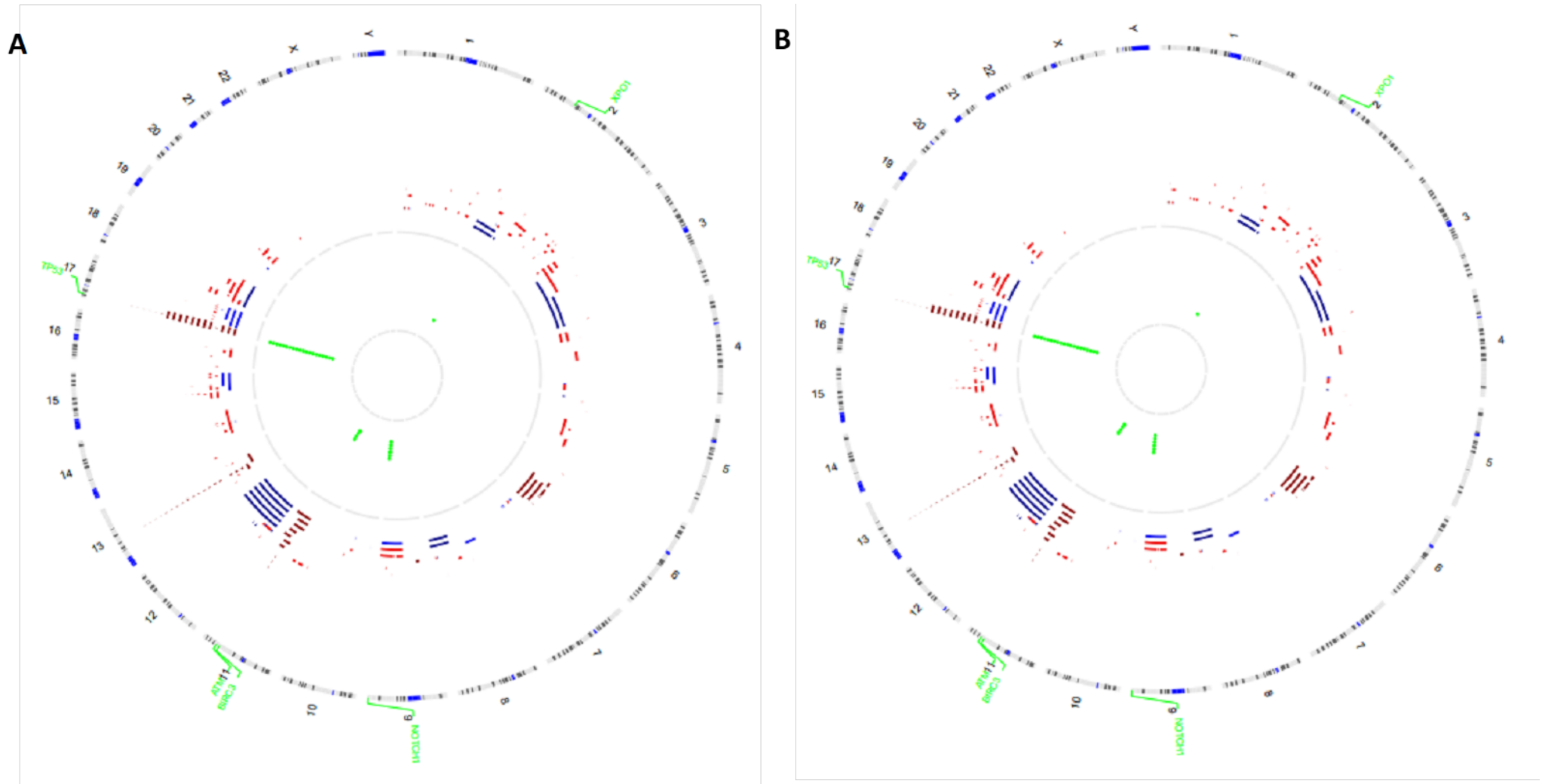
Figure 64. Comparison of CNA detection in CLL patients using conventional techniques and the designed NGS panel. The CNA detected by conventional techniques (**A**) and by NGS (**B**) are shown in two circos plots: the chromosomes are represented in the outermost circumference; then, the CNA identified for each chromosome are shown in red for losses, and blue for gains; finally, the SNV detected are displayed as green dots in the center.

4.3.2. CNA detected in splenic marginal zone lymphoma

SMZL presents great clinical and biological heterogeneity, since although approximately 70% to 80% of cases show cytogenetic and/or molecular genetic abnormalities, none of them is disease-specific (Matutes et al., 2008). However, there are some well-established cytogenetic abnormalities commonly detected in SMZL, including deletions in 7q (30-40% of patients) and gains in chromosome 3/3q (30% of patients). Besides these abnormalities, which have a role as diagnostic markers, the cytogenetic testing currently performed in clinical practice also includes the detection of 17p deletions (*TP53*), asssociated with poor prognosis (Salido et al., 2010; Arcaini et al., 2016).

When comparing the ability of the panel to detect these three target abnormalities with routine techniques, a sensitivity of 95% (n = 20/21) was observed (Table 66). The only missed alteration was a case with a deletion in 7q, detected at 18% of nuclei by FISH, which is consistent with the sensitivity cut-off described for other target alterations. Furthermore, 100% specificity was achieved for all target alterations. Besides, the panel also provided information about the specific region involved in the alterations, which coincided with great accuracy with that previously described by genomic arrays (Figure 65). With regard to 7q deletion, in our study it was found to span from 7q22.1 to 7q36.3, with 7q31.2-q34 identified as the smallest overlapping region of imbalance, which is slightly larger than that reported by previous studies (Gruszka-Westwood et al., 2003; Andersen et al., 2005; Salido et al., 2010). As for gain 3q, although it has been reported that in SMZL the gain in chromosome 3 occurs specifically in 3q (Salido et al., 2010), in our series great variability was observed, which ranged from a more reduced genomic region of 62.2Mb to the entire chromosome. However, 3q21.2-q27.2 was identified as the smallest overlapping region of imbalance, which involves the relevant regions of overrepresentation previously described by other authors (3q21-23, 3q25-29 and 3q23-25) (Gazzo et al., 2003). Finally, unlike the previous alterations, the losses of *TP53* detected involved the whole 17p arm (spanning from 17p13.3 to 17p11.2), except for one case that displayed a smaller deletion of 6Mb which only involved 17p13 cytoband.

Table 66. The alterations identified by conventional techniques (CBA and/or FISH) and detected by NGS for each of the target alterations in SMZL.

Target alteration	Routine techniques		NGS		
	N of cases altered	N of cases not altered	Sensitivity	Specificity	Alteration size (Kb)
Gain 3q	7	32	100% (7/7)	100% (32/32)	10,100,000 – 197,500,000
Loss 7q	8	31	88% (7/8)	100% (31/31)	43,800,000 – 58,200,000
Loss 17p13 (TP53)	8	31	100% (8/8)	100% (31/31)	6,000,000 – 19,800,000
Total			**95.4%**	**98.9%**	

On the other hand, results obtained regarding the detection of non-target CNA were similar to those previously shown in CLL. Although there was a greater concordance between the CNA profiles obtained by the panel and by routine techniques (CBA or genomic arrays) (Figure 65), more strict filtering criteria were needed to avoid false positive results in these regions. Thus, results from SMZL cohort also indicate that the panel is not equivalent to conventional techniques to assess global genomic complexity.

Finally, the panel also provided information on mutations otherwise not tested in our laboratory routine procedures, this being especially important for *KLF2, NOTCH2*, and *TP53* genes (Figure 65). *KLF2* lesions have been identified as the most frequent somatic changes in SMZL (up to 20%) and appear to comprise a subset with a distinct genotype (Clipson et al., 2015; Campos-Martín et al., 2017). Lesions in *KLF2* have been described as early clonal events, generally associated with the 7q deletion (53% 7q-with aberrant *KLF2* versus 11% 7q- without aberrant *KLF2*), and are an independent factor of poor prognosis in SMZL. In our series *KLF2* lesions were detected in four patients, of which three also harbored a 7q deletion (75%). These findings show the possibility that the poor prognosis conferred on 7q deletion by some authors could be due to its association. On the other hand, although it has been described that *KLF2* mutations do not present with deletions in the remaining allele (Piva et al., 2015; Campos-Martín et al., 2017), this evidence could not be

verified in our study. Although the *KLF2* region is covered by the designed panel, and we did not detect any cases with this deletion, we had no cases with a *KLF2* deletion identified by conventional techniques, so we could not verify that the panel detected this alteration.

Furthermore, *NOTCH2* variants have been reported as independent markers of poor outcome in SMZL, and have also been associated with 7q deletions (Parry et al., 2015). In our study, of the five patients identified with *NOTCH2* variants, three (60%) showed 7q deletions, and of these, two cases also harbored lesions in *KLF2*. Of note, 7q deletions combined with somatic variants in *KLF2* and/or *NOTCH2* has recently been described as highly specific for SMZL. On the other hand, abnormalities of the *TP53* gene have also shown to be independent markers of poor prognosis in SMZL (Gruszka-Westwood et al., 2001; Chacón et al. 2002; Parry et al., 2015). In our study, eight cases were identified with *TP53* variants. It is noteworthy that, of these, six cases (75%) also harbored deletions in 17p. To the best of our knowledge, the clinical impact of *TP53* abnormalities when both alleles are altered (by mutations and deletions) has not been described. However, studies carried out in other neoplasms such as CLL support the importance of evaluating *TP53* in an integrated way, both for the analysis of possible deletions and mutations (Yuan et al., 2019). All together, these data reflect the importance of knowing and integrating the information of the different alterations both to perform a differential diagnosis and to establish the real clinical impact that these alterations will have on the patient. In this sense, the panel provides an advantage over the techniques currently available in the diagnostic routine.

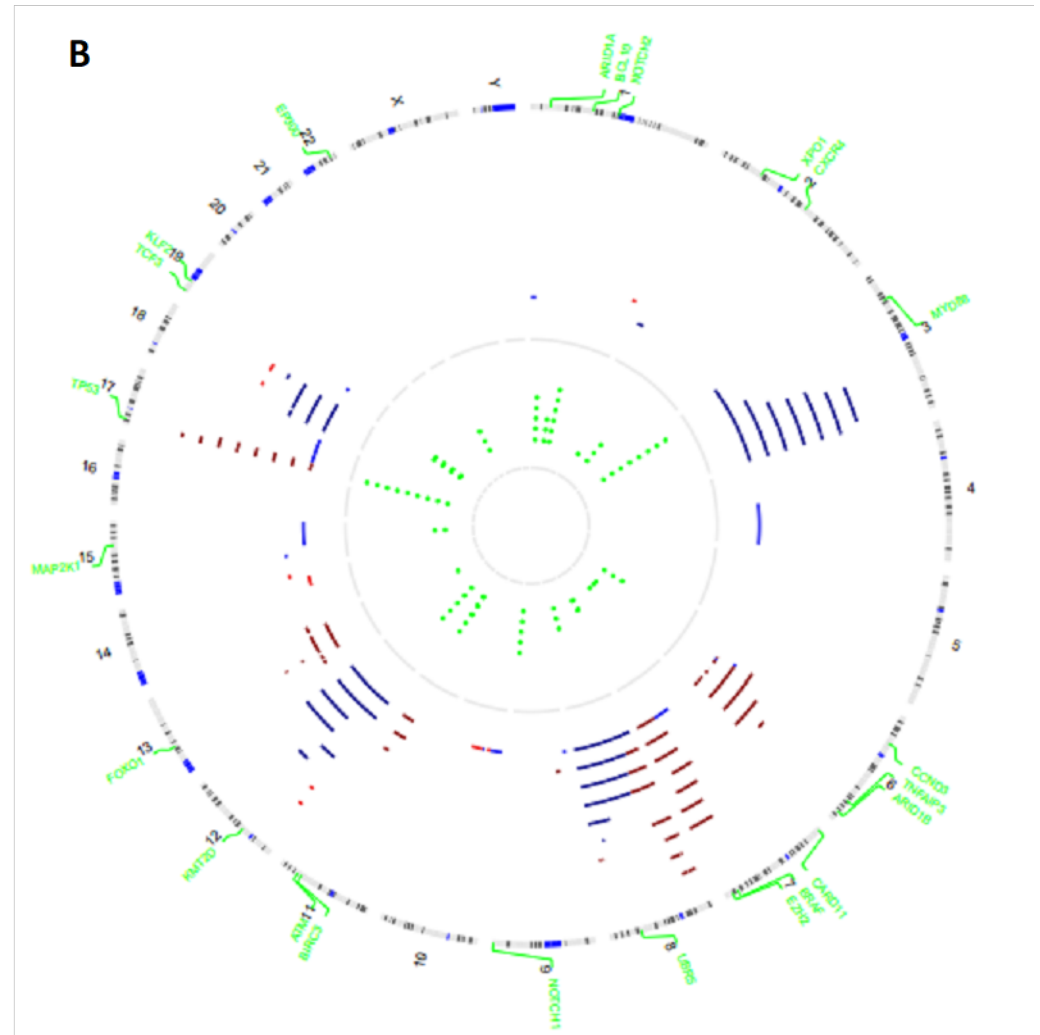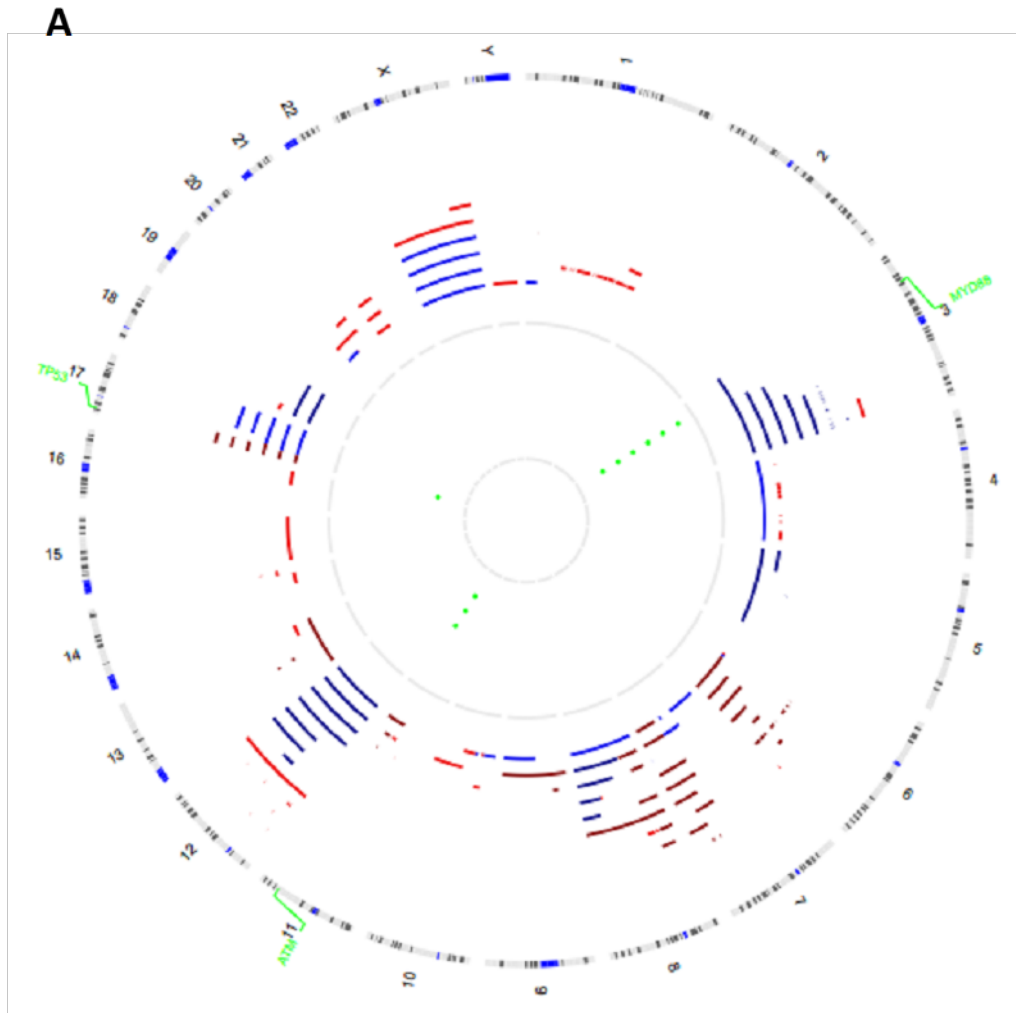**Figure 65. Comparison of CNA detection in SMZL patients using conventional techniques and the designed NGS panel.** The CNA detected by conventional techniques (A) and by NGS (B) are shown in two circos plots: the chromosomes are present in the outermost circumference; then, the CNA identified for each chromosome are shown in red for losses, and blue for gains; finally, the SNV detected are displayed as green dots in the center.

5. Assessment of the panel applicability in FFPE sample sequencing

To evaluate the ability of the panel to analyze FFPE samples, a preliminary study was carried out in four patients (three DLBCL and one FL) that were sequenced in parallel with a fresh tissue sample from the same tumor. Results obtained for each type of abnormality in FFPE samples were compared with both, the data obtained from the sequencing of fresh tissue from the same patient, and with that previously obtained by conventional techniques, when available.

Regarding SNV and indel detection, comparative analysis with the 16 variants previously identified in fresh tissues revealed that all of them were also detected with highly similar VAF in the FFPE sample, reaching a sensitivity of 100% (Figure 66). Besides, none of the detected variants had to be excluded from the analysis for not presenting the established quality parameters. Of note, two additional frameshift variants in the *KMT2D* gene were identified in two patients in the FFPE samples compared to the variants detected in fresh tissue for this gene. These two variants had been previously observed in fresh tissue, further with allelic frequencies similar to those obtained in FFPE (46.7% in FFPE vs 33.4% in fresh tissue; and 16.1% in FFPE vs 13.8% in fresh tissue), but were eliminated from the analysis because they presented <10 supporting variant reads. This discordance is a consequence of the greater coverage depth reached in the sequencing of the FFPE tissue compared to the fresh tissue samples (mean of coverage of 883.84x vs 470.32x). Globally, no false positives were detected, achieving a specificity of 100% (confirmed normal results in 124 and 16 mutation calls, respectively).

With regard to rearrangements, detection was optimal in FFPE tissue samples since 100% of previously identified rearrangements, both by routine techniques and in fresh tissue samples, were detected (n=4/4). Furthermore, a specificity of 100% was achieved. On the other hand, it is noteworthy that the breakpoints detected were located exactly in the same genomic coordinates where they were identified in their corresponding fresh tissue simple (Table 67).
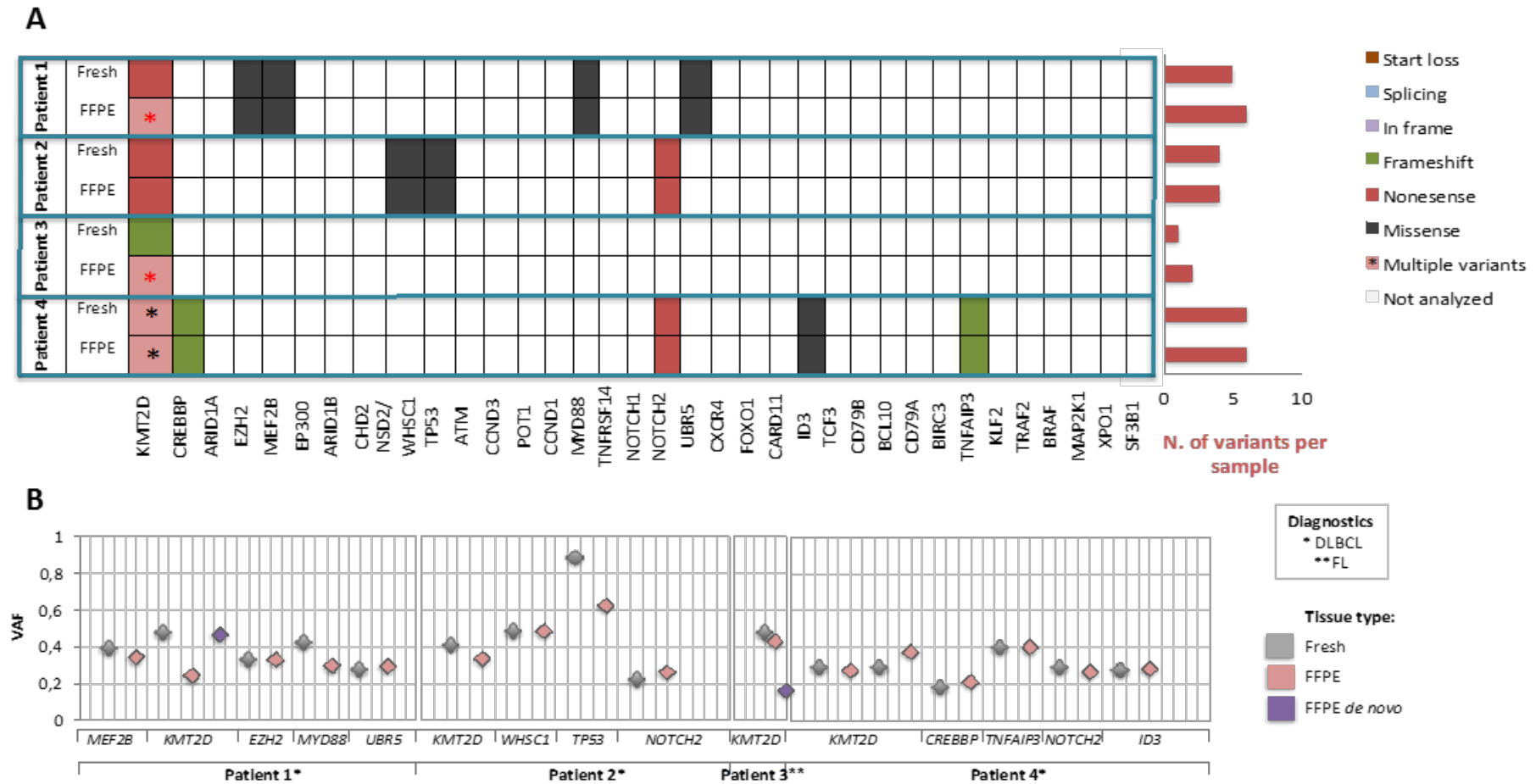
Figure 66. SNV and small indel variants detected the four patients sequenced with paired samples (FFPE and fresh tissue samples). (A) The variants detected in each patient for each of the two tissue types (fresh and FFPE) are shown. Genes with multiple variants are identified with a (*); (B) For each patient, the VAF of the variants detected in each gene with each tissue type (fresh tissue in gray and FFPE tissue in light red) are shown. Those variants identified only in FFPE tissue (de novo FFPE) are outlined in purple.

Table 67. Genomic rearrangements detected in sequenced paired samples

Patient ID	Tissue type	Genomic coordinates in the target gene	Breakpoint location in the target gene	Genomic coordinates in the patner gene	Breakpoint location in the partner gene
Patient 1	Fresh	18:60793600	Exon 3 of *BCL2* (MBR)	14:106329465	IGHJ6
	FFPE	18:60793600	Exon 3 of *BCL2* (MBR)	14:106329465	IGHJ6
Patient 2	Fresh	18:60793548	Exon 3 of *BCL2* (MBR)	14:106382688	IGHJ6
				14:106329462	IGHD2-2
	FFPE	18:60793548	Exon 3 of *BCL2* (MBR)	14:106382688	IGHJ6
				14:106329464	IGHD2-2
Patient 3	Fresh	18:60793447	Exon 3 of *BCL2* (MBR)	14:106363817	IGHD2-15
	FFPE	18:60793447	Exon 3 of *BCL2* (MBR)	14:106363817	IGHD2-15
Patient 4	Fresh	3:187461906	Intron 1 of *BCL6* (MBR)	14:106327132	14q32.33 (*IGH*)
	FFPE	3:187461906	Intron 1 of *BCL6* (MBR)	14:106327132	14q32.33 (*IGH*)

For each of the four patients, the rearrangements identified in the fresh and in the FFPE tissue are shown. One of the patients (Patient 2) shows a double breakpoint in IGH detected in both tissues.

In contrast, some discrepancies were found regarding the detection of CNA in FFPE tissues compared with their paired biopsies. Results obtained showed a slightly inferior sensitivity regarding the identification of the 13 known abnormalities (54% vs 46% in fresh and FFPE samples, respectively) as well as a false positive result not observed in the parallel analysis (specificity: 97.7%) (Figure 67). As shown in the whole cohort, the alterations identified in FFPE tissue samples were largerly concordant in size and location with those previously defined by genomic arrays. Thus, the median overlap between the exact coordinates established by both methods was 96.7% (range: 85.15%-100%) and the size of the CNA detected ranged from 18Mb to complete chromosomal abnormalities.

On the other hand, regarding the undetected alterations, they were mostly found in Patient 3 (Figure 67). However, as the quality parameters of the sequencing obtained for the FFPE tissue of this patient were good, discrepancies could not be attributed to a poorer sequencing performance in this patient. Therefore, since the results from genomic array came from the fresh sample, we could hypothesize that the FFPE tissue presents a smaller fraction of tumor cells with these alterations. However, FISH assessment on FFPE tissue is needed to corroborate this hypothesis.

Regarding the remaining patients, while some CNA were not detected in the FFPE tissue, other alterations not observed in the fresh tissue were identified (Figure 67). In addition, some of the variants considered negative in FFPE tissue were filtered after applying more restrictive values of seg.mean (>2.3 for gains; and <1.7 for losses) defined for the whole cohort of patients with fresh or frozen samples (Figure 67). Therefore, the sensitivity in the detection of CNA could be improved after adjusting the seg.mean values for the analysis of FFPE tissue samples. However, since our cohort was not large enough, the seg.mean values could not be optimized.

Overall, these results suggest the panel's ability to sequence and analyze FFPE tissue samples. However, for CNA, unlike for SNV and translocations, although the results are promising, the analysis should be improved. On the other hand, since this is a preliminary study, more samples should be included in the analysis to draw definitive conclusions.

	Patient 1			Patient 2			Patient 3			Patient 4		
	Routine	Fresh	FFPE	Routine	Fresh	FFPE	Routine	Fresh	FFPE	Routine	Fresh	FFPE
Loss 1p36	■	░	░									
Gain 2p12-2p16												
Gain chr 3												
Gain chr 5/5q	■	█	█									█ (red)
Loss 6q13-q26	■	█	█	■	█	█						
Loss 7q22.1-q35												
Loss 8p21-p23												
Gain 8q24 (MYC)							■	░*	░	■	░	█
Loss 9p21												
Loss 11q22.1-q24							■	░	░*			
Gain chr 12	■	█	█				■	█	░			
Loss 13q14										■	░	░*
Loss 17p13 (TP53)										■	░	█
Gain chr18				■	█	░	■	█	░*			

Legend:
- ■ Altered by routine
- █ Known alteration detected by NGS
- ░ Known alteration missed by NGS
- * Known alteration missed by NGS (filtered by Segment.mean)
- █ (red) False positive by NGS

Figure 67. CNA detected in the four patients sequenced with paired samples (FFPE and fresh tissue samples). The rows show the CNA target that have been analyzed. The columns it is identified if the sample has been altered (for each CNA target) by routine techniques, in the fresh tissue and in the FFPE tissue, for each of the four patients.

6. Global analysis of the designed panel and its application in clinical practice

Taking into account all the results previously exposed, both the designed panel and the pipeline developed for the analysis allow to detect SNV and indels, translocations and CNA in the target regions with a high specificity (nearly 100%) and sensitivity (nearly 90%), being the sensitivity limit ≥ 20-30% of abnormal cells for CNA and rearrangements and ≥ 5% of VAF for SNV and indels. Nonetheless, although the designed panel could potentially be implemented in the clinical practice, some additional studies are still required to improve its performance. First, while PV2 has been optimized for SNV and rearrangements, panel design for CNA detection could be improved. Results obtained for CNA revealed, on the one hand, the need to enrich the target regions of CNA with SNP, and on the other, the futility of the inclusion of SNP in non-target regions. Thus, a third version of the panel that combines the target regions included in PV2 and those SNP of PV1 that were located in the CNA target regions should be evaluated. Second, although the results obtained in PV2 showed an improvement in the detection of *MYC* rearrangements (for which the design had been optimized) compared to PV1, the number of sequenced cases that presented with this gene rearranged was scarce. Therefore, it would be advisable to sequence a greater number of cases with this rearrangement in a third version of the panel to reaffirm the optimization of the *MYC* design carried out for PV2. Finally, although the sensitivity of the designed panel is acceptable and is in line with that obtained by other techniques currently used and that aim to detect this type of alterations (such as Sanger sequencing to detect SNV, or genomic arrays to detect CNA), due to the importance of detecting these aberrations in clinical practice, the sensitivity limit of the panel should be improved in a third design, probably increasing the depth of coverage of sequencing. On the other hand, the preliminary study carried out with paired samples of fresh and FFPE tissue from the same patient also suggests the ability of the panel to efficiently analyze FFPE tissue samples. However, the results obtained for CNA, although promising, indicated the need to optimize the analysis for the study of this particular tissue type. Therefore, a study including a larger cohort of samples must be conducted to apply the panel in this setting.

In recent years, numerous studies have been published characterizing the different hematological malignancies using NGS techniques. Moreover, studies prior to ours have demonstrated the possibility of detecting complex genetic alterations, such as translocations, in the neoplastic process using NGS (Duncavage et al., 2012; Morin et al., 2013). In this regard, targeted panels which allow the identification of translocations in myeloid neoplasms have been described (Grossmann et al., 2011; He et al., 2016; McKerrell et al,. 2016; Stengel et al., 2018; Prieto-Conde et al., 2020) (Table 68). Nontheles, in mature B-cell neoplasms, the application of these techniques has been mainly focused on the description of somatic mutation profiles (Love et al., 2012; Beà et al., 2013; Gaidano & Rossi, 2017; Sujobert et al., 2019) (Table 68). On the other hand, a large number of targeted NGS panels have been commercialized for the detection of SNV and small indel variants, mainly, in solid tumors and hematological neoplasms, with a greater focus on myeloid neoplasms. Besides, some of these panels also allow the analysis of CNA and fusion genes, the latter by amplicon-based sequencing using RNA (TruSight RNA Fusion, Illumina; The Oncomine™ Focus Assay, and The Oncomine™ Precision Assay, Thermo Fisher Scientific; SOPHiA Solid Tumor Plus, SOPHIA GENETICS; and AVENIO NGS Oncology Assay, Nimblegen, among others). However, regarding MBCN, the NGS panels available on the market are scarce. Moreover, due to the technical complexity involved in detecting the characteristic translocations of MBCN (since these translocations do not generate fusion genes, and therefore cannot be analyzed using RNA), the commercialized panels in MBCN have focused on the study of SNV, and recently also in CNA, in hematological neoplasms such as CLL, mainly (Table 69). In this sense, to date, none of the panels available on the market for the study of MBCN can detect structural chromosomal alterations, such as translocations. Furthermore, to our knowledge, this is the first panel developed that allows the integrated analysis of SNV and small indels, CNA and chromosomal rearrangements in MBCN.

In this sense, our panel has the ability, on the one hand, to analyze the different clinically relevant alterations detected today by the diverse genetic techniques applied in routine laboratories, and on the other, to obtain additional information, clinically relevant, and that, otherwise, would not be studied. Thus, this design facilitates its application in routine laboratories, compared to the use of panels aimed exclusively at the study of genetic mutation profiles, since a large part of the information provided by these panels is currently not applied in the clinical management of patients. Furthermore, it provides greater efficiency and scalability than the combination of genetic techniques currently used in the diagnosis of mature B-cell neoplasms, since it allows the simultaneous study of multiple genetic alterations, as well as the parallel analysis of different patients, using a single methodology. On the other hand, MBCN may pose a challenge in their diagnosis, since they often have overlapping morphological and clinical characteristics (Momose et al., 2015; Swerdlow et al., 2016; Wagener et al., 2019). Most of the panels developed focus on the study of a single entity. However, the use of a panel that comprises the genetic alterations involved in different MBCN could help to make a differential diagnosis, in addition to predicting the progression or transformation to a more aggressive lymphoma. Furthermore, this design also allows the genetic characterization of less frequent neoplasms for which designing a specific panel would not be cost effective in clinical practice.

Otherwise, due to the additional financial burden that the use of NGS implies, which includes the expenses related to technical equipment and personnel, composed of a variety of professionals such as molecular and computational biologists, genetic counselors and specialized doctors, in addition to the volume of samples required to keep sequencing costs and turnaround time affordable, we consider that this panel could be applied in reference laboratories or those with a relatively high sequencing activity, since the samples can be multiplexed together with other libraries for their sequencing. Furthermore, given that the panel includes alterations of many neoplasms, it allows to recruit more patients than if it were specific to a single pathology.

Altogether, we have designed a panel with the ability to analyze the distinct clinically relevant alterations currently detected by different genetic techniques, in addition to providing additional, clinically relevant information that, is not currently analyzed in clinical practice, since it does not present actionability in patients. Nevertheless, the incorporation of these additional regions in the design of the panel only represents a few Kb of the total size, so their study does not imply a limitation associated with the increase in the size of the panel. Furthermore, whether or not to analyze this information will be optional and laboratory dependent. However, having this information available and integrating it into the clinical context of patients will help deepen our understanding of these diseases.

Table 68. Targeted NGS panels published in the literature similar to the designed in the thesis.

Reference	Type of sequencing	Target regions included	Diagnosis to which it is directed	Type of sample	Alterations analyzed	Quality criteria	Outstanding results
Grossmann et al., 2011	Targeted sequence capture microarray (Roche Applied Science)	Exons from 92 genes relevant in leukemias; also three array designed to cover *MLL*, *RUNX1* and *PDGFRB* genes	AML, ALL and MPN	DNA	SNV, small indels, fusion genes, and balanced translocations and inversions	-	• Median per-base coverage for each sample: 19x • On-target read percentage: 2.9%- 63% (varied according to the microarray design)
Duncavage et al., 2012	Hybridization-based sequencing (Illumina)	Exons from nine genes of interest in leukemia; and introns and exons from 11 genes frequently translocated in these pathologies	AML	DNA (from five previously-characterized AML cell lines and the BM of a patient with AML	SNV, small indels, and rearrangements	The established parameters were predetermined by the different programs applied	• Total on-target read percentage: 5.7-14.9% • Depth of coverage in each gene: 150±69x (range: 0–2215x) • Identified all published mutations within genes on the capture panel, without false positives • A 177x depth of coverage was sufficient for reliable translocation detection; lower-fold coverage (150x) allowed to detect SNV and indels
Frampton et al., 2013	Hybridization-based sequencing (Illumina)	Exons from 287 cancer-related genes and introns from 19 genes frequently rearranged in solid tumors	Different solid tumors	DNA from routine FFPE samples	SNV, small indels, CNA and fusion genes	>500× by non-PCR duplicate read pairs; >99% exons at >100× Rearrangements: ≥10 chimeric pairs identified Filtering of candidates performed by MQ >30	Detection rate: • SNV: >99% if VAF ≥10% and 99% when VAF <10%; two false-positive calls at VAF <5% • Indels: 97-98% at VAF 10-20%, 88% at 5-10 (Few false-positive calls at VAF <20%) • CNA (copy number ≥8): 99% (91/92) if tumor purity ≥30%; 80% at 20–30% of tumor burden • FFPE: 95% of known alterations identified No information regarding fusion gene detection
He et al., 2016	Hybridization-based sequencing (Illumina)	Genes (265 for RNA and 405 for DNA and target genomic regions of alterations specific for these pathologies	ALL, AML, MDS, NHL, and MM	DNA/RNA from PB, BM, and FFPE	SNV, small indels, CNA and fusion genes	DNA:~500x average unique coverage RNA: >3M unique pairs Median exon coverage of 150-250 were considered qualified; SNV considered: VAF ≥1% for known, ≥5% for novel Indels considered: VAF ≥3% for known, ≥10% for novel Rearrangements: stablished a MQ=30	They identified fusion transcripts present in 10% to 20% of cells

Reference	Type of sequencing	Target regions included	Diagnosis to which it is directed	Type of sample	Alterations analyzed	Quality criteria	Outstanding results
Bolli et al., 2016	Hybridization-based sequencing (Illumina)	246 genes implicated in MM and/or cancer in general 2538 SNPs to detect CNA and allelic frequency changes (more densely tiled in target CNA regions) Whole IGH locus	MM	DNA (from hematopoietic cells lines and patient samples)	SNV, CNA, IGH translocations and tumor-specific V(D)J rearrange-ments	V(D)J rearrangements: • ⩾10 reads in the test sample and none in the controls • Supported by reads mapped with a quality of ⩾27 on one end and >0 on the other • Around a deletion of at least 8Kb of length • Start and end coordinates not duplicated in the control cohort, suggestive of potential artifacts	• Mean coverage of the target region was 155.48x, resulting from a rather low average on-target efficiency of 36.3% • IGH locus can be targeted with good efficiency (130x) • Overall, 90% and 63% of the target region was covered on average >10x or >30x, respectively. • Exons were covered at an average depth of 181x, with 99% and 71% of bases covered at more than 30x and 100x.
McKerrell et al., 2016	Hybridization-based sequencing (Illumina)	All exons from 49 genes known to be recurrently mutated in myeloid malignancies. Previously identified intronic breakpoints at both partner genes of *PML-RARA*, *CBFB-MYH11*, *RUNX1-RUNXT1*, and at the *MLL* gene	AML, MDS.	DNA from PB and BM	SNV and indels, tandem duplication, fusion genes, CNA and CN-LOH variants	Translocations: ⩾3 independent supporting chimeric reads Sequencing and mapping quality >20 and no additional mismatches or indels in the same read SNV and indel: ≥5% VAF and ⩾5 independent reads Missense, frameshift, or nonsense variants: VAF >10%, only if they affected genes known to be targeted by somatic variants	• Average coverage was ≥30× for 94% of target exons and 98% of target SNP • 75% of exons/SNPs covered at ≥70× • They detected CNA involving >20% of cells
Prieto-Conde et al., 2019	Hybridization-based sequencing (Illumina)	32 genes (9 complete coding sequence and 23 hotspots) Intronic or exonic breakpoints of one or both partners of the most frequent translocations	AML and CML	DNA from BM	SNV and rearrange-ments	SNV: genotype quality GQX >80, read depth >100, VAF >5%, and no synonymous variants Rearrangements: ⩾4-5 supporting paired-end reads and at least a mean mapping quality of 20 (filter PASS) (only applied to Delly2) were considered as reliable	• They identified all SNV and small indels • It detected two of the three FLT3-ITD cases previously identified by fragment analysis • Comparing three different open-access algorithms (Lumpy, GASV, and DELLY2), the Lumpy algorithm provided the best results

| Current study | Hybridization-based sequencing (Illumina) | PV1: 11 genes involved in translocations, 33 frequently mutated genes, and 17 CNA regions (with 9,111 SNP)
PV2: 4 genes involved in translocations, 35 frequently mutated genes, and 13 regions involved in CNA. | MBCN | DNA from fresh (from PB, BM, and tissue samples), frozen and FFPE tissue samples | SNV and small indel variants, CNA, and rearrange-ments | Average coverage of >110x
SNV: ≥10 supporting variant reads (unless validated), VAF ≥ 6.4% for PV1 and ≥ 4.2% for PV2
Rearrangements: Non-specific rearrangements were manually filtered
CNA: in PV1: 100Kb bin and segment.mean <1.8 or >2.2; in PV2: 500Kb bin and segment.mean. <1.7 or >2.3; regions with the same copy number state separated by <100Kb were also joined | SNV:
• Specificity of 100% for both, PV1 and PV2; sensitivity of 100% (75/75) for PV1 and 97% (32/33) for PV2
Rearrangements:
• Specificity of 100% for both, PV1 and PV2; sensitivity of 50%-100% for PV1, and 75%-100% for PV2
CNA:
• PV1 displayed a high sensitivity (70.1% to 100%) and a specificity of almost 100% for all known target alterations (97.6% to 100%). Highly concordant in terms of size and location with those previously defined by genomic arrays; PV2 displayed a high specificity (99%) and a sensitivity of 68.8% (16.7%-100%) |

Table 69. Commercial panels currently available for the study of the MBCN.

Commercial platform	Type of sequencing	Source material	Target regions included	Diagnosis to which it is directed	Alterations analyzed	Additional information
Oncomine Lymphoma Panel (ThermoFisher SCIENTIFIC)	Amplicon-based	Specific for DNA extracted from FFPE	25-gene panel	Major B-cell lymphomas, including non-Hodgkin's lymphomas (DLBCL, FL, MCL, and MZL) as well as Hodgkin's lymphoma	SNV	• It allows to modify the design including certain genes (from the inventory in Ion AmpliSeq™ Designer) • Low sample input requirement and robust performance on FFPE tissue • End-to-end research workflow, including bioinformatics and a reporting solution • Specialized support for assay verification
SOPHiA LYMPHOMA SOLUTION™ (SOPHiA GENETICS)	Hybridization-based	DNA	54 relevant genes associated with many B- and T-cell lymphomas	Major B- and T-cell lymphomas	SNV, indels and gene amplifications	• Performance also in FFPE • Specialized support in variant analysis and interpretation • The gene panel can be adapted according to experts' needs
SOPHIA CLL SOLUTION™	Hybridization-based	DNA	Exons of 13 genes with potential prognostic and predictive value of response to treatment in CLL, and genes in CNA regions of CLL [del(11q); del(13q); del(17p); +12	CLL	SNV, indels, CNA and inversions	• Specialized support in variant analysis and interpretation
CLL MASTR Plus (Agilent)	Amplicon-based	DNA	9 selected genes associated with CLL	CLL	SNV and CNA	

Conclusions

1. We have developed a targeted NGS panel that allows the study of genetic alterations with diagnostic, prognostic and therapeutic value in mature B-cell neoplasms, currently determined with different techniques, using a single methodology.

2. We have designed a bioinformatic algorithm to analyze the results of the sequencing panel. For each type of alteration, this algorithm interrogates different bioinformatics tools and performs a comparative analysis of the results obtained. This approach allows increasing the specificity and sensitivity of the panel.

3. The panel design allows the integrated analysis of SNV and small indels, CNA and chromosomal rearrangements of special relevance in the different MBCN, this being the first panel developed with this capacity.

4. In addition to the study of the alterations included in the current clinical guidelines, the designed panel includes other relevant genetic alterations, which may have an impact on the diagnostic routine in the near future.

5. In this sense, the panel also provides relevant information, currently not accessible with other methodologies, such as the identification of rearrangement partners and the exact location of the breakpoints involved in the rearrangement. This information may have clinical implications.

6. We have successfully validated most of the genomic alterations detected by the designed panel using different conventional genetic techniques (CBA, FISH, genomic microarrays, PCR and Sanger sequencing) in addition to other massive sequencing panels for SNV validation. Furthermore, this has allowed us to compare the sensitivity and specificity of the panel with respect to genetic techniques performed in routine laboratories in addition to other NGS panels.

7. Although the designed panel could potentially be implemented in clinical practice, some additional studies are still required to improve its performance, so it would be advisable to carry out a third panel design that integrates the improvements of the two previously designed panels.

References

Adessi, Celine, Gilles Matton, Guidon Ayala, Gerardo Turcatti, Jean Jacques Mermod, Pascal Mayer, and Eric Kawashima. 2000. "Solid Phase DNA Amplification: Characterisation of Primer Attachment and Amplification Mechanisms." *Nucleic Acids Research* 28 (20): 87. https://doi.org/10.1093/nar/28.20.e87.

Adey, Andrew, Hilary G. Morrison, Asan, Xu Xun, Jacob O. Kitzman, Emily H. Turner, Bethany Stackhouse, et al. 2010. "Rapid, Low-Input, Low-Bias Construction of Shotgun Fragment Libraries by High-Density in Vitro Transposition." *Genome Biology* 11 (12). https://doi.org/10.1186/gb-2010-11-12-r119.

Ahmed, Makhdum, Leo Zhang, Krystle Nomie, Laura Lam, and Michael Wang. n.d. "Oncotarget 58638 Www.Impactjournals.Com/Oncotarget Gene Mutations and Actionable Genetic Lesions in Mantle Cell Lymphoma." *Oncotarget*. Vol. 7. www.impactjournals.com/oncotarget/.

Ajay, Subramanian S., Stephen C.J. Parker, Hatice Ozel Abaan, Karin V. Fuentes Fajardo, and Elliott H. Margulies. 2011. "Accurate and Comprehensive Sequencing of Personal Genomes." *Genome Research* 21 (9): 1498–1505. https://doi.org/10.1101/gr.123638.111.

Akasaka, Takashi, Hiroshi Akasaka, Noboru Yonetani, Hitoshi Ohno, Hirohiko Yamabe, Shirou Fukuhara, and Minoru Okuma. 1998. "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*. Vol. 18.

Akasaka, Takashi, Chiyoko Ueda, Masayuki Kurata, Hiroshi Akasaka, Hirohiko Yamabe, Takashi Uchiyama, and Hitoshi Ohno. 2000. "Nonimmunoglobulin (Non-Ig)/BCL6 Gene Fusion in Diffuse Large B-Cell Lymphoma Results in Worse Prognosis than Ig/BCL6." www.genome.wi.mit.edu.

Akhter, Ariz, Etienne Mahe, Lesley Street, Payam Pournazari, Marco Perizzolo, Meer Taher Shabani-Rad, Douglas A. Stewart, and Adnan Mansoor. 2015. "CD10-Positive Mantle Cell Lymphoma: Biologically Distinct Entity or an Aberrant Immunophenotype? Insight, through Gene Expression Profile in a Unique Case Series." *Journal of Clinical Pathology* 68 (10): 844–48. https://doi.org/10.1136/jclinpath-2015-202955.

Akyurek, Nalan, Aysegul Uner, Mustafa Benekli, and Ibrahim Barista. 2012. "Prognostic Significance of MYC, BCL2, and BCL6 Rearrangements in Patients with Diffuse Large B-Cell Lymphoma Treated with Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone plus Rituximab." *Cancer* 118 (17): 4173–83. https://doi.org/10.1002/cncr.27396.

Alizadeh, Ash A, Michael B Eisen, R Eric Davis, Chi Ma, Izidore S Lossos, Andreas Rosenwald, Jennifer C Boldrick, et al. 2000. "Distinct Types of Diffuse Large B-Cell Lymphoma Identi®ed by Gene Expression Pro®ling." *NATURE*. Vol. 403. www.nature.com.

Allen, Eliezer M. Van, Nikhil Wagle, Petar Stojanov, Danielle L. Perrin, Kristian Cibulskis, Sara Marlow, Judit Jane-Valbuena, et al. 2014. "Whole-Exome Sequencing and

Clinical Interpretation of Formalin-Fixed, Paraffin-Embedded Tumor Samples to Guide Precision Cancer Medicine." *Nature Medicine* 20 (6): 682–88. https://doi.org/10.1038/nm.3559.

Amarasinghe, Shanika L., Shian Su, Xueyi Dong, Luke Zappia, Matthew E. Ritchie, and Quentin Gouil. 2020. "Opportunities and Challenges in Long-Read Sequencing Data Analysis." *Genome Biology*. BioMed Central Ltd. https://doi.org/10.1186/s13059-020-1935-5.

Amendola, Laura M., Gail P. Jarvik, Michael C. Leo, Heather M. McLaughlin, Yassmine Akkari, Michelle D. Amaral, Jonathan S. Berg, et al. 2016. "Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium." *American Journal of Human Genetics* 98 (6): 1067–76. https://doi.org/10.1016/j.ajhg.2016.03.024.

Andersen, Claus L., Alicja Gruszka-Westwood, Shayne Atkinson, Estella Matutes, Daniel Catovsky, Rikke K. Pedersen, Bjarne B. Pedersen, et al. 2005. "Recurrent Genomic Imbalances in B-Cell Splenic Marginal-Zone Lymphoma Revealed by Comparative Genomic Hybridization." *Cancer Genetics and Cytogenetics* 156 (2): 122–28. https://doi.org/10.1016/j.cancergencyto.2004.04.026.

Anderson, Mary Ann, Constantine Tam, Thomas E. Lew, Surender Juneja, Manu Juneja, David Westerman, Meaghan Wall, et al. 2017. "Clinicopathological Features and Outcomes of Progression of CLL on the BCL2 Inhibitor Venetoclax." *Blood* 129 (25): 3362–70. https://doi.org/10.1182/blood-2017-01-763003.

Arcaini, Luca, Davide Rossi, and Marco Paulli. 2016. "Review Series INDOLENT B-CELL LYMPHOMA Splenic Marginal Zone Lymphoma: From Genetics to Management." https://doi.org/10.1182/blood-2015-11.

Arons, Evgeny, Laura Roth, Jeffrey Sapolsky, Tara Suntum, Maryalice Stetler-Stevenson, and Robert J. Kreitman. 2011. "Evidence of Canonical Somatic Hypermutation in Hairy Cell Leukemia." *Blood* 117 (18): 4844–51. https://doi.org/10.1182/blood-2010-11-316737.

Asslaber, Daniela, Nathalie Wacht, Michael Leisch, Yuan Qi, Nicole Maeding, Clemens Hufnagl, Bettina Jansko, et al. 2019. "BIRC3 Expression Predicts CLL Progression and Defines Treatment Sensitivity via Enhanced NF-KB Nuclear Translocation." *Clinical Cancer Research* 25 (6): 1901–12. https://doi.org/10.1158/1078-0432.CCR-18-1548.

Autore, Francesco, Paolo Strati, Luca Laurenti, and Alessandra Ferrajoli. 2018. "Morphological, Immunophenotypic, and Genetic Features of Chronic Lymphocytic Leukemia with Trisomy 12: A Comprehensive Review." *Haematologica*. Ferrata Storti Foundation. https://doi.org/10.3324/haematol.2017.186684.

Bachman, Julia. 2013. "Reverse-Transcription PCR (RT-PCR)." In *Methods in Enzymology*, 530:67–74. Academic Press Inc. https://doi.org/10.1016/B978-0-12-420037-1.00002-6.

Balatti, Veronica, Arianna Bottoni, Alexey Palamarchuk, Hansjuerg Alder, Laura Z. Rassenti, Thomas J. Kipps, Yuri Pekarsky, and Carlo M. Croce. 2012. "NOTCH1 Mutations in CLL Associated with Trisomy 12." *Blood* 119 (2): 329–31.

https://doi.org/10.1182/blood-2011-10-386144.

Baliakas, P., A. Hadzidimitriou, L. A. Sutton, D. Rossi, E. Minga, N. Villamor, M. Larrayoz, et al. 2015. "Recurrent Mutations Refine Prognosis in Chronic Lymphocytic Leukemia." *Leukemia* 29 (2): 329–36. https://doi.org/10.1038/leu.2014.196.

Baliakas, Panagiotis, Sabine Jeromin, Michalis Iskas, Anna Puiggros, Karla Plevova, Florence Nguyen-Khac, Zadie Davis, et al. 2019. "Cytogenetic Complexity in Chronic Lymphocytic Leukemia: Definitions, Associations, and Clinical Impact." *Blood* 133 (11): 1205–16. https://doi.org/10.1182/blood-2018-09-873083.

Barber, Domingo F., Mathias Faure, and Eric O. Long. 2004. "LFA-1 Contributes an Early Signal for NK Cell Cytotoxicity." *The Journal of Immunology* 173 (6): 3653–59. https://doi.org/10.4049/jimmunol.173.6.3653.

Barrans, Sharon L., Sheila J.M. O'Connor, Paul A.S. Evans, Faith E. Davies, Roger G. Owen, Andrew P. Haynes, Gareth J. Morgan, and Andrew S. Jack. 2002. "Rearrangement of the BCL6 Locus at 3q27 Is an Independent Poor Prognostic Factor in Nodal Diffuse Large B-Cell Lymphoma." *British Journal of Haematology* 117 (2): 322–32. https://doi.org/10.1046/j.1365-2141.2002.03435.x.

Baseggio, Lucile, Marie Odile Geay, Sophie Gazzo, Françoise Berger, Alexandra Traverse-Glehen, Martine Ffrench, Sandrine Hayette, et al. 2012. "In Non-Follicular Lymphoproliferative Disorders, IGH/BCL2-Fusion Is Not Restricted to Chronic Lymphocytic Leukaemia." *British Journal of Haematology* 158 (4): 489–98. https://doi.org/10.1111/j.1365-2141.2012.09178.x.

Basso, Katia, and Riccardo Dalla-Favera. 2015. "Germinal Centres and B Cell Lymphomagenesis." *Nature Reviews Immunology* 15 (3): 172–84. https://doi.org/10.1038/nri3814.

Basso, Katia, Adam A. Margolin, Gustavo Stolovitzky, Ulf Klein, Riccardo Dalla-Favera, and Andrea Califano. 2005. "Reverse Engineering of Regulatory Networks in Human B Cells." *Nature Genetics* 37 (4): 382–90. https://doi.org/10.1038/ng1532.

Batstone, P. J., and J. R. Goodlad. 2005. "Efficacy of Screening the Intermediate Cluster Region of the Bcl2 Gene in Follicular Lymphomas by PCR." *Journal of Clinical Pathology* 58 (1): 81–82. https://doi.org/10.1136/jcp.2004.018135.

Beà, Sílvia, Itziar Salaverria, Lluís Armengol, Magda Pinyol, Verónica Fernández, Elena M. Hartmann, Pedro Jares, et al. 2009. "Uniparental Disomies, Homozygous Deletions, Amplifications, and Target Genes in Mantle Cell Lymphoma Revealed by Integrative High-Resolution Whole-Genome Profiling." *Blood* 113 (13): 3059–69. https://doi.org/10.1182/blood-2008-07-170183.

Beà, Sílvia, Rafael Valdés-Mas, Alba Navarro, Itziar Salaverria, David Martín-Garcia, Pedro Jares, Eva Giné, et al. 2013. "Landscape of Somatic Mutations and Clonal Evolution in Mantle Cell Lymphoma." *Proceedings of the National Academy of Sciences of the United States of America* 110 (45): 18250–55. https://doi.org/10.1073/pnas.1314608110.

Béguelin, Wendy, Relja Popovic, Matt Teater, Yanwen Jiang, Karen L. Bunting, Monica

Rosen, Hao Shen, et al. 2013. "EZH2 Is Required for Germinal Center Formation and Somatic EZH2 Mutations Promote Lymphoid Transformation." *Cancer Cell* 23 (5): 677–92. https://doi.org/10.1016/j.ccr.2013.04.011.

Behdad, Amir, and Nathanael G. Bailey. 2014. "Diagnosis of Splenic B-Cell Lymphomas in the Bone Marrow: A Review of Histopathologic, Immunophenotypic, and Genetic Findings." *Archives of Pathology and Laboratory Medicine* 138 (10): 1295–1301. https://doi.org/10.5858/arpa.2014-0291-CC.

Bellan, Cristiana, Stefano Lazzi, Michael Hummel, Nazzareno Palummo, Margherita De Santi, Teresa Amato, Joshua Nyagol, et al. 2005. "Immunoglobulin Gene Analysis Reveals 2 Distinct Cells of Origin for EBV-Positive and EBV-Negative Burkitt Lymphomas." *Blood* 106 (3): 1031–36. https://doi.org/10.1182/blood-2005-01-0168.

Bellan, Cristiana, Lazzi Stefano, De Falco Giulia, Emily A. Rogena, and Leoncini Lorenzo. 2009. "Burkitt Lymphoma versus Diffuse Large B-Cell Lymphoma: A Practical Approach." *Hematological Oncology*. https://doi.org/10.1002/hon.914.

Benavides, Mario, Lai Fong Chow-Tsang, Jinsong Zhang, and Hualin Zhong. 2013. "The Novel Interaction between Microspherule Protein Msp58 and Ubiquitin E3 Ligase EDD Regulates Cell Cycle Progression." *Biochimica et Biophysica Acta - Molecular Cell Research* 1833 (1): 21–32. https://doi.org/10.1016/j.bbamcr.2012.10.007.

Benjamin, David, Takuto Sato, Kristian Cibulskis, Gad Getz, Chip Stewart, and Lee Lichtenstein. 2019. "Calling Somatic SNVs and Indels with Mutect2." *BioRxiv*. bioRxiv. https://doi.org/10.1101/861054.

Berger, Michael F., and Elaine R. Mardis. 2018. "The Emerging Clinical Relevance of Genomics in Cancer Medicine." *Nature Reviews Clinical Oncology*. Nature Publishing Group. https://doi.org/10.1038/s41571-018-0002-6.

Bertrand, P., C. Bastard, C. Maingonnat, F. Jardin, C. Maisonneuve, M. N. Courel, P. Ruminy, J. M. Picquenot, and H. Tilly. 2007. "Mapping of MYC Breakpoints in 8q24 Rearrangements Involving Non-Immunoglobulin Partners in B-Cell Lymphomas." *Leukemia* 21 (3): 515–23. https://doi.org/10.1038/sj.leu.2404529.

Bickmore, Wendy A. n.d. "Karyotype Analysis and Chromosome Banding." www.els.net.

Birren, B W, Y Tachi-Iri, U-J Kim, M Nguyen, H Shizuya, J R Korenberg, and M I Simon. 1996. "A Human Chromosome 22 Fosmid Resource: Mapping and Analysis of 96 Clones." *GENOMICS*. Vol. 34.

Bödör, Csaba, Vera Grossmann, Nikolay Popov, Jessica Okosun, Ciarán O'Riain, King Tan, Jacek Marzec, et al. 2013. "EZH2 Mutations Are Frequent and Represent an Early Event in Follicular Lymphoma." *Blood* 122 (18): 3165–68. https://doi.org/10.1182/blood-2013-04-496893.

Bolli, N., Y. Li, V. Sathiaseelan, K. Raine, D. Jones, P. Ganly, F. Cocito, et al. 2016. "A DNA Target-Enrichment Approach to Detect Mutations, Copy Number Changes and Immunoglobulin Translocations in Multiple Myeloma." *Blood Cancer Journal* 6 (9). https://doi.org/10.1038/bcj.2016.72.

Bouska, A., W. Zhang, Q. Gong, J. Iqbal, A. Scuto, J. Vose, M. Ludvigsen, et al. 2017. "Combined Copy Number and Mutation Analysis Identifies Oncogenic Pathways Associated with Transformation of Follicular Lymphoma." *Leukemia* 31 (1): 83–91. https://doi.org/10.1038/leu.2016.175.

Buccheri, Valeria, Wolney Gois Barreto, Laura Maria Fogliatto, Marcelo Capra, Mariana Marchiani, and Vanderson Rocha. 2018. "Prognostic and Therapeutic Stratification in CLL: Focus on 17p Deletion and P53 Mutation." *Annals of Hematology*. Springer Verlag. https://doi.org/10.1007/s00277-018-3503-6.

Buchonnet, G, P Lenain, P Ruminy, S Lepretre, A Stamatoullas, F Parmentier, F Jardin, et al. 2000. "Characterisation of BCL2-JH Rearrangements in Follicular Lymphoma: PCR Detection of 39 BCL2 Breakpoints and Evidence of a New Cluster." *Leukemia*. Vol. 14. www.nature.com/leu.

Calin, George Adrian, Calin Dan Dumitru, Masayoshi Shimizu, Roberta Bichi, Simona Zupo, Evan Noch, Hansjuerg Aldler, et al. 2002. "Frequent Deletions and Down-Regulation of Micro-RNA Genes MiR15 and MiR16 at 13q14 in Chronic Lymphocytic Leukemia." *Proceedings of the National Academy of Sciences of the United States of America* 99 (24): 15524–29. https://doi.org/10.1073/pnas.242606799.

Campo, Elias, and Simon Rule. 2015. "Review Series AGGRESSIVE B-CELL LYMPHOMAS Mantle Cell Lymphoma: Evolving Management Strategies." https://doi.org/10.1182/blood-2014-05.

Campos-Martín, Yolanda, Nerea Martínez, Azahara Martínez-López, Laura Cereceda, Felipe Casado, Patrocinio Algara, David Oscier, et al. 2017. "Clinical and Diagnostic Relevance of NOTCH2 and KLF2 Mutations in Splenic Marginal Zone Lymphoma." *Haematologica*. Ferrata Storti Foundation. https://doi.org/10.3324/haematol.2016.161711.

Cascione, Luciano, Andrea Rinaldi, Alessio Bruscaggin, Chiara Tarantelli, Alberto J. Arribas, Ivo Kwee, Lorenza Pecciarini, et al. 2019. "Novel Insights into the Genetics and Epigenetics of MALT Lymphoma Unveiled by next Generation Sequencing Analyses." *Haematologica*. Ferrata Storti Foundation. https://doi.org/10.3324/haematol.2018.214957.

Chacón, Jose I., Manuela Mollejo, Enriqueta Muñoz, Patricia Algara, Marisol Mateo, Luis Lopez, Jesús Andrade, et al. 2002. "Splenic Marginal Zone Lymphoma: Clinical Characteristics and Prognostic Factors in a Series of 60 Patients." *Blood* 100 (5): 1648–54. https://doi.org/10.1182/blood.v100.5.1648.h81702001648_1648_1654.

Chanan-Khan, Asher, Paula Cramer, Fatih Demirkan, Graeme Fraser, Rodrigo Santucci Silva, Sebastian Grosicki, Aleksander Pristupa, et al. 2016. "Ibrutinib Combined with Bendamustine and Rituximab Compared with Placebo, Bendamustine, and Rituximab for Previously Treated Chronic Lymphocytic Leukaemia or Small Lymphocytic Lymphoma (HELIOS): A Randomised, Double-Blind, Phase 3 Study." *The Lancet Oncology* 17 (2): 200–211. https://doi.org/10.1016/S1470-2045(15)00465-9.

Chang, Sandy. 2013. "Cancer Chromosomes Going to POT1." *Nature Genetics*.

https://doi.org/10.1038/ng.2617.

Chen, Zhong, Rodman Morgan, Carol S Berger, and Avery A Sandberg. 1992. "Application of Fluorescence In Situ Hybridization in Hematological Disorders."

Cheng, Donavan T., Talia N. Mitchell, Ahmet Zehir, Ronak H. Shah, Ryma Benayed, Aijazuddin Syed, Raghu Chandramohan, et al. 2015. "Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology." *Journal of Molecular Diagnostics* 17 (3): 251–64. https://doi.org/10.1016/j.jmoldx.2014.12.006.

Chong, Lauren C., Susana Ben-Neriah, Graham W. Slack, Ciara Freeman, Daisuke Ennishi, Anja Mottok, Brett Collinge, et al. 2018. "High-Resolution Architecture and Partner Genes of MYC Rearrangements in Lymphoma with DLBCL Morphology." *Blood Advances* 2 (20): 2755–65. https://doi.org/10.1182/bloodadvances.2018023572.

Cleary, Michael L., Stephen D. Smith, and Jeffrey Sklar. 1986. "Cloning and Structural Analysis of CDNAs for Bcl-2 and a Hybrid Bcl-2/Immunoglobulin Transcript Resulting from the t(14;18) Translocation." *Cell* 47 (1): 19–28. https://doi.org/10.1016/0092-8674(86)90362-4.

Clipson, A., M. Wang, L. De Leval, M. Ashton-Key, A. Wotherspoon, G. Vassiliou, N. Bolli, et al. 2015. "KLF2 Mutation Is the Most Frequent Somatic Change in Splenic Marginal Zone Lymphoma and Identifies a Subset with Distinct Genotype." *Leukemia* 29 (5): 1177–85. https://doi.org/10.1038/leu.2014.330.

Colt, Joanne S., Scott Davis, Richard K. Severson, Charles F. Lynch, Wendy Cozen, David Camann, Eric A. Engels, Aaron Blair, and Patricia Hartge. 2006. "Residential Insecticide Use and Risk of Non-Hodgkin's Lymphoma." *Cancer Epidemiology Biomarkers and Prevention* 15 (2): 251–57. https://doi.org/10.1158/1055-9965.EPI-05-0556.

Conconi, Annarita, S. Franceschetti, K. Aprile von Hohenstaufen, G. Margiotta-Casaluci, A. Stathis, A. A. Moccia, F. Bertoni, et al. 2015. "Histologic Transformation in Marginal Zone Lymphomas." *Annals of Oncology* 26 (11): 2329–35. https://doi.org/10.1093/annonc/mdv368.

Cotter, F. E., C. Price, J. Meerabux, E. Zucca, and B. D. Young. 1991. "Direct Sequence Analysis of 14(Q+) and 18(q-) Chromosome Junctions at the MBR and MCR Revealing Clustering within the MBR in Follicular Lymphoma." In *Annals of Oncology*, 2:93–97. https://doi.org/10.1093/annonc/2.suppl_2.93.

Cottrell, Catherine E., Hussam Al-Kateb, Andrew J. Bredemeyer, Eric J. Duncavage, David H. Spencer, Haley J. Abel, Christina M. Lockwood, et al. 2014. "Validation of a Next-Generation Sequencing Assay for Clinical Molecular Oncology." *Journal of Molecular Diagnostics* 16 (1): 89–105. https://doi.org/10.1016/j.jmoldx.2013.10.002.

Cuneo, Antonio, G. M. Rigolin, R. Bigoni, A. De Angeli, A. Veronese, F. Cavazzani, A. Bardi, et al. 2004. "Chronic Lymphocytic Leukemia with 6q- Shows Distinct Hematological Features and Intermediate Prognosis." *Leukemia* 18 (3): 476–83.

https://doi.org/10.1038/sj.leu.2403242.

Curiel-Olmo, Soraya, Rufino Mondéjar, Mond´ Mondéjar, Carmen Almaraz, Manuela
Mollejo, Laura Cereceda, Roso Marès, et al. n.d. "Splenic Diffuse Red Pulp Small B-Cell Lymphoma Displays Increased Expression of Cyclin D3 and Recurrent CCND3
Mutations." www.bloodjournal.org.

D'Agaro, Tiziana, Tamara Bittolo, Vanessa Bravin, Michele Dal Bo, Federico Pozzo, Pietro
Bulian, Francesca M. Rossi, et al. 2018. "NOTCH1 Mutational Status in Chronic
Lymphocytic Leukaemia: Clinical Relevance of Subclonal Mutations and Mutation
Types." *British Journal of Haematology*. Blackwell Publishing Ltd.
https://doi.org/10.1111/bjh.14843.

Dankner, Matthew, April A.N. Rose, Shivshankari Rajkumar, Peter M. Siegel, and Ian R.
Watson. 2018. "Classifying BRAF Alterations in Cancer: New Rational Therapeutic
Strategies for Actionable Mutations." *Oncogene* 37 (24): 3183–99.
https://doi.org/10.1038/s41388-018-0171-x.

Davis, R. Eric, Vu N. Ngo, Georg Lenz, Pavel Tolar, Ryan M. Young, Paul B. Romesser,
Holger Kohlhammer, et al. 2010. "Chronic Active B-Cell-Receptor Signalling in
Diffuse Large B-Cell Lymphoma." *Nature* 463 (7277): 88–92.
https://doi.org/10.1038/nature08638.

Deimling, F. Von, J. M. Scharf, T. Liehr, M. Rothe, A. R. Kelter, P. Albers, W. F. Dietrich, L.
M. Kunkel, N. Wernert, and B. Wirth. 1999. "Human and Mouse RAD17 Genes:
Identification, Localization, Genomic Structure and Histological Expression Pattern
in Normal Testis and Seminoma." *Human Genetics* 105 (1–2): 17–27.
https://doi.org/10.1007/s004399900067.

Deutsch, A. J.A., A. Aigelsreiter, E. Steinbauer, M. Frühwirth, H. Kerl, C. Beham-Schmid,
H. Schaider, and P. Neumeister. 2008. "Distinct Signatures of B-Cell Homeostatic
and Activation-Dependent Chemokine Receptors in the Development and
Progression of Extragastric MALT Lymphomas." *Journal of Pathology* 215 (4): 431–
44. https://doi.org/10.1002/path.2372.

Do, Hongdo, and Alexander Dobrovic. 2015. "Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization." *Clinical Chemistry*. American
Association for Clinical Chemistry Inc.
https://doi.org/10.1373/clinchem.2014.223040.

Dohm, Juliane C., Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. 2008.
"Substantial Biases in Ultra-Short Read Data Sets from High-Throughput DNA
Sequencing." *Nucleic Acids Research* 36 (16). https://doi.org/10.1093/nar/gkn425.

Döhner, Hartmut, Konstanze Fischer, Martin Bentz, Katrin Hansen, Axel Benner, Georges
Cabot, Daniela Diehl, et al. 1995. "P53 Gene Deletion Predicts for Poor Survival and
Non-Response to Therapy with Purine Analogs in Chronic B-Cell Leukemias." *Blood*
85 (6): 1580–89. https://doi.org/10.1182/blood.v85.6.1580.bloodjournal8561580.

Döhner, Hartmut, Stephan Stilgenbauer, Axel Benner, Elke Leupolt, Alexander Kröber,
Lars Bullinger, Konstanze Döhner, Martin Bentz, and Peter Lichter. 2000. "Genomic
Aberrations and Survival in Chronic Lymphocytic Leukemia." *New England Journal*

of Medicine 343 (26): 1910–16. https://doi.org/10.1056/nejm200012283432602.

Döhner, Hartmut, Stephan Stilgenbauer, Michael R. James, Axel Benner, Traudel Weilguni, Martin Bentz, Konstanze Fischer, Werner Hunstein, and Peter Lichter. 1997. "11q Deletions Identify a New Subset of B-Cell Chronic Lymphocytic Leukemia Characterized by Extensive Nodal Involvement and Inferior Prognosis." *Blood* 89 (7): 2516–22. https://doi.org/10.1182/blood.v89.7.2516.

Drmanac, Radoje, Andrew B. Sparks, Matthew J. Callow, Aaron L. Halpern, Norman L. Burns, Bahram G. Kermani, Paolo Carnevali, et al. 2010. "Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays." *Science* 327 (5961): 78–81. https://doi.org/10.1126/science.1181498.

Du, Hui, Lei Gao, Jing Luan, Hangfan Zhang, and Taiwu Xiao. 2019. "C-X-C Chemokine Receptor 4 in Diffuse Large B Cell Lymphoma: Achievements and Challenges." *Acta Haematologica*. S. Karger AG. https://doi.org/10.1159/000497430.

Duncavage, Eric J., Haley J. Abel, Philippe Szankasi, Todd W. Kelley, and John D. Pfeifer. 2012. "Targeted next Generation Sequencing of Clinically Significant Gene Mutations and Translocations in Leukemia." *Modern Pathology* 25 (6): 795–804. https://doi.org/10.1038/modpathol.2012.29.

Durham, Benjamin H, Bartlomiej Getta, Sascha Dietrich, Justin Taylor, Helen Won, James M Bogenberger, Sasinya Scott, et al. 2017. "Brief Report LYMPHOID NEOPLASIA Genomic Analysis of Hairy Cell Leukemia Identifies Novel Recurrent Genetic Alterations." *Blood* 130 (14): 1644–48. https://doi.org/10.1182/blood-2017-01.

Edelmann, Jennifer, Karlheinz Holzmann, Florian Miller, Dirk Winkler, Andreas Bühler, Thorsten Zenz, Lars Bullinger, et al. 2012. "High-Resolution Genomic Profiling of Chronic Lymphocytic Leukemia Reveals New Recurrent Genomic Alterations." *Blood* 120 (24): 4783–94. https://doi.org/10.1182/blood-2012-04-423517.

Einaga, Naoki, Akio Yoshida, Hiroko Noda, Masaaki Suemitsu, Yuki Nakayama, Akihisa Sakurada, Yoshiko Kawaji, et al. 2017. "Assessment of the Quality of DNA from Various Formalin-Fixed Paraffin-Embedded (FFPE) Tissues and the Use of This DNA for next-Generation Sequencing (NGS) with No Artifactual Mutation." *PLoS ONE* 12 (5). https://doi.org/10.1371/journal.pone.0176280.

Einerson, R. R., M. E. Law, H. E. Blair, P. J. Kurtin, R. F. McClure, R. P. Ketterling, H. C. Flynn, A. Dogan, and E. D. Remstein. 2006. "Novel FISH Probes Designed to Detect IGK-MYC and IGL-MYC Rearrangements in B-Cell Lineage Malignancy Identify a New Breakpoint Cluster Region Designated BVR2." *Leukemia* 20 (10): 1790–99. https://doi.org/10.1038/sj.leu.2404340.

Eskelund, Christian W., Christina Dahl, Jakob W. Hansen, Maj Westman, Arne Kolstad, Lone B. Pedersen, Carmen P. Montano-Almendras, et al. 2017. "TP53 Mutations Identify Younger Mantle Cell Lymphoma Patients Who Do Not Benefit from Intensive Chemoimmunotherapy." *Blood* 130 (17): 1903–10. https://doi.org/10.1182/blood-2017-04-779736.

Espinet, Blanca, Beatriz Bellosillo, Carme Melero, M. Carmen Vela, Carmen Pedro, Marta Salido, Lara Pijuan, et al. 2008. "FISH Is Better than BIOMED-2 PCR to Detect

IgH/BCL2 Translocation in Follicular Lymphoma at Diagnosis Using Paraffin-Embedded Tissue Sections." *Leukemia Research* 32 (5): 737–42. https://doi.org/10.1016/j.leukres.2007.09.010.

Fabbri, Giulia, Silvia Rasi, Davide Rossi, Vladimir Trifonov, Hossein Khiabanian, Jing Ma, Adina Grunn, et al. 2011. "Analysis of the Chronic Lymphocytic Leukemia Coding Genome: Role of NOTCH1 Mutational Activation." *Journal of Experimental Medicine* 208 (7): 1389–1401. https://doi.org/10.1084/jem.20110921.

Fedurco, Milan, Anthony Romieu, Scott Williams, Isabelle Lawrence, and Gerardo Turcatti. 2006. "BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies." *Nucleic Acids Research* 34 (3). https://doi.org/10.1093/nar/gnj023.

Fernàndez, Veronica, Elena Hartmann, German Ott, Elias Campo, and Andreas Rosenwald. 2005. "Pathogenesis of Mantle-Cell Lymphoma: All Oncogenic Roads Lead to Dysregulation of Cell Cycle and DNA Damage Response Pathways." *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2005.05.019.

Flodr, P., P. Latalova, M. Tichy, Z. Kubova, T. Papajik, M. Svachova, K. Vrzalikova, L. Radova, M. Jarosova, and P. Murray. 2014. "Diffuse Large B-Cell Lymphoma: The History, Current View and New Perspectives." *Neoplasma* 61 (5): 494–504. https://doi.org/10.4149/neo_2014_062.

Forrest, Matthew S., Christine F. Skibola, Tracy J. Lightfoot, Paige M. Bracci, Eleanor V. Willett, Martyn T. Smith, Elizabeth A. Holly, and Eve Roman. 2006. "Polymorphisms in Innate Immunity Genes and Risk of Non-Hodgkin Lymphoma." *British Journal of Haematology* 134 (2): 180–83. https://doi.org/10.1111/j.1365-2141.2006.06141.x.

Frampton, Garrett M., Alex Fichtenholtz, Geoff A. Otto, Kai Wang, Sean R. Downing, Jie He, Michael Schnall-Levin, et al. 2013. "Development and Validation of a Clinical Cancer Genomic Profiling Test Based on Massively Parallel DNA Sequencing." *Nature Biotechnology* 31 (11): 1023–31. https://doi.org/10.1038/nbt.2696.

Francke, Uta (1994). Digitized and differentially shaded human chromosome ideograms for genomic applications. Cytogenetic and Genome Research, 65(3), 206–219. doi:10.1159/000133633

Gaidano, Gianluca, and Davide Rossi. n.d. "The Mutational Landscape of Chronic Lymphocytic Leukemia and Its Impact on Prognosis and Treatment." https://ashpublications.org/hematology/article-pdf/2017/1/329/1250189/hem00046.pdf.

Gao, Juehua, Lo Ann Peterson, Beverly Nelson, Charles Goolsby, and Yi Hua Chen. 2009. "Immunophenotypic Variations in Mantle Cell Lymphoma." *American Journal of Clinical Pathology* 132 (5): 699–706. https://doi.org/10.1309/AJCPV8LN5ENMZOVY.

García-Ramírez, Idoia, Saber Tadros, Inés González-Herrero, Alberto Martín-Lorenzo, Guillermo Rodríguez-Hernández, Dalia Moore, Lucía Ruiz-Roca, et al. 2017. "Crebbp Loss Cooperates with Bcl2 Overexpression to Promote Lymphoma in Mice." *Blood* 129 (19): 2645–56. https://doi.org/10.1182/blood-2016-08-733469.

Garcia, Joe G.N., and Shwu Fan Ma. 2005. "Polymerase Chain Reaction: A Landmark in the History of Gene Technology." *Critical Care Medicine* 33 (12 SUPPL.). https://doi.org/10.1097/01.CCM.0000186782.93865.00.

Garibyan, Lilit, and Nidhi Avashia. 2013. "Polymerase Chain Reaction." *Journal of Investigative Dermatology* 133 (3): 1–4. https://doi.org/10.1038/jid.2013.1.

Godon, A., A. Moreau, P. Talmant, L. Baranger-Papot, F. Geneviève, N. Milpied, M. Zandecki, and H. Avet-Loiseau. 2003. "Is t(14;18) (Q32;Q21) a Constant Finding in Follicular Lymphoma? An Interphase FISH Study on 63 Patients." *Leukemia* 17 (1): 255–59. https://doi.org/10.1038/sj.leu.2402739.

González-Gascón y Marín, Isabel, María Hernández-Sánchez, Ana Eugenia Rodríguez-Vicente, Carmen Sanzo, Anna Aventín, Anna Puiggros, Rosa Collado, et al. 2016. "A High Proportion of Cells Carrying Trisomy 12 Is Associated with a Worse Outcome in Patients with Chronic Lymphocytic Leukemia." *Hematological Oncology* 34 (2): 84–92. https://doi.org/10.1002/hon.2196.

Gonzalez, David, Pilar Martinez, Rachel Wade, Sarah Hockley, David Oscier, Estella Matutes, Claire E. Dearden, Sue M. Richards, Daniel Catovsky, and Gareth J. Morgan. 2011. "Mutational Status of the TP53 Gene as a Predictor of Response and Survival in Patients with Chronic Lymphocytic Leukemia: Results from the LRF CLL4 Trial." *Journal of Clinical Oncology* 29 (16): 2223–29. https://doi.org/10.1200/JCO.2010.32.0838.

Goodwin, Sara, John D. McPherson, and W. Richard McCombie. 2016. "Coming of Age: Ten Years of next-Generation Sequencing Technologies." *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg.2016.49.

Goy, Andre, and Brad Kahl. 2011. "Mantle Cell Lymphoma: The Promise of New Treatment Options." *Critical Reviews in Oncology/Hematology*. https://doi.org/10.1016/j.critrevonc.2010.09.003.

Gozzetti, Ahsundra, and Micbelk M Le Beuu. 2000. "Fluorescence In Situ Hybridization: Uses and Limitations." http://www.

Grande, Bruno M., Daniela S. Gerhard, Aixiang Jiang, Nicholas B. Griner, Jeremy S. Abramson, Thomas B. Alexander, Hilary Allen, et al. 2019. "Genome-Wide Discovery of Somatic Coding and Noncoding Mutations in Pediatric Endemic and Sporadic Burkitt Lymphoma." *Blood* 133 (12): 1313–24. https://doi.org/10.1182/blood-2018-09-871418.

Greenough, Adrienne, and Sandeep S. Dave. 2014. "New Clues to the Molecular Pathogenesis of Burkitt Lymphoma Revealed through Next-Generation Sequencing." *Current Opinion in Hematology* 21 (4): 326–32. https://doi.org/10.1097/MOH.0000000000000059.

Greisman, Harvey A., Zhengfei Lu, Albert G. Tsai, Timothy C. Greiner, Hye Son Yi, and Michael R. Lieber. 2012. "IgH Partner Breakpoint Sequences Provide Evidence That AID Initiates t(11;14) and t(8;14) Chromosomal Breaks in Mantle Cell and Burkitt Lymphomas." *Blood* 120 (14): 2864–67. https://doi.org/10.1182/blood-2012-02-412791.

Grossmann, V., A. Kohlmann, H. U. Klein, S. Schindela, S. Schnittger, F. Dicker, M. Dugas, W. Kern, T. Haferlach, and C. Haferlach. 2011. "Targeted Next-Generation Sequencing Detects Point Mutations, Insertions, Deletions and Balanced Chromosomal Rearrangements as Well as Identifies Novel Leukemia-Specific Fusion Genes in a Single Procedure." *Leukemia* 25 (4): 671–80. https://doi.org/10.1038/leu.2010.309.

Groves, Frank D, Martha S Linet, Lois B Travis, and Susan S Devesa. n.d. "Cancer Surveillance Series: Non-Hodgkin's Lymphoma Incidence by Histologic Subtype in the United States From 1978 Through 1995." http://jnci.oxfordjournals.org/.

Gruszka-Westwood, Alicja M., Rifat A. Hamoudi, Estella Matutes, Esperanza Tuset, and Daniel Catovsky. 2001. "P53 Abnormalities in Splenic Lymphoma With Villous Lymphocytes." *Blood* 97 (11): 3552–58. https://doi.org/10.1182/blood.V97.11.3552.

Gruszka-Westwood, Alicja M., Rifat Hamoudi, Lucy Osborne, Estella Matutes, and Daniel Catovsky. 2003. "Deletion Mapping on the Long Arm of Chromosome 7 in Splenic Lymphoma with Villous Lymphocytes." *Genes Chromosomes and Cancer* 36 (1): 57–69. https://doi.org/10.1002/gcc.10142.

Guikema, Jeroen E.J., Conny De Boer, Eugenia Haralambieva, Laura A. Smit, Carel J.M. Van Noesel, Ed Schuuring, and Philip M. Kluin. 2006. "IGH Switch Breakpoints in Burkitt Lymphoma: Exclusive Involvement of Noncanonical Class Switch Recombination." *Genes Chromosomes and Cancer* 45 (9): 808–19. https://doi.org/10.1002/gcc.20345.

Guo, Jia, Ning Xu, Zengmin Li, Shenglong Zhang, Jian Wu, Hyun Kim Dae, Sano Marma Mong, et al. 2008. "Four-Color DNA Sequencing with 3'-O-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides." *Proceedings of the National Academy of Sciences of the United States of America* 105 (27): 9145–50. https://doi.org/10.1073/pnas.0804023105.

Guruharsha, K. G., Mark W. Kankel, and Spyros Artavanis-Tsakonas. 2012. "The Notch Signalling System: Recent Insights into the Complexity of a Conserved Pathway." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3272.

Hallek, Michael, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillaume Dighiero, Hartmut Döhner, Peter Hillmen, et al. 2008. "Guidelines for the Diagnosis and Treatment of Chronic Lymphocytic Leukemia: A Report from the International Workshop on Chronic Lymphocytic Leukemia Updating the National Cancer Institute-Working Group 1996 Guidelines." *Blood*. https://doi.org/10.1182/blood-2007-06-093906.

Hallek, Michael, Bruce D Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Peter Hillmen, Michael Keating, et al. 2018. "Special Report IwCLL Guidelines for Diagnosis, Indications for Treatment, Response Assessment, and Supportive Management of CLL." http://ashpublications.org/blood/article-pdf/131/25/2745/1465960/blood806398.pdf.

Haralambieva, Eugenia, Ed Schuuring, Stefano Rosati, Carel Van Noesel, Patty Jansen,

Inge Appel, Jeroen Guikema, et al. 2004. "Interphase Fluorescence In Situ Hybridization for Detection of 8q24/MYC Breakpoints on Routine Histologic Sections: Validation in Burkitt Lymphomas from Three Geographic Regions." *Genes Chromosomes and Cancer* 40 (1): 10–18. https://doi.org/10.1002/gcc.20009.

Hartge, Patricia, Joanne S Colt, Richard K Severson, James R Cerhan, Wendy Cozen, David Camann, Shelia Hoar Zahm, and Scott Davis. 2005. "Residential Herbicide Use and Risk of Non-Hodgkin Lymphoma."

Hartmann, Elena M., Elias Campo, George Wright, Georg Lenz, Itziar Salaverria, Pedro Jares, Wenming Xiao, et al. 2010. "Pathway Discovery in Mantle Cell Lymphoma by Integrated Analysis of High-Resolution Gene Expression and Copy Number Profiling." *Blood* 116 (6): 953–61. https://doi.org/10.1182/blood-2010-01-263806.

He, Jie, Omar Abdel-Wahab, Michelle K Nahas, Kai Wang, Raajit K Rampal, Andrew M Intlekofer, Jay Patel, et al. 2016. "Integrated Genomic DNA/RNA Profiling of Hematologic Malignancies in the Clinical Setting Key Points." https://doi.org/10.1182/blood-2015-08.

Head, Steven R., H. Kiyomi Komori, Sarah A. LaMere, Thomas Whisenant, Filip Van Nieuwerburgh, Daniel R. Salomon, and Phillip Ordoukhanian. 2014. "Library Construction for Next-Generation Sequencing: Overviews and Challenges." *BioTechniques* 56 (2): 61–77. https://doi.org/10.2144/000114133.

Healy, Jane A, Adrienne Nugent, Rachel E Rempel, Andrea B Moffitt, Nicholas S Davis, Xiaoyu Jiang, Jennifer R Shingleton, et al. 2016. "Key Points." https://doi.org/10.1182/blood-2015-07.

Henderson, Michelle J., Marcia A. Munoz, Darren N. Saunders, Jennifer L. Clancy, Amanda J. Russell, Brandi Williams, Darryl Pappin, et al. 2006. "EDD Mediates DNA Damage-Induced Activation of CHK2." *Journal of Biological Chemistry* 281 (52): 39990–0. https://doi.org/10.1074/jbc.M602818200.

Herling, Carmen Diana, Marion Klaumünzer, Cristiano Krings Rocha, Janine Altmüller, Holger Thiele, Jasmin Bahlo, Sandra Kluth, et al. 2016. "Complex Karyotypes and KRAS and POT1 Mutations Impact Outcome in CLL after Chlorambucil-Based Chemotherapy or Chemoimmunotherapy." *Blood* 128 (3): 395–404. https://doi.org/10.1182/blood-2016-01-691550.

Hernández, José Ángel, María Hernández-Sánchez, Ana Eugenia Rodríguez-Vicente, Vera Grossmann, Rosa Collado, Cecilia Heras, Anna Puiggros, et al. 2015. "A Low Frequency of Losses in 11q Chromosome Is Associated with Better Outcome and Lower Rate of Genomic Mutations in Patients with Chronic Lymphocytic Leukemia." *PLoS ONE* 10 (12). https://doi.org/10.1371/journal.pone.0143073.

Hockley, S. L., G. J. Morgan, P. E. Leone, B. A. Walker, A. Morilla, M. Else, A. Wotherspoon, et al. 2011. "High-Resolution Genomic Profiling in Hairy Cell Leukemia-Variant Compared with Typical Hairy Cell Leukemia." *Leukemia* 25 (7): 1189–92. https://doi.org/10.1038/leu.2011.47.

Hodges, Emily, Zhenyu Xuan, Vivekanand Balija, Melissa Kramer, Michael N. Molla, Steven W. Smith, Christina M. Middle, et al. 2007. "Genome-Wide in Situ Exon

Capture for Selective Resequencing." *Nature Genetics* 39 (12): 1522–27. https://doi.org/10.1038/ng.2007.42.

Horsman M D, Douglas E, Randy D Gascoyne, Robert W Coupland, Andrew J Coldman, and Sheryle A Adomat. n.d. "Original Article." http://ajcp.oxfordjournals.org/.

Horton, Sarah J., George Giotopoulos, Haiyang Yun, Shabana Vohra, Olivia Sheppard, Rachael Bashford-Rogers, Mamunur Rashid, et al. 2017. "Early Loss of Crebbp Confers Malignant Stem Cell Properties on Lymphoid Progenitors." *Nature Cell Biology* 19 (9): 1093–1104. https://doi.org/10.1038/ncb3597.

Horwitz, Steven M, Andrew D Zelenetz, ; Leo, I Gordon, William G Wierda, Jeremy S Abramson, Ranjana H Advani, et al. 2016. "16 ; Matthew Lunning, DO 17 ; Auayporn Nademanee, MD 18." *JNCCN-Journal of the National Comprehensive Cancer Network |*. Vol. 14. http://education.nccn.org/node/79463;

Hsu, Ronald H. 2016. "Letter to the Editor: To the Editor." *Pediatric Dentistry* 38 (1): 12–13.

Hu, Zhihong, L Jeffrey Medeiros, Zi Chen, Weina Chen, Shaoying Li, Sergej N Konoplev, Xinyan Lu, et al. 2016. "Mantle Cell Lymphoma With MYC Rearrangement A Report of 17 Patients." www.ajsp.com.

Huber, Wolfgang, Vincent J. Carey, Robert Gentleman, Simon Anders, Marc Carlson, Benilton S. Carvalho, Hector Corrada Bravo, et al. 2015. "Orchestrating High-Throughput Genomic Analysis with Bioconductor." *Nature Methods* 12 (2): 115–21. https://doi.org/10.1038/nmeth.3252.

Hudson, Thomas J., Warwick Anderson, Axel Aretz, Anna D. Barker, Cindy Bell, Rosa R. Bernabé, M. K. Bhan, et al. 2010. "International Network of Cancer Genome Projects." *Nature*. https://doi.org/10.1038/nature08987.

Huet, Sarah, Pierre Sujobert, and Gilles Salles. 2018. "From Genetics to the Clinic: A Translational Perspective on Follicular Lymphoma." *Nature Reviews Cancer*. Nature Publishing Group. https://doi.org/10.1038/nrc.2017.127.

Hung, Stacy S., Barbara Meissner, Elizabeth A. Chavez, Susana Ben-Neriah, Daisuke Ennishi, Martin R. Jones, Hennady P. Shulha, et al. 2018. "Assessment of Capture and Amplicon-Based Approaches for the Development of a Targeted Next-Generation Sequencing Pipeline to Personalize Lymphoma Management." *Journal of Molecular Diagnostics* 20 (2): 203–14. https://doi.org/10.1016/j.jmoldx.2017.11.010.

Iacoangeli, Anna, Linda Adzovic, Emily Q. Chen, Rabia Latif Cattie, Gerald A. Soff, and Henri Tiedge. 2018. "Regulatory BC200 RNA in Peripheral Blood of Patients with Invasive Breast Cancer." *Journal of Investigative Medicine* 66 (7): 1055–63. https://doi.org/10.1136/jim-2018-000717.

Iqbal, J., T. C. Greiner, K. Patel, B. J. Dave, L. Smith, J. Ji, G. Wright, et al. 2007. "Distinctive Patterns of BCL6 Molecular Alterations and Their Functional Consequences in Different Subgroups of Diffuse Large B-Cell Lymphoma." *Leukemia* 21 (11): 2332–43. https://doi.org/10.1038/sj.leu.2404856.

Iqbal, Javeed, Warren G. Sanger, Douglas E. Horsman, Andreas Rosenwald, Diane L. Pickering, Bhavana Dave, Sandeep Dave, et al. 2004. "BCL2 Translocation Defines a Unique Tumor Subset within the Germinal Center B-Cell-like Diffuse Large B-Cell Lymphoma." *American Journal of Pathology* 165 (1): 159–66. https://doi.org/10.1016/S0002-9440(10)63284-1.

Isaacson, Peter G.; Du, Ming-Qing (2004). Timeline: MALT lymphoma: from morphology to molecules. , 4(8), 644–653. doi:10.1038/nrc1409

James Watkins, A., Yuanxue Huang, Hongtao Ye, Estelle Chanudet, Nicola Johnson, Rifat Hamoudi, Hongxiang Liu, et al. 2010. "Splenic Marginal Zone Lymphoma: Characterization of 7q Deletion and Its Value in Diagnosis." *Journal of Pathology* 220 (4): 461–74. https://doi.org/10.1002/path.2665.

Janz, Siegfried. 2006. "Myc Translocations in B Cell and Plasma Cell Neoplasms." *DNA Repair* 5 (9–10): 1213–24. https://doi.org/10.1016/j.dnarep.2006.05.017.

Jaramillo Oquendo, Carolina, Helen Parker, David Oscier, Sarah Ennis, Jane Gibson, and Jonathan C. Strefford. 2019. "Systematic Review of Somatic Mutations in Splenic Marginal Zone Lymphoma." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-46906-1.

Jardin, Fabrice, Philippe Ruminy, Christian Bastard, and Hervé Tilly. 2007. "The BCL6 Proto-Oncogene: A Leading Role during Germinal Center Development and Lymphomagenesis." *Pathologie Biologie* 55 (1): 73–83. https://doi.org/10.1016/j.patbio.2006.04.001.

Jares Pedro, and Elías Campo. 2008. "Advances in the Understanding of Mantle Cell Lymphoma." *British Journal of Haematology*. https://doi.org/10.1111/j.1365-2141.2008.07124.x.

Jares Pedro, Dolors Colomer, and Elias Campo. 2007. "Genetic and Molecular Pathogenesis of Mantle Cell Lymphoma: Perspectives for New Targeted Therapeutics." *Nature Reviews Cancer*. https://doi.org/10.1038/nrc2230.

Jares Pedro, Dolors Colomer, and Elias Campo. 2012. "Molecular Pathogenesis of Mantle Cell Lymphoma." *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI61272.

Jarosova, Marie, Eva Kriegova, Petra Schneiderova, Regina Fillerova, Vit Prochazka, Michaela Mikesova, Patrik Flodr, Karel Indrak, and Tomas Papajik. 2016. "A Novel Non-Immunoglobulin (Non-Ig)/BCL6 Translocation in Diffuse Large B-Cell Lymphoma Involving Chromosome 10q11.21 Loci and Review on Clinical Consequences of BCL6 Rearrangements." *Pathology and Oncology Research*. Springer Netherlands. https://doi.org/10.1007/s12253-015-9972-1.

Jennings, Lawrence J., Maria E. Arcila, Christopher Corless, Suzanne Kamel-Reid, Ira M. Lubin, John Pfeifer, Robyn L. Temple-Smolkin, Karl V. Voelkerding, and Marina N. Nikiforova. 2017. "Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists." *Journal of Molecular Diagnostics*. Elsevier B.V. https://doi.org/10.1016/j.jmoldx.2017.01.011.

Jeromin, S., S. Weissmann, C. Haferlach, F. Dicker, K. Bayer, V. Grossmann, T. Alpermann, et al. 2014. "SF3B1 Mutations Correlated to Cytogenetics and Mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 Untreated CLL Patients." *Leukemia* 28 (1): 108–17. https://doi.org/10.1038/leu.2013.263.

Jo Molina, Thierry, Danielle Canioni, Christiane Copie-Bergman, Christian Recher, Josette Brière, Corinne Haioun, Françoise Berger, et al. 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." *Journal of Clinical Oncology* 32 (35): 3996–4003. https://doi.org/10.1200/JCO.2013.54.9493.

Johnson, Nathalie A., Graham W. Slack, Kerry J. Savage, Joseph M. Connors, Susana Ben-Neriah, Sanja Rogic, David W. Scott, et al. 2012. "Concurrent Expression of MYC and BCL2 in Diffuse Large B-Cell Lymphoma Treated with Rituximab plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone." *Journal of Clinical Oncology* 30 (28): 3452–59. https://doi.org/10.1200/JCO.2011.41.0985.

Joseph, Loren, Milena Cankovic, Samuel Caughron, Pranil Chandra, Rajyasree Emmadi, Jill Hagenkord, Stephanie Hallam, et al. 2016. "The Spectrum of Clinical Utilities in Molecular Pathology Testing Procedures for Inherited Conditions and Cancer: A Report of the Association for Molecular Pathology." *Journal of Molecular Diagnostics* 18 (5): 605–19. https://doi.org/10.1016/j.jmoldx.2016.05.007.

Kamps, Rick, Rita D. Brandão, Bianca J. van den Bosch, Aimee D.C. Paulussen, Sofia Xanthoulea, Marinus J. Blok, and Andrea Romano. 2017. "Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification." *International Journal of Molecular Sciences* 18 (2). https://doi.org/10.3390/ijms18020308.

Karger, Barry L., and András Guttman. 2009. "DNA Sequencing by CE." *Electrophoresis*. https://doi.org/10.1002/elps.200900218.

Karube, Kennosuke, and Elias Campo. 2015. "MYC Alterations in Diffuse Large B-Cell Lymphomas." *Seminars in Hematology*. W.B. Saunders. https://doi.org/10.1053/j.seminhematol.2015.01.009.

Karube, Kennosuke, Ying Guo, Junji Suzumiya, Yasuo Sugita, Yuko Nomura, Kohei Yamamoto, Kei Shimizu, et al. 2007. "CD10 MUM1 Follicular Lymphoma Lacks BCL2 Gene Translocation and Shows Characteristic Biologic and Clinical Features." https://doi.org/10.1182/blood.

Kenter, Amy L., Robert Wuerffel, Satyendra Kumar, and Fernando Grigera. 2013. "Genomic Architecture May Influence Recurrent Chromosomal Translocation Frequency in the Igh Locus." *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2013.00500.

Kiel, Mark J., Thirunavukkarasu Velusamy, Bryan L. Betz, Lili Zhao, Helmut G. Weigelin, Mark Y. Chiang, David R. Huebner-Chan, et al. 2012. "Whole-Genome Sequencing Identifies Recurrent Somatic NOTCH2 Mutations in Splenic Marginal Zone

Lymphoma." *Journal of Experimental Medicine* 209 (9): 1553–65. https://doi.org/10.1084/jem.20120910.

Klein, Ulf, and Riccardo Dalla-Favera. 2008. "Germinal Centres: Role in B-Cell Physiology and Malignancy." *Nature Reviews Immunology*. https://doi.org/10.1038/nri2217.

Kluin, Philip M, and Ed Schuuring. 2011. "Molecular Cytogenetics of Lymphoma: Where Do We Stand in 2010?" *Histopathology* 58 (1): 128–44. https://doi.org/10.1111/j.1365-2559.2010.03700.x.

Koh, Dong In, Won Il Choi, Bu Nam Jeon, Choong Eun Lee, Chae Ok Yun, and Man Wook Hur. 2009. "A Novel POK Family Transcription Factor, ZBTB5, Represses Transcription of P21CIP1 Gene." *Journal of Biological Chemistry* 284 (30): 19856–66. https://doi.org/10.1074/jbc.M109.025817.

Koning, Tom J. De, Jan D.H. Jongbloed, Birgit Sikkema-Raddatz, and Richard J. Sinke. 2015. "Targeted Next-Generation Sequencing Panels for Monogenetic Disorders in Clinical Diagnostics: The Opportunities and Challenges." *Expert Review of Molecular Diagnostics* 15 (1): 61–70. https://doi.org/10.1586/14737159.2015.976555.

Korenberg, Julie R.; Rykowski, Mary C. (1988). *Human genome organization: Alu, LINES, and the molecular structure of metaphase chromosome bands. Cell, 53(3), 391–400.* doi:10.1016/0092-8674(88)90159-6

Kozarewa, Iwanka, Javier Armisen, Andrew F. Gardner, Barton E. Slatko, and C. L. Hendrickson. 2015. "Overview of Target Enrichment Strategies." *Current Protocols in Molecular Biology* 2015 (October). https://doi.org/10.1002/0471142727.mb0721s112.

Kridel, Robert, Fong Chun Chan, Anja Mottok, Merrill Boyle, Pedro Farinha, King Tan, Barbara Meissner, et al. 2016. "Histological Transformation and Progression in Follicular Lymphoma: A Clonal Evolution Study." *PLoS Medicine* 13 (12). https://doi.org/10.1371/journal.pmed.1002197.

Kridel, Robert, Barbara Meissner, Sanja Rogic, Merrill Boyle, Adele Telenius, Bruce Woolcock, Jay Gunawardana, et al. 2012. "Whole Transcriptome Sequencing Reveals Recurrent NOTCH1 Mutations in Mantle Cell Lymphoma." *Blood* 119 (9): 1963–71. https://doi.org/10.1182/blood.

Krysiak, Kilannin, Felicia Gomez, Brian S. White, Matthew Matlock, Christopher A. Miller, Lee Trani, Catrina C. Fronick, et al. 2017. "No Title." *Blood* 129 (4): 473–83. http://ashpublications.org/blood/article-pdf/129/4/473/1401948/blood729954.pdf.

Kuilman, Thomas, Arno Velds, Kristel Kemper, Marco Ranzani, Lorenzo Bombardelli, Marlous Hoogstraat, Ekaterina Nevedomskaya, et al. 2015. "CopywriteR: DNA Copy Number Detection from off-Target Sequence Data." *Genome Biology* 16 (1). https://doi.org/10.1186/s13059-015-0617-1.

Kuper-Hommel, Marion J.J., and J. Han J.M. Van Krieken. 2012. "Molecular Pathogenesis and Histologic and Clinical Features of Extranodal Marginal Zone Lymphomas of Mucosa-Associated Lymphoid Tissue Type." *Leukemia and Lymphoma*.

https://doi.org/10.3109/10428194.2011.631157.

Küppers, Ralf. 2005. "Mechanisms of B-Cell Lymphoma Pathogenesis." *Nature Reviews Cancer*. https://doi.org/10.1038/nrc1589.

Küppers, Ralf, Ana B. Sousa, Audrey S. Baur, John G. Strickler, Klaus Rajewsky, and Martin Leo Hansmann. 2001. "Common Germinal-Center B-Cell Origin of the Malignant Cells in Two Composite Lymphomas, Involving Classical Hodgkin's Disease and Either Follicular Lymphoma or B-CLL." *Molecular Medicine* 7 (5): 285–92. https://doi.org/10.1007/bf03402211.

Lai, Raymond, Daniel A Arber, Karen L Chang, C S Wilson, Lawrence M Weiss. 1998. " Frequency of bcl-2 expression in non-Hodgkin's lymphoma: a study of 778 cases with comparison of marginal zone lymphoma and monocytoid B-cell hyperplasia " *Am J Clin Pathol*. Vol. 111. http://ajcp.oxfordjournals.org/.

Lai, Raymond, Lawrence M Weiss, Karen L Chang, and Daniel A Arber. 1999. "Frequency of CD43 Expression in Non-Hodgkin Lymphoma A Survey of 742 Cases and Further Characterization of Rare CD43 + Follicular Lymphomas Downloaded From." *Am J Clin Pathol*. Vol. 111. http://ajcp.oxfordjournals.org/.

Lam, Lloyd T, George Wright, R Eric Davis, Georg Lenz, Pedro Farinha, Lenny Dang, John W Chan, Andreas Rosenwald, Randy D Gascoyne, and Louis M Staudt. n.d. "Cooperative Signaling through the Signal Transducer and Activator of Transcription 3 and Nuclear Factor-B Pathways in Subtypes of Diffuse Large B-Cell Lymphoma." https://doi.org/10.1182/blood.

Landau, Dan A., Eugen Tausch, Amaro N. Taylor-Weiner, Chip Stewart, Johannes G. Reiter, Jasmin Bahlo, Sandra Kluth, et al. 2015. "Mutations Driving CLL and Their Evolution in Progression and Relapse." *Nature* 526 (7574): 525–30. https://doi.org/10.1038/nature15395.

Landegren, Ulf, Robert Kaiser, Jane Sanders, and Leroy Hood. 1988. "A Ligase-Mediated Gene Detection Technique." *Science* 241 (4869): 1077–80. https://doi.org/10.1126/science.3413476.

Langabeer, Stephen E, David O Brien, Anthony M Mcelligott, Michelle Lavin, and Paul V Browne. 2013. "Case Report BRAF V600E-Negative Hairy Cell Leukaemia" 2013 (June 2012): 1–4.

Layer, Ryan M., Colby Chiang, Aaron R. Quinlan, and Ira M. Hall. 2014. "LUMPY: A Probabilistic Framework for Structural Variant Discovery." *Genome Biology* 15 (6). https://doi.org/10.1186/gb-2014-15-6-r84.

Leeksma, Alexander C., Panagiotis Baliakas, Theodoros Moysiadis, Anna Puiggros, Karla Plevova, Anne Marie van der Kevie-Kersemaekers, Hidde Posthuma, et al. 2020. "Genomic Arrays Identify High-Risk Chronic Lymphocytic Leukemia with Genomic Complexity: A Multi-Center Study." *Haematologica* 105 (5). https://doi.org/10.3324/HAEMATOL.2019.239947.

Leng, Wendy W.J. De, Christa G. Gadellaa-Van Hooijdonk, Françoise A.S. Barendregt-

Smouter, Marco J. Koudijs, Ies Nijman, John W.J. Hinrichs, Edwin Cuppen, et al. 2016. "Targeted next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material." *PLoS ONE* 11 (2): 1–18. https://doi.org/10.1371/journal.pone.0149405.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, Marilyn M., Michael Datto, Eric J. Duncavage, Shashikant Kulkarni, Neal I. Lindeman, Somak Roy, Apostolia M. Tsimberidou, et al. 2017. "Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists." *Journal of Molecular Diagnostics* 19 (1): 4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.

Li, Po E., Chien Chi Lo, Joseph J. Anderson, Karen W. Davenport, Kimberly A. Bishop-Lilly, Yan Xu, Sanaa Ahmed, Shihai Feng, Vishwesh P. Mokashi, and Patrick S.G. Chain. 2017. "Enabling the Democratization of the Genomics Revolution with a Fully Integrated Web-Based Bioinformatics Platform." *Nucleic Acids Research* 45 (1): 67–80. https://doi.org/10.1093/nar/gkw1027.

Liehr, T, A Weise, A Heller, H Starke, K Mrasek, A Kuechler, H.-U G Weier, U Claussen, and Thomas Liehr. 2002. "Multicolor Chromosome Banding (MCB) with YAC/BAC-Based Probes and Region-Specific Microdissection DNA Libraries Request Reprints from Dr." *Cytogenet Genome Res*. Vol. 97. www.karger.comwww.karger.com/journals/cgr.

Liu, Delong, Josif Shimonov, Suneeta Primanneni, Yongrong Lai, Tauseef Ahmed, and Karen Seiter. 2007. "T(8;14;18): A 3-Way Chromosome Translocation in Two Patients with Burkitt's Lymphoma/Leukemia." *Molecular Cancer* 6 (June). https://doi.org/10.1186/1476-4598-6-35.

Liu, Yang, and Stefan Klaus Barta. 2019. "Diffuse Large B-Cell Lymphoma: 2019 Update on Diagnosis, Risk Stratification, and Treatment." *American Journal of Hematology*. Wiley-Liss Inc. https://doi.org/10.1002/ajh.25460.

Lohr, Jens G., Petar Stojanov, Michael S. Lawrence, Daniel Auclair, Bjoern Chapuy, Carrie Sougnez, Peter Cruz-Gordillo, et al. 2012. "Discovery and Prioritization of Somatic Mutations in Diffuse Large B-Cell Lymphoma (DLBCL) by Whole-Exome Sequencing." *Proceedings of the National Academy of Sciences of the United States of America* 109 (10): 3879–84. https://doi.org/10.1073/pnas.1121343109.

Lossos, Izidore S, Debra K Czerwinski, Ash A Alizadeh, Mark A Wechser, Rob Tibshirani, David Botstein, and Ronald Levy. 2004. "Prediction of Survival in Diffuse Large-B-Cell Lymphoma Based on the Expression of Six Genes." *N Engl J Med*. Vol. 18. www.nejm.org.

Lossos, Izidore S, Carol D Jones, Roger Warnke, Yasodha Natkunam, Herbert Kaizer, James L Zehnder, Rob Tibshirani, and Ronald Levy. 2001. "Expression of a Single

Gene, BCL-6, Strongly Predicts Survival in Patients with Diffuse Large B-Cell Lymphoma." www.bloodjournal.org.

Love, Cassandra, Zhen Sun, Dereje Jima, Guojie Li, Jenny Zhang, Rodney Miles, Kristy L. Richards, et al. 2012. "The Genetic Landscape of Mutations in Burkitt Lymphoma." *Nature Genetics* 44 (12): 1321–25. https://doi.org/10.1038/ng.2468.

Lu, Zhengfei, Albert G Tsai, Takashi Akasaka, Hitoshi Ohno, Yanwen Jiang, Ari M Melnick, Harvey A Greisman, and Michael R Lieber. 2013. "Brief Report LYMPHOID NEOPLASIA BCL6 Breaks Occur at Different AID Sequence Motifs in Ig-BCL6 and Non-Ig-BCL6 Rearrangements." https://doi.org/10.1182/blood-2012-10.

Mandelker, Diana, Sami S. Amr, Trevor Pugh, Sivakumar Gowrisankar, Rimma Shakhbatyan, Elizabeth Duffy, Mark Bowser, et al. 2014. "Comprehensive Diagnostic Testing for Stereocilin: An Approach for Analyzing Medically Important Genes with High Homology." *Journal of Molecular Diagnostics* 16 (6): 639–47. https://doi.org/10.1016/j.jmoldx.2014.06.003.

Marasca, Roberto, Rossana Maffei, Silvia Martinelli, Stefania Fiorcari, Jenny Bulgarelli, Giulia Debbia, Davide Rossi, et al. 2013. "Clinical Heterogeneity of de Novo 11q Deletion Chronic Lymphocytic Leukaemia: Prognostic Relevance of Extent of 11q Deleted Nuclei inside Leukemic Clone." *Hematological Oncology* 31 (2): 348–55. https://doi.org/10.1002/hon.2028.

Mardis, Elaine R. 2011. "A Decade's Perspective on DNA Sequencing Technology." *Nature* 470 (7333): 198–203. https://doi.org/10.1038/nature09796.

Mardis, Elaine R. 2013. "Next-Generation Sequencing Platforms." *Annual Review of Analytical Chemistry* 6 (March): 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628.

Mark, Hong Fong L., H. Wyandt, X. L. Huang, and J. M. Milunsky. 2005. "Delineation of a Supernumerary Marker Chromosome Utilizing a Multimodal Approach of G-Banding, Flourescent in Situ Hybridization, Confirmatory P1 Artificial Chromosome Flourescent in Situ Hybridization, and High-Resolution Comparative Genomic Hydribization." *Clinical Genetics* 68 (2): 146–51. https://doi.org/10.1111/j.1399-0004.2005.00466.x.

Masir, Noraidah, Lisa J. Campbell, Lindsey K. Goff, Margaret Jones, Teresa Marafioti, Jacqueline Cordell, Andrew J. Clear, T. Andrew Lister, David Y. Mason, and Abigail M. Lee. 2009. "BCL2 Protein Expression in Follicular Lymphomas with t(14;18) Chromosomal Translocations." *British Journal of Haematology* 144 (5): 716–25. https://doi.org/10.1111/j.1365-2141.2008.07528.x.

Matutes, E., D. Oscier, C. Montalban, F. Berger, E. Callet-Bauchu, A. Dogan, P. Felman, et al. 2008. "Splenic Marginal Zone Lymphoma Proposals for a Revision of Diagnostic, Staging and Therapeutic Criteria." *Leukemia*. Nature Publishing Group. https://doi.org/10.1038/sj.leu.2405068.

Matutes, Estella, Alejandra Martínez-Trillos, and Elias Campo. 2015. "Hairy Cell Leukaemia-Variant: Disease Features and Treatment." *Best Practice and Research: Clinical Haematology* 28 (4): 253–63. https://doi.org/10.1016/j.beha.2015.09.002.

Matutes, Estella, A. Wotherspoon, and D. Catovsky. 2003. "The Variant Form of Hairy-Cell Leukaemia." *Best Practice and Research: Clinical Haematology* 16 (1): 41–56. https://doi.org/10.1016/S1521-6926(02)00086-5.

McCabe, Michael T., Heidi M. Ott, Gopinath Ganji, Susan Korenchuk, Christine Thompson, Glenn S. Van Aller, Yan Liu, et al. 2012. "EZH2 Inhibition as a Therapeutic Strategy for Lymphoma with EZH2-Activating Mutations." *Nature*. https://doi.org/10.1038/nature11606.

McKerrell, Thomas, Thaidy Moreno, Hannes Ponstingl, Niccolo Bolli, Joã M L Dias, German Tischler, Vincenza Colonna, et al. 2016. "E-Blood MYELOID NEOPLASIA Development and Validation of a Comprehensive Genomic Diagnostic Tool for Myeloid Malignancies." https://doi.org/10.1182/blood-2015-11.

Mcneil, Nicole, and Thomas Ried. 2000. "Novel Molecular Cytogenetic Techniques for Identifying Complex Chromosomal Rearrangements: Technology and Applications in Molecular Medicine." http://www-ermm.cbcu.cam.ac.uk.

Menendez, P., A. Vargas, C. Bueno, S. Barena, J. Almeida, M. de Santiago, A. López, S. Roa, J. F. San Miguel, and A. Orfao. 2004. "Quantitative Analysis of Bcl-2 Expression in Normal and Leukemic Human B-Cell Differentiation." *Leukemia* 18 (3): 491–98. https://doi.org/10.1038/sj.leu.2403231.

Metzker, Michael L. 2010. "Sequencing Technologies the next Generation." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2626.

Meyerson, Matthew, Stacey Gabriel, and Gad Getz. 2010. "Advances in Understanding Cancer Genomes through Second-Generation Sequencing." *Nature Reviews Genetics* 11 (10): 685–96. https://doi.org/10.1038/nrg2841.

Meynert, Alison M, Louise S Bicknell, Matthew E Hurles, Andrew P Jackson, and Martin S Taylor. 2013. "Quantifying Single Nucleotide Variant Detection Sensitivity in Exome Sequencing." *BMC Bioinformatics*. Vol. 14. http://www.biomedcentral.com/1471-2105/14/195http://www.biomedcentral.com/1471-2105/14/195.

Milhollen, Michael A., Tary Traore, Jennifer Adams-Duffy, Michael P. Thomas, Allison J. Berger, Lenny Dang, Lawrence R. Dick, et al. 2010. "MLN4924, a NEDD8-Activating Enzyme Inhibitor, Is Active in Diffuse Large B-Cell Lymphoma Models: Rationale for Treatment of NF-KB-Dependent Lymphoma." *Blood* 116 (9): 1515–23. https://doi.org/10.1182/blood-2010-03-272567.

Mohanty, Atish, Natalie Sandoval, Manasi Das, Raju Pillai, Lu Chen, Robert W Chen, Hesham M Amin, et al. n.d. "CCND1 Mutations Increase Protein Stability and Promote Ibrutinib Resistance in Mantle Cell Lymphoma." Vol. 7. www.impactjournals.com/oncotarget.

Momose, S., S. Weißbach, J. Pischimarov, T. Nedeva, E. Bach, M. Rudelius, E. Geissinger, A. M. Staiger, G. Ott, and A. Rosenwald. 2015. "The Diagnostic Gray Zone between Burkitt Lymphoma and Diffuse Large B-Cell Lymphoma Is Also a Gray Zone of the Mutational Spectrum." *Leukemia*. Nature Publishing Group. https://doi.org/10.1038/leu.2015.34.

Monti, Paola, Marta Lionetti, Giuseppa De Luca, Paola Menichini, Anna Grazia Recchia, Serena Matis, Monica Colombo, et al. 2020. "Time to First Treatment and P53 Dysfunction in Chronic Lymphocytic Leukaemia: Results of the O-CLL1 Study in Early Stage Patients." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-75364-3.

Mora, Alba, Rosa Bosch, Carolina Cuellar, Eva Puy Vicente, Laura Blanco, Rodrigo Martino, José M. Ubeda, Jorge Sierra, Carol Moreno, and Josep Nomdedeu. 2019. "CD200 Is a Useful Marker in the Diagnosis of Chronic Lymphocytic Leukemia." *Cytometry Part B - Clinical Cytometry* 96 (2): 143–48. https://doi.org/10.1002/cyto.b.21722.

Morgan, Gareth, Eileen M Boyle, Christopher Wardell, Brian A Walker, Faith E Davies, and Xavier Leleu. 2014. "B-Cell Malignancies: Capture-Sequencing Strategies for Identification of Gene Rearrangements and Translocations into Immunoglobulin Gene Loci." *Blood and Lymphatic Cancer: Targets and Therapy*, October, 107. https://doi.org/10.2147/blctt.s51503.

Morin, Ryan D., Sarit Assouline, Miguel Alcaide, Arezoo Mohajeri, Rebecca L. Johnston, Lauren Chong, Jasleen Grewal, et al. 2016. "Genetic Landscapes of Relapsed and Refractory Diffuse Large B-Cell Lymphomas." *Clinical Cancer Research* 22 (9): 2290–2300. https://doi.org/10.1158/1078-0432.CCR-15-2123.

Morin, Ryan D., Nathalie A. Johnson, Tesa M. Severson, Andrew J. Mungall, Jianghong An, Rodrigo Goya, Jessica E. Paul, et al. 2010. "Somatic Mutations Altering EZH2 (Tyr641) in Follicular and Diffuse Large B-Cell Lymphomas of Germinal-Center Origin." *Nature Genetics* 42 (2): 181–85. https://doi.org/10.1038/ng.518.

Morin, Ryan D., Maria Mendez-Lago, Andrew J. Mungall, Rodrigo Goya, Karen L. Mungall, Richard D. Corbett, Nathalie A. Johnson, et al. 2011. "Frequent Mutation of Histone-Modifying Genes in Non-Hodgkin Lymphoma." *Nature* 476 (7360): 298–303. https://doi.org/10.1038/nature10351.

Morin, Ryan D., Karen Mungall, Erin Pleasance, Andrew J. Mungall, Rodrigo Goya, Ryan D. Huff, David W. Scott, et al. 2013. "Mutational and Structural Analysis of Diffuse Large B-Cell Lymphoma Using Whole-Genome Sequencing." *Blood* 122 (7): 1256–65. https://doi.org/10.1182/blood-2013-02-483727.

Mosca, Laura, Sonia Fabris, Marta Lionetti, Katia Todoerti, Luca Agnelli, Fortunato Morabito, Giovanna Cutrona, et al. 2010. "Integrative Genomics Analyses Reveal Molecularly Distinct Subgroups of B-Cell Chronic Lymphocytic Leukemia Patients with 13q14 Deletion." *Clinical Cancer Research* 16 (23): 5641–53. https://doi.org/10.1158/1078-0432.CCR-10-0151.

Mullis, Kary B. 1990. "The Unusual Origin of the Polymerase Chain Reaction."

Nadeu, Ferran, Julio Delgado, Cristina Royo, Tycho Baumann, Tatjana Stankovic, Magda Pinyol, Pedro Jares, et al. 2016. "Clinical Impact of Clonal and Subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM Mutations in Chronic Lymphocytic Leukemia." *Blood* 127 (17): 2122–30. https://doi.org/10.1182/blood-2015-07.

Naresh, Kikkeri N., Hazem A.H. Ibrahim, Stefano Lazzi, Patricia Rince, Monica Onorati,

Maria R. Ambrosio, Chrystèle Bilhou-Nabera, et al. 2011. "Diagnosis of Burkitt Lymphoma Using an Algorithmic Approach - Applicable in Both Resource-Poor and Resource-Rich Countries." *British Journal of Haematology* 154 (6): 770–76. https://doi.org/10.1111/j.1365-2141.2011.08771.x.

Nedomova, Radka, Tomas Papajik, Vit Prochazka, Karel Indrak, and Marie Jarosova. 2013. "Cytogenetics and Molecular Cytogenetics in Diffuse Large B-Cell Lymphoma (DLBCL)." *Biomedical Papers* 157 (3): 239–47. https://doi.org/10.5507/bp.2012.085.

Ngo, Vu N., Ryan M. Young, Roland Schmitz, Sameer Jhavar, Wenming Xiao, Kian Huat Lim, Holger Kohlhammer, et al. 2011. "Oncogenically Active MYD88 Mutations in Human Lymphoma." *Nature* 470 (7332): 115–21. https://doi.org/10.1038/nature09671.

Nguyen, Lynh, Peter Papenhausen, and Haipeng Shao. 2017. "The Role of C-MYC in B-Cell Lymphomas: Diagnostic and Molecular Aspects." *Genes* 8 (4): 2–22. https://doi.org/10.3390/genes8040116.

Nord, Alex S., Ming Lee, Mary Claire King, and Tom Walsh. 2011. "Accurate and Exact CNV Identification from Targeted High-Throughput Sequence Data." *BMC Genomics* 12 (April). https://doi.org/10.1186/1471-2164-12-184.

O'Rawe, Jason, Tao Jiang, Guangqing Sun, Yiyang Wu, Wei Wang, Jingchu Hu, Paul Bodily, et al. 2013. "Low Concordance of Multiple Variant-Calling Pipelines: Practical Implications for Exome and Genome Sequencing." *Genome Medicine* 5 (3). https://doi.org/10.1186/gm432.

Ohno, Hitoshi. 2006. "Pathogenetic and Clinical Implications of Non-Immunoglobulin ; BCL6 Translocations in B-Cell Non-Hodgkin's Lymphoma." *〖J Clin Exp Hematopathol*. Vol. 46. www.genecards.org/cgi-bin/carddisp?BCL6.

Okosun, Jessica, Csaba Bödör, Jun Wang, Shamzah Araf, Cheng Yuan Yang, Chenyi Pan, Sören Boller, et al. 2014. "Integrated Genomic Analysis Identifies Recurrent Mutations and Evolution Patterns Driving the Initiation and Progression of Follicular Lymphoma." *Nature Genetics* 46 (2): 176–81. https://doi.org/10.1038/ng.2856.

Olszewski, Adam J., and Jorge J. Castillo. 2013. "Survival of Patients with Marginal Zone Lymphoma: Analysis of the Surveillance, Epidemiology, and End Results Database." *Cancer* 119 (3): 629–38. https://doi.org/10.1002/cncr.27773.

Onaindia, Arantza, L. Jeffrey Medeiros, and Keyur P. Patel. 2017. "Clinical Utility of Recently Identified Diagnostic, Prognostic, and Predictive Molecular Biomarkers in Mature B-Cell Neoplasms." *Modern Pathology* 30 (10): 1338–66. https://doi.org/10.1038/modpathol.2017.58.

Ott, German, Tiemo Katzenberger, Andreas Lohr, Steffi Kindelberger, Thomas Rü, Martin Wilhelm, Jö Rg Kalla, et al. 2002. "Cytomorphologic, Immunohistochemical, and Cytogenetic Profiles of Follicular Lymphoma: 2 Types of Follicular Lymphoma Grade 3." www.bloodjournal.org.

Ouillette, Peter, Roxane Collins, Sajid Shakhan, Jinghui Li, Cheng Li, Kerby Shedden, and Sami N. Malek. 2011. "The Prognostic Significance of Various 13q14 Deletions in

Chronic Lymphocytic Leukemia." *Clinical Cancer Research* 17 (21): 6778–90. https://doi.org/10.1158/1078-0432.CCR-11-0785.

Ouillette, Peter, Harry Erba, Lisa Kujawski, Mark Kaminski, Kerby Shedden, and Sami N. Malek. 2008. "Integrated Genomic Profiling of Chronic Lymphocytic Leukemia Identifies Subtypes of Deletion 13q14." *Cancer Research* 68 (4): 1012–21. https://doi.org/10.1158/0008-5472.CAN-07-3105.

Pakneshan, Sahar, Ali Salajegheh, Robert Anthony Smith, and Alfred King Yin Lam. 2013. "Clinicopathological Relevance of BRAF Mutations in Human Cancer." *Pathology* 45 (4): 346–56. https://doi.org/10.1097/PAT.0b013e328360b61d.

Palomo, Laura, Mariam Ibáñez, María Abáigar, Iria Vázquez, Sara Álvarez, Marta Cabezón, Bárbara Tazón-Vega, et al. 2020. "Spanish Guidelines for the Use of Targeted Deep Sequencing in Myelodysplastic Syndromes and Chronic Myelomonocytic Leukaemia." *British Journal of Haematology*. Blackwell Publishing Ltd. https://doi.org/10.1111/bjh.16175.

Panea, Razvan I., Cassandra L. Love, Jennifer R. Shingleton, Anupama Reddy, Jeffrey A. Bailey, Ann M. Moormann, Juliana A. Otieno, et al. 2019. "The Whole-Genome Landscape of Burkitt Lymphoma Subtypes." *Blood* 134 (19): 1598–1607. https://doi.org/10.1182/blood.2019001880.

Parry, Marina, Matthew J.J. Rose-Zerilli, Jane Gibson, Sarah Ennis, Renata Walewska, Jade Forster, Helen Parker, et al. 2013. "Whole Exome Sequencing Identifies Novel Recurrently Mutated Genes in Patients with Splenic Marginal Zone Lymphoma." *PLoS ONE* 8 (12): 1–5. https://doi.org/10.1371/journal.pone.0083244.

Parry, Marina, Matthew J.J. Rose-Zerilli, Viktor Ljungström, Jane Gibson, Jun Wang, Renata Walewska, Helen Parker, et al. 2015. "Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing." *Clinical Cancer Research* 21 (18): 4174–83. https://doi.org/10.1158/1078-0432.CCR-14-2759.

Pasqualucci, Laura. 2013. "The Genetic Basis of Diffuse Large B-Cell Lymphoma." *Current Opinion in Hematology*. https://doi.org/10.1097/MOH.0b013e3283623d7f.

Pasqualucci, Laura, David Dominguez-Sola, Annalisa Chiarenza, Giulia Fabbri, Adina Grunn, Vladimir Trifonov, Lawryn H. Kasper, et al. 2011. "Inactivating Mutations of Acetyltransferase Genes in B-Cell Lymphoma." *Nature* 471 (7337): 189–96. https://doi.org/10.1038/nature09730.

Pasqualucci, Laura, Hossein Khiabanian, Marco Fangazio, Mansi Vasishtha, Monica Messina, Antony B. Holmes, Peter Ouillette, et al. 2014. "Genetics of Follicular Lymphoma Transformation." *Cell Reports* 6 (1): 130–40. https://doi.org/10.1016/j.celrep.2013.12.027.

Pastore, Alessandro, Vindi Jurinovic, Robert Kridel, Eva Hoster, Annette M. Staiger, Monika Szczepanowski, Christiane Pott, et al. 2015. "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." *The Lancet Oncology* 16 (9): 1111–22.

https://doi.org/10.1016/S1470-2045(15)00169-2.

Pé Rez-Galán, Patricia, Martin Dreyling, and Adrian Wiestner. n.d. "Mantle Cell Lymphoma: Biology, Pathogenesis, and the Molecular Basis of Treatment in the Genomic Era." https://doi.org/10.1182/blood-2010-04.

Perry, Anamarija M., Teresa M. Cardesa-Salzmann, Paul N. Meyer, Luis Colomo, Lynette M. Smith, Kai Fu, Timothy C. Greiner, et al. 2012. "A New Biologic Prognostic Model Based on Immunohistochemistry Predicts Survival in Patients with Diffuse Large B-Cell Lymphoma." *Blood* 120 (11): 2290–96. https://doi.org/10.1182/blood-2012-05-430389.

Peterson, Jess F., Beth A. Pitel, Stephanie A. Smoley, George Vasmatzis, James B. Smadbeck, Patricia T. Greipp, Rhett P. Ketterling, William R. Macon, and Linda B. Baughn. 2019. "Elucidating a False-Negative MYC Break-Apart Fluorescence in Situ Hybridization Probe Study by next-Generation Sequencing in a Patient with High-Grade B-Cell Lymphoma with IGH/MYC and IGH/BCL2 Rearrangements." *Cold Spring Harbor Molecular Case Studies* 5 (3). https://doi.org/10.1101/mcs.a004077.

Peveling-Oberhag, Jan, Franziska Wolters, Claudia Döring, Dirk Walter, Ludger Sellmann, René Scholtysik, Marco Lucioni, et al. 2015. "Whole Exome Sequencing of Microdissected Splenic Marginal Zone Lymphoma: A Study to Discover Novel Tumor-Specific Mutations." *BMC Cancer* 15 (1): 1–10. https://doi.org/10.1186/s12885-015-1766-z.

Pinzaru, Alexandra M., Robert A. Hom, Angela Beal, Aaron F. Phillips, Eric Ni, Timothy Cardozo, Nidhi Nair, et al. 2016. "Telomere Replication Stress Induced by POT1 Inactivation Accelerates Tumorigenesis." *Cell Reports* 15 (10): 2170–84. https://doi.org/10.1016/j.celrep.2016.05.008.

Pittaluga~, Stefania, Torik A Y Ayoubi, Iwona Wlodarska, Michel Stul, Jean-Jacques Cassiman, Cristina Mecucci, Herman Van Den Berghe, Wim J M Van De, and Chris De Wolf-Peeters~. 1996. "BCL-6 EXPRESSION IN REACTIVE LYMPHOID TISSUE AND IN B-CELL NON-HODGKIN'S LYMPHOMAS*." *JOURNAL OF PATHOLOGY*. Vol. 179.

Piva, R., S. Deaglio, R. Famà, R. Buonincontri, I. Scarfò, A. Bruscaggin, E. Mereu, et al. 2015. "The Krüppel-like Factor 2 Transcription Factor Gene Is Recurrently Mutated in Splenic Marginal Zone Lymphoma." *Leukemia*. Nature Publishing Group. https://doi.org/10.1038/leu.2014.294.

Pospisilova, S., D. Gonzalez, J. Malcikova, M. Trbusek, D. Rossi, A. P. Kater, F. Cymbalista, et al. 2012. "ERIC Recommendations on TP53 Mutation Analysis in Chronic Lymphocytic Leukemia." *Leukemia*. https://doi.org/10.1038/leu.2012.25.

Potapov, Vladimir, and Jennifer L. Ong. 2017. "Examining Sources of Error in PCR by Single-Molecule Sequencing." *PLoS ONE* 12 (1): 1–19. https://doi.org/10.1371/journal.pone.0169774.

Prakash, Gaurav, Anupriya Kaur, Pankaj Malhotra, Alka Khadwal, Prashant Sharma, Vikas Suri, Neelam Varma, and Subhash Varma. 2016. "Current Role of Genetics in Hematologic Malignancies." *Indian Journal of Hematology and Blood Transfusion*. Springer India. https://doi.org/10.1007/s12288-015-0584-4.

Prieto-Conde, M. Isabel, Luis A. Corchete, María García-Álvarez, Cristina Jiménez, Alejandro Medina, Ana Balanzategui, Montserrat Hernández-Ruano, et al. 2020. "A New Next-Generation Sequencing Strategy for the Simultaneous Analysis of Mutations and Chromosomal Rearrangements at DNA Level in Acute Myeloid Leukemia Patients." *Journal of Molecular Diagnostics* 22 (1): 60–71. https://doi.org/10.1016/j.jmoldx.2019.08.002.

Puente, Xose S., Silvia Beà, Rafael Valdés-Mas, Neus Villamor, Jesús Gutiérrez-Abril, José I. Martín-Subero, Marta Munar, et al. 2015. "Non-Coding Recurrent Mutations in Chronic Lymphocytic Leukaemia." *Nature* 526 (7574): 519–24. https://doi.org/10.1038/nature14666.

Puente, Xose S, Pedro Jares, and Elias Campo. 2018. "Review Series Chronic Lymphocytic Leukemia and Mantle Cell Lymphoma: Crossroads of Genetic and Microenvironment Interactions." https://ashpublications.org/blood/article-pdf/131/21/2283/1468260/blood764373.pdf.

Pugh, Trevor J., Sami S. Amr, Mark J. Bowser, Sivakumar Gowrisankar, Elizabeth Hynes, Lisa M. Mahanta, Heidi L. Rehm, Birgit Funke, and Matthew S. Lebo. 2016. "VisCap: Inference and Visualization of Germ-Line Copy-Number Variants from Targeted Clinical Sequencing Data." *Genetics in Medicine* 18 (7): 712–19. https://doi.org/10.1038/gim.2015.156.

Puiggros, Anna, Gonzalo Blanco, and Blanca Espinet. 2014. "Genetic Abnormalities in Chronic Lymphocytic Leukemia: Where We Are and Where We Go." *BioMed Research International* 2014. https://doi.org/10.1155/2014/435983.

Puiggros, Anna, Rosa Collado, Maria José Calasanz, Margarita Ortega, Neus Ruiz-Xivillé, Alfredo Rivas-Delgado, Elisa Luño, et al. 2017. "Patients with Chronic Lymphocytic Leukemia and Complex Karyotype Show an Adverse Outcome Even in Absence of TP53/ATM FISH Deletions." *Oncotarget* 8 (33): 54297–303. https://doi.org/10.18632/oncotarget.17350.

Puiggros, Anna, Julio Delgado, Ana Rodriguez-Vicente, Rosa Collado, Anna Aventín, Elisa Luño, Javier Grau, et al. 2013. "Biallelic Losses of 13q Do Not Confer a Poorer Outcome in Chronic Lymphocytic Leukaemia: Analysis of 627 Patients with Isolated 13q Deletion." *British Journal of Haematology* 163 (1): 47–54. https://doi.org/10.1111/bjh.12479.

Quesada, Víctor, Laura Conde, Neus Villamor, Gonzalo R. Ordóñez, Pedro Jares, Laia Bassaganyas, Andrew J. Ramsay, et al. 2012. "Exome Sequencing Identifies Recurrent Mutations of the Splicing Factor SF3B1 Gene in Chronic Lymphocytic Leukemia." *Nature Genetics* 44 (1): 47–52. https://doi.org/10.1038/ng.1032.

Rack, K. A., E. van den Berg, C. Haferlach, H. B. Beverloo, D. Costa, B. Espinet, N. Foot, et al. 2019. "European Recommendations and Quality Assurance for Cytogenomic Analysis of Haematological Neoplasms." *Leukemia*. Nature Publishing Group. https://doi.org/10.1038/s41375-019-0378-z.

Rahal, Rami, Mareike Frick, Rodrigo Romero, Joshua M. Korn, Robert Kridel, Fong Chun Chan, Barbara Meissner, et al. 2014. "Pharmacological and Genomic Profiling

Identifies NF-KB-Targeted Treatment Strategies for Mantle Cell Lymphoma." *Nature Medicine* 20 (1): 87–92. https://doi.org/10.1038/nm.3435.

Ramsay, Andrew J., Víctor Quesada, Miguel Foronda, Laura Conde, Alejandra Martínez-Trillos, Neus Villamor, David Rodríguez, et al. 2013. "POT1 Mutations Cause Telomere Dysfunction in Chronic Lymphocytic Leukemia." *Nature Genetics* 45 (5): 526–30. https://doi.org/10.1038/ng.2584.

Rausch, Tobias, Thomas Zichner, Andreas Schlattl, Adrian M. Stütz, Vladimir Benes, and Jan O. Korbel. 2012. "DELLY: Structural Variant Discovery by Integrated Paired-End and Split-Read Analysis." *Bioinformatics* 28 (18): 333–39. https://doi.org/10.1093/bioinformatics/bts378.

Reddy, Anupama, Jenny Zhang, Nicholas S. Davis, Andrea B. Moffitt, Cassandra L. Love, Alexander Waldrop, Sirpa Leppa, et al. 2017. "Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma." *Cell* 171 (2): 481-494.e15. https://doi.org/10.1016/j.cell.2017.09.027.

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. 2015. "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." *Genetics in Medicine* 17 (5): 405–24. https://doi.org/10.1038/gim.2015.30.

Richter, Julia, Matthias Schlesner, Steve Hoffmann, Markus Kreuz, Ellen Leich, Birgit Burkhardt, MacIej Rosolowski, et al. 2012. "Recurrent Mutation of the ID3 Gene in Burkitt Lymphoma Identified by Integrated Genome, Exome and Transcriptome Sequencing." *Nature Genetics* 44 (12): 1316–20. https://doi.org/10.1038/ng.2469.

Rickert, Robert C. 2013. "New Insights into Pre-BCR and BCR Signalling with Relevance to B Cell Malignancies." *Nature Reviews Immunology*. https://doi.org/10.1038/nri3487.

Rigby, P, M Dieckmanen, C Rhodes, and P Berg. 1977. "Labeling DNA to High Specific Activity in Vitro by Nick Translation and DNA Polymerase I ." *J. Mol. Biol.* 113: 237–45.

Robinson, Jacob E, and Christine E Cutucache. n.d. "Deciphering Splenic Marginal Zone Lymphoma Pathogenesis: The Proposed Role of MicroRNA." www.oncotarget.com.

Robledo, Cristina, Juan L. García, Rocío Benito, Teresa Flores, Manuela Mollejo, José Ángel Martínez-Climent, Eva García, Norma C. Gutiérrez, Miguel A. Piris, and Jesús M. Hernández. 2011. "Molecular Characterization of the Region 7q22.1 in Splenic Marginal Zone Lymphomas." *PLoS ONE* 6 (9). https://doi.org/10.1371/journal.pone.0024939.

Rodríguez-Vicente, Ana E, Vasilis Bikos, María Hernández-Sánchez, Jitka Malcikova, Jesús-María Hernández-Rivas, and Sarka Pospisilova. 2017. "Next-Generation Sequencing in Chronic Lymphocytic Leukemia: Recent Findings and New Horizons." *Oncotarget*. Vol. 8. www.impactjournals.com/oncotarget/.

Rohde, Marius, Bettina R. Bonn, Martin Zimmermann, Jonas Lange, Anja Möricke,

Wolfram Klapper, Ilske Oschlies, et al. 2017. "Relevance of ID3-TCF3-CCND3 Pathway Mutations in Pediatric Aggressive B-Cell Lymphoma Treated According to the Non-Hodgkin Lymphoma Berlin-Frankfurt-Münster Protocols." *Haematologica* 102 (6): 1091–98. https://doi.org/10.3324/haematol.2016.156885.

Rosati, Emanuela, Stefano Baldoni, Filomena De Falco, Beatrice Del Papa, Erica Dorillo, Chiara Rompietti, Elisa Albi, Franca Falzetti, Mauro Di Ianni, and Paolo Sportoletti. 2018. "NOTCH1 Aberrations in Chronic Lymphocytic Leukemia." *Frontiers in Oncology*. Frontiers Media S.A. https://doi.org/10.3389/fonc.2018.00229.

Rosenquist, Richard, Andreas Rosenwald, Ming Qing Du, Gianluca Gaidano, Patricia Groenen, Andrew Wotherspoon, Paolo Ghia, Philippe Gaulard, Elias Campo, and Kostas Stamatopoulos. 2016. "Clinical Impact of Recurrently Mutated Genes on Lymphoma Diagnostics: State-of-the-Art and Beyond." *Haematologica*. Ferrata Storti Foundation. https://doi.org/10.3324/haematol.2015.134510.

Rosenthal, Allison, and Anas Younes. 2017. "High Grade B-Cell Lymphoma with Rearrangements of MYC and BCL2 and/or BCL6: Double Hit and Triple Hit Lymphomas and Double Expressing Lymphoma." *Blood Reviews*. Churchill Livingstone. https://doi.org/10.1016/j.blre.2016.09.004.

Rosenwald, Andreas, Susanne Bens, Ranjana Advani, Sharon Barrans, Christiane Copie-Bergman, Mad-Helenie Elsensohn, Yaso Natkunam, et al. 2019. "Prognostic Significance of MYC Rearrangement and Translocation Partner in Diffuse Large B-Cell Lymphoma: A Study by the Lunenburg Lymphoma Biomarker Consortium." *J Clin Oncol*. Vol. 37. https://doi.org/10.

Ross, Michael G., Carsten Russ, Maura Costello, Andrew Hollinger, Niall J. Lennon, Ryan Hegarty, Chad Nusbaum, and David B. Jaffe. 2013. "Characterizing and Measuring Bias in Sequence Data." *Genome Biology* 14 (5). https://doi.org/10.1186/gb-2013-14-5-r51.

Rossi, Davide, Alessio Bruscaggin, Valeria Spina, Silvia Rasi, Hossein Khiabanian, Monica Messina, Marco Fangazio, et al. 2011. "Mutations of the SF3B1 Splicing Factor in Chronic Lymphocytic Leukemia: Association with Progression and Fludarabine-Refractoriness." *Blood* 118 (26): 6904–8. https://doi.org/10.1182/blood-2011-08-373159.

Rossi, Davide, Hossein Khiabanian, Valeria Spina, Carmela Ciardullo, Alessio Bruscaggin, Rosella Famà, Silvia Rasi, et al. 2014. "Clinical Impact of Small TP53 Mutated Subclones in Chronic Lymphocytic Leukemia." *Blood* 123 (14): 2139–47. https://doi.org/10.1182/blood-2013-11-539726.

Rossi, Davide, Silvia Rasi, Valeria Spina, Alessio Bruscaggin, Sara Monti, Carmela Ciardullo, Clara Deambrogi, et al. 2013. "Integrated Mutational and Cytogenetic Analysis Identifies New Prognostic Subgroups in Chronic Lymphocytic Leukemia." *Blood* 121 (8): 1403–12. https://doi.org/10.1182/blood-2012-09-458265.

Rossi, Davide, Vladimir Trifonov, Marco Fangazio, Alessio Bruscaggin, Silvia Rasi, Valeria Spina, Sara Monti, et al. 2012. "The Coding Genome of Splenic Marginal Zone Lymphoma: Activation of NOTCH2 and Other Pathways Regulating Marginal Zone

Development." *Journal of Experimental Medicine* 209 (9): 1537–51. https://doi.org/10.1084/jem.20120904.

Rovira, Jordina, Kennosuke Karube, Alexandra Valera, Dolors Colomer, Anna Enjuanes, Lluís Colomo, Alejandra Martínez-Trillos, et al. 2016. "MYD88 L265P Mutations, but No Other Variants, Identify a Subpopulation of DLBCL Patients of Activated B-Cell Origin, Extranodal Involvement, and Poor Outcome." *Clinical Cancer Research* 22 (11): 2755–64. https://doi.org/10.1158/1078-0432.CCR-15-1525.

Ruan, Jia, Peter Martin, Bijal Shah, Stephen J. Schuster, Sonali M. Smith, Richard R. Furman, Paul Christos, et al. 2015. "Lenalidomide plus Rituximab as Initial Treatment for Mantle-Cell Lymphoma." *New England Journal of Medicine* 373 (19): 1835–44. https://doi.org/10.1056/nejmoa1505237.

Rudolph, C., T. Liehr, D. Steinemann, M. Emura, M. Daibata, Y. Matsuo, N. Emi, et al. 2006. "Different Breakage-Prone Regions on Chromosome 1 Detected in t(11;14)-Positive Mantle Cell Lymphoma Cell Lines and Multiple Myeloma Cell Lines Are Associated with Different Tumor Progession-Related Mechanisms." *Cytogenetic and Genome Research* 112 (3–4): 213–21. https://doi.org/10.1159/000089873.

Ryan, Russell J.H., Yotam Drier, Holly Whitton, M. Joel Cotton, Jasleen Kaur, Robbyn Issner, Shawn Gillespie, et al. 2015. "Detection of Enhancer-Associated Rearrangements Reveals Mechanisms of Oncogene Dysregulation in B-Cell Lymphoma." *Cancer Discovery* 5 (10): 1058–71. https://doi.org/10.1158/2159-8290.CD-15-0370.

Salaverria, Itziar, Blanca Espinet, Ana Carrió, Dolors Costa, Laura Astier, Julia Slotta-Huspenina, Leticia Quintanilla-Martinez, et al. 2008. "Multiple Recurrent Chromosomal Breakpoints in Mantle Cell Lymphoma Revealed by a Combination of Molecular Cytogenetic Techniques." *Genes Chromosomes and Cancer* 47 (12): 1086–97. https://doi.org/10.1002/gcc.20609.

Salaverria, Itziar, David Martín-Garcia, Cristina López, Guillem Clot, Manel García-Aragonés, Alba Navarro, Julio Delgado, et al. 2015. "Detection of Chromothripsis-like Patterns with a Custom Array Platform for Chronic Lymphocytic Leukemia." *Genes Chromosomes and Cancer* 54 (11): 668–80. https://doi.org/10.1002/gcc.22277.

Salaverria, Itziar, Idoia Martin-Guerrero, Rabea Wagener, Markus Kreuz, Christian W Kohler, Julia Richter, Barbara Pienkowska-Grela, et al. 2014. "A Recurrent 11q Aberration Pattern Characterizes a Subset of MYC-Negative High-Grade B-Cell Lymphomas Resembling Burkitt Lymphoma." *Blood* 123 (8): 1187–98. https://doi.org/10.1182/blood-2013-06.

Salaverria, Itziar, Cristina Royo, Alejandra Carvajal-Cuenca, Guillem Clot, Alba Navarro, Alejandra Valera, Joo Y Song, et al. n.d. "CCND2 Rearrangements Are the Most Frequent Genetic Events in Cyclin D1 Mantle Cell Lymphoma." https://doi.org/10.1182/blood.

Salaverria, Itziar, Andreas Zettl, Silvia Beà, Elena M. Hartmann, Sandeep S. Dave, George W. Wright, Evert Jan Boerma, et al. 2008. "Chromosomal Alterations Detected by

Comparative Genomic Hybridization in Subgroups of Gene Expression-Defined Burkitt's Lymphoma." *Haematologica* 93 (9): 1327–34. https://doi.org/10.3324/haematol.13071.

Salavoura, Katerina, Aggeliki Kolialexi, George Tsangaris, and Ariadni Mavrou. 2008. "Development of Cancer in Patients with Primary Immunodeficiencies." *Anticancer Research* 28 (2 B): 1263–69.

Salido, Marta, Cristina Baró, David Oscier, Kostas Stamatopoulos, Judith Dierlamm, Estela Matutes, Alexandra Traverse-Glehen, et al. 2010. "Cytogenetic Aberrations and Their Prognostic Value in a Series of 330 Splenic Marginal Zone B-Cell Lymphomas: A Multicenter Study of the Splenic B-Cell Lymphoma Group." *Blood* 116 (9): 1479–88. https://doi.org/10.1182/blood-2010-02-267476.

Samorodnitsky, Eric, Benjamin M. Jewell, Raffi Hagopian, Jharna Miya, Michele R. Wing, Ezra Lyon, Senthilkumar Damodaran, et al. 2015. "Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing." *Human Mutation* 36 (9): 903–14. https://doi.org/10.1002/humu.22825.

Samuels, David C., Leng Han, Jiang Li, Sheng Quanghu, Travis A. Clark, Yu Shyr, and Yan Guo. 2013. "Finding the Lost Treasures in Exome Sequencing Data." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2013.07.006.

Sander, Sandrine, Dinis P. Calado, Lakshmi Srinivasan, Karl Köchert, Baochun Zhang, Maciej Rosolowski, Scott J. Rodig, et al. 2012. "Synergy between PI3K Signaling and MYC in Burkitt Lymphomagenesis." *Cancer Cell* 22 (2): 167–79. https://doi.org/10.1016/j.ccr.2012.06.012.

Sanger, F, S Nicklen, and A R Coulson. 1977. "DNA Sequencing with Chain-Terminating Inhibitors (DNA Polymerase/Nucleotide Sequences/Bacteriophage 4X174)." Vol. 74.

Schmitt, Adam M., and Howard Y. Chang. 2016. "Long Noncoding RNAs in Cancer Pathways." *Cancer Cell*. Cell Press. https://doi.org/10.1016/j.ccell.2016.03.010.

Schmitz, Roland, George W. Wright, Da Wei Huang, Calvin A. Johnson, James D. Phelan, James Q. Wang, Sandrine Roulland, et al. 2018. "Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma." *New England Journal of Medicine* 378 (15): 1396–1407. https://doi.org/10.1056/nejmoa1801445.

Schmitz, Roland, Ryan M. Young, Michele Ceribelli, Sameer Jhavar, Wenming Xiao, Meili Zhang, George Wright, et al. 2012. "Burkitt Lymphoma Pathogenesis and Therapeutic Targets from Structural and Functional Genomics." *Nature* 490 (7418): 116–20. https://doi.org/10.1038/nature11378.

Schnaiter, Andrea, Peter Paschka, Marianna Rossi, Thorsten Zenz, Andreas Bühler, Dirk Winkler, Mario Cazzola, et al. 2013. "NOTCH1, SF3B1, and TP53 Mutations in Fludarabine-Refractory CLL Patients Treated with Alemtuzumab: Results from the CLL2H Trial of the GCLLSG." *Blood* 122 (7): 1266–70. https://doi.org/10.1182/blood-2013-03-488197.

Schoenbrunner, Nancy J., Amar P. Gupta, Karen K.Y. Young, and Stephen G. Will. 2017. "Covalent Modification of Primers Improves PCR Amplification Specificity and

Yield." *Biology Methods and Protocols* 2 (1). https://doi.org/10.1093/biomethods/bpx011.

Scholtysik, René, Inga Nagel, Markus Kreuz, Inga Vater, Maciej Giefing, Carsten Schwaenen, Swen Wessendorf, et al. 2012. "Recurrent Deletions of the TNFSF7 and TNFSF9 Genes in 19p13.3 in Diffuse Large B-Cell and Burkitt Lymphomas." *International Journal of Cancer* 131 (5). https://doi.org/10.1002/ijc.27416.

Schoumans, Jacqueline, Javier Suela, Ros Hastings, Dominique Muehlematter, Katrina Rack, Eva van den Berg, H. Berna Beverloo, and Marian Stevens-Kroef. 2016. "Guidelines for Genomic Array Analysis in Acquired Haematological Neoplastic Disorders." *Genes Chromosomes and Cancer* 55 (5): 480–91. https://doi.org/10.1002/gcc.22350.

Schraders, Margit, Daphne de Jong, Philip Kluin, Patricia Groenen, and Han van Krieken. 2005. "Lack of Bcl-2 Expression in Follicular Lymphoma May Be Caused by Mutations in the BCL2 Gene or by Absence of the t(14;18) Translocation." *Journal of Pathology* 205 (3): 329–35. https://doi.org/10.1002/path.1689.

Schüler, Frank, Lars Dölken, Carsten Hirt, Thomas Kiefer, Tobias Berg, Gerhard Fusch, K. Weitmann, et al. 2009. "Prevalence and Frequency of Circulating (14;18)-MBE Translocation Carrying Cells in Healthy Individuals." *International Journal of Cancer* 124 (4): 958–63. https://doi.org/10.1002/ijc.23958.

Seité, Paule, Dominique Leroux, Josette Hillion, Michèle Monteil, Roland Berger, Danièle Mathieu-Mahul, and Christian-Jacques -J Larsen. 1993. "Molecular Analysis of a Variant 18;22 Translocation in a Case of Lymphocytic Lymphoma." *Genes, Chromosomes and Cancer* 6 (1): 39–44. https://doi.org/10.1002/gcc.2870060108.

Seitz, Volkhard, Peter Butzhammer, Burkhard Hirsch, Jochen Hecht, Ines Gütgemann, Anke Ehlers, Dido Lenze, et al. 2011. "Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma." *PLoS ONE* 6 (11). https://doi.org/10.1371/journal.pone.0026837.

Setoodeh, Reza, Stuart Schwartz, Peter Papenhausen, Ling Zhang, Elizabeth M Sagatys, Lynn C Moscinski, and Haipeng Shao. 2013. "Double-Hit Mantle Cell Lymphoma with MYC Gene Rearrangement or Amplification: A Report of Four Cases and Review of the Literature." *Int J Clin Exp Pathol*. Vol. 6. www.ijcep.com/.

Shin, Sunguk, and Joonhong Park. 2016. "Characterization of Sequence-Specific Errors in Various next-Generation Sequencing Systems." *Molecular BioSystems* 12 (3): 914–22. https://doi.org/10.1039/c5mb00750j.

Shirai, Cara Lunn, Brian S. White, Manorama Tripathi, Roberto Tapia, James N. Ley, Matthew Ndonwi, Sanghyun Kim, et al. 2017. "Mutant U2AF1-Expressing Cells Are Sensitive to Pharmacological Modulation of the Spliceosome." *Nature Communications* 8. https://doi.org/10.1038/ncomms14060.

Sikkema-Raddatz, Birgit, Lennart F. Johansson, Eddy N. de Boer, Rowida Almomani, Ludolf G. Boven, Maarten P. van den Berg, Karin Y. van Spaendonck-Zwarts, et al. 2013. "Targeted Next-Generation Sequencing Can Replace Sanger Sequencing in Clinical Diagnostics." *Human Mutation* 34 (7): 1035–42.

https://doi.org/10.1002/humu.22332.

Simons, Annet, Birgit Sikkema-Raddatz, Nicole de Leeuw, Nicole Claudia Konrad, Rosalind J. Hastings, and Jacqueline Schoumans. 2012. "Genome-Wide Arrays in Routine Diagnostics of Hematological Malignancies." *Human Mutation*. https://doi.org/10.1002/humu.22057.

Sims, David, Ian Sudbery, Nicholas E. Ilott, Andreas Heger, and Chris P. Ponting. 2014. "Sequencing Depth and Coverage: Key Considerations in Genomic Analyses." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3642.

Skowronska, Anna, Anton Parker, Gulshanara Ahmed, Ceri Oldreive, Zadie Davis, Sue Richards, Martin Dyer, et al. 2012. "Biallelic ATM Inactivation Significantly Reduces Survival in Patients Treated on the United Kingdom Leukemia Research Fund Chronic Lymphocytic Leukemia 4 Trial." *Journal of Clinical Oncology* 30 (36): 4524–32. https://doi.org/10.1200/JCO.2011.41.0852.

Smith, Cindy J., and A. Mark Osborn. 2009. "Advantages and Limitations of Quantitative PCR (Q-PCR)-Based Approaches in Microbial Ecology." *FEMS Microbiology Ecology*. https://doi.org/10.1111/j.1574-6941.2008.00629.x.

Speicher, Michael R., and Nigel P. Carter. 2005. "The New Cytogenetics: Blurring the Boundaries with Molecular Biology." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg1692.

Spina, Valeria, and Davide Rossi. 2017. "Molecular Pathogenesis of Splenic and Nodal Marginal Zone Lymphoma." *Best Practice and Research: Clinical Haematology* 30 (1–2): 5–12. https://doi.org/10.1016/j.beha.2016.09.004.

Stengel, Anna, Niroshan Nadarajah, Torsten Haferlach, Frank Dicker, Wolfgang Kern, Manja Meggendorfer, and Claudia Haferlach. 2018. "Detection of Recurrent and of Novel Fusion Transcripts in Myeloid Malignancies by Targeted RNA Sequencing." *Leukemia* 32 (5): 1229–38. https://doi.org/10.1038/s41375-017-0002-z.

Stilgenbauer, Stephan, Andrea Schnaiter, Peter Paschka, Thorsten Zenz, Marianna Rossi, Matthias Ritgen, Michael Kneba, et al. 2014. "Gene Mutations and Treatment Outcome in Chronic Lymphocytic Leukemia: Results from the CLL8 Trial." https://doi.org/10.1182/blood-2014-01.

Straus, Stephen E., Elaine S. Jaffe, Jennifer M. Puck, Janet K. Dale, Keith B. Elkon, Angela Rösen-Wolff, Anke M.J. Peters, et al. 2001. "The Development of Lymphomas in Families with Autoimmune Lymphoproliferative Syndrome with Germline Fas Mutations and Defective Lymphocyte Apoptosis." *Blood* 98 (1): 194–200. https://doi.org/10.1182/blood.V98.1.194.

Streubel, B., I. Simonitsch-Klupp, L. Müllauer, A. Lamprecht, D. Huber, R. Siebert, M. Stolte, et al. 2004. "Variable Frequencies of MALT Lymphoma-Associated Genetic Aberrations in MALT Lymphomas of Different Sites." *Leukemia* 18 (10): 1722–26. https://doi.org/10.1038/sj.leu.2403501.

Strom, Samuel P. 2016. "Current Practices and Guidelines for Clinical Next-Generation Sequencing Oncology Testing." *Cancer Biology and Medicine*. Cancer Biology and

Medicine. https://doi.org/10.28092/j.issn.2095-3941.2016.0004.

Sujobert, Pierre, Yannick Le Bris, Laurence De Leval, Audrey Gros, Jean Philippe Merlio, Cedric Pastoret, Sarah Huet, et al. 2019. "The Need for a Consensus Next-Generation Sequencing Panel for Mature Lymphoid Malignancies." *HemaSphere*. Wolters Kluwer Health. https://doi.org/10.1097/HS9.0000000000000169.

Sutton, Lesley Ann, Viktor Ljungström, Larry Mansouri, Emma Young, Diego Cortese, Veronika Navrkalova, Jitka Malcikova, et al. 2015. "Targeted Next-Generation Sequencing in Chronic Lymphocytic Leukemia: A High-Throughput yet Tailored Approach Will Facilitate Implementation in a Clinical Setting." *Haematologica* 100 (3): 370–76. https://doi.org/10.3324/haematol.2014.109777.

Swenson, Samantha. n.d. "E3 Ligase, UBR5, HECT Domain Mutations in Lymphoma Control Maturation of B Cells." https://doi.org/10.1182/blood.2019002102/1725064/blood.2019002102.pdf.

Swerdlow, Steven H., Elias Campo, Stefano A. Pileri, Nancy Lee Harris, Harald Stein, Reiner Siebert, Ranjana Advani, et al. 2016. *The 2016 Revision of the World Health Organization Classification of Lymphoid Neoplasms*. *Blood*. Vol. 127. https://doi.org/10.1182/blood-2016-01-643569.

Swerdlow, Steven H. n.d. "Diagnosis of 'double Hit' Diffuse Large B-Cell Lymphoma and B-Cell Lymphoma, Unclassifiable, with Features Intermediate between DLBCL and Burkitt Lymphoma: When and How, FISH versus IHC." https://ashpublications.org/hematology/article-pdf/2014/1/90/1251400/bep00114000090.pdf.

Taji, Sawako, Kenichi Nomura, Yosuke Matsumoto, Hideaki Sakabe, Naohisa Yoshida, Shoji Mitsufuji, Kazuhiro Nishida, et al. 2005. "Trisomy 3 May Predict a Poor Response of Gastric MALT Lymphoma to Helicobacter Pylori Eradication Therapy." *World Journal of Gastroenterology* 11 (1). http://www.wjgnet.com/1007-9327/11/89.asp.

Tapia, Gustavo, Raquel Lopez, Ana M. Muñoz-Mármol, José L. Mate, Carolina Sanz, Ruth Marginet, José Tomás Navarro, Josep Maria Ribera, and Aurelio Ariza. 2011. "Immunohistochemical Detection of MYC Protein Correlates with MYC Gene Status in Aggressive B Cell Lymphomas." *Histopathology* 59 (4): 672–78. https://doi.org/10.1111/j.1365-2559.2011.03978.x.

Thompson, Philip A., Susan M. O'Brien, William G. Wierda, Alessandra Ferrajoli, Francesco Stingo, Susan C. Smith, Jan A. Burger, et al. 2015. "Complex Karyotype Is a Stronger Predictor than Del(17p) for an Inferior Outcome in Relapsed or Refractory Chronic Lymphocytic Leukemia Patients Treated with Ibrutinib-Based Regimens." *Cancer* 121 (20): 3612–21. https://doi.org/10.1002/cncr.29566.

Tiacci, Enrico, Vladimir Trifonov, D Ph, Gianluca Schiavoni, D Ph, Antony Holmes, D Ph, et al. 2013. "NIH Public Access" 364 (24): 2305–15. https://doi.org/10.1056/NEJMoa1014209.BRAF.

Traverse-Glehen, A., L. Baseggio, G. Salles, P. Felman, and F. Berger. 2011. "Splenic Marginal Zone B-Cell Lymphoma: A Distinct Clinicopathological and Molecular

Entity. Recent Advances in Ontogeny and Classification." *Current Opinion in Oncology* 23 (5): 441–48. https://doi.org/10.1097/CCO.0b013e328349ab8d.

Treangen, Todd J., and Steven L. Salzberg. 2012. "Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3117.

Trinh, Diane L, David W Scott, Ryan D Morin, Maria Mendez-Lago, Jianghong An, Steven J M Jones, Andrew J Mungall, et al. 2013. "Analysis of FOXO1 Mutations in Diffuse Large B-Cell Lymphoma." *Blood* 121 (18): 3666–74. https://doi.org/10.1182/blood-2013.

Troppan, Katharina, Kerstin Wenzl, Peter Neumeister, and Alexander Deutsch. 2015. "Molecular Pathogenesis of MALT Lymphoma." *Gastroenterology Research and Practice*. Hindawi Publishing Corporation. https://doi.org/10.1155/2015/102656.

Tschernitz, Sebastian, Lucia Flossbach, Margrit Bonengel, Sabine Roth, Andreas Rosenwald, and Eva Geissinger. 2014. "Alternative BRAF Mutations in BRAF V600E-Negative Hairy Cell Leukaemias." *British Journal of Haematology* 165 (4): 529–33. https://doi.org/10.1111/bjh.12735.

Tsuchiya, Dai, Aya Matsumoto, Sarah F Covert, Charlotte R Bronson, and Masatoki Taga. n.d. "Physical Mapping of Plasmid and Cosmid Clones in Filamentous Fungi by Fiber-FISH." www.academicpress.com.

Tsujimoto, Y; Cossman, J; Jaffe, E; Croce, C. (1985). *Involvement of the bcl-2 gene in human follicular lymphoma. , 228(4706), 1440–1443.* doi:10.1126/science.3874430 Ueda, Chiyoko, Takashi Uchiyama, and Hitoshi Ohno. 2002. "Immunoglobulin (Ig)/BCL6 versus Non-Ig/BCL6 Gene Fusion in Diffuse Large B-Cell Lymphoma Corresponds to a High-versus Low-Level Expression of BCL6 MRNA." *2624 BLOOD*. Vol. 99. https://ashpublications.org/blood/article-pdf/99/7/2624/1682354/2624.pdf.

Tsujimoto, Y; Gorham, J; Cossman, J; Jaffe, E; Croce, C. (1985). *The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. Science, 229(4720), 1390–1393.* doi:10.1126/science.3929382

Tsujimoto, Y.; Jaffe, E.; Cossman, J.; Gorham, J.; Nowell, P. C.; Croce, C. M. (1985). *Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11 ; 14) chromosome translocation. , 315(6017), 340–343.* doi:10.1038/315340a0

Ulahannan, D., M. B. Kovac, P. J. Mulholland, J. B. Cazier, and I. Tomlinson. 2013. "Technical and Implementation Issues in Using Next-Generation Sequencing of Cancers in Clinical Practice." *British Journal of Cancer* 109 (4): 827–35. https://doi.org/10.1038/bjc.2013.416.

Vaandrager, Jan Willem, Ed Schuuring, Ton Raap, Katja Philippo, Karin Kleiverda, and Philip Kluin. 2000. "Interphase FISH Detection of BCL2 Rearrangement in Follicular Lymphoma Using Breakpoint-Flanking Probes." *Genes Chromosomes and Cancer* 27 (1): 85–94. https://doi.org/10.1002/(SICI)1098-2264(200001)27:1<85::AID-GCC11>3.0.CO;2-9.

VanGuilder, Heather D., Kent E. Vrana, and Willard M. Freeman. 2008. "Twenty-Five Years of Quantitative PCR for Gene Expression Analysis." *BioTechniques*. https://doi.org/10.2144/000112776.

Vardi, Anna, Andreas Agathangelidis, Lesley Ann Sutton, Paolo Ghia, Richard Rosenquist, and Kostas Stamatopoulos. 2014. "Immunogenetic Studies of Chronic Lymphocytic Leukemia: Revelations and Speculations about Ontogeny and Clinical Evolution." *Cancer Research*. American Association for Cancer Research Inc. https://doi.org/10.1158/0008-5472.CAN-14-0630.

Vermeesch, Joris R., Cindy Melotte, Guy Froyen, Steven Van Vooren, Binita Dutta, Nicole Maas, Stefan Vermeulen, et al. 2005. "Molecular Karyotyping: Array CGH Quality Criteria for Constitutional Genetic Diagnosis." In *Journal of Histochemistry and Cytochemistry*, 53:413–22. https://doi.org/10.1369/jhc.4A6436.2005.

Victora, Gabriel D., and Michel C. Nussenzweig. 2012. "Germinal Centers." *Annual Review of Immunology*. https://doi.org/10.1146/annurev-immunol-020711-075032.

Vogt, Niklas, Beiying Dai, Tabea Erdmann, Wolfgang E. Berdel, and Georg Lenz. 2017. "The Molecular Pathogenesis of Mantle Cell Lymphoma." *Leukemia and Lymphoma*. Taylor and Francis Ltd. https://doi.org/10.1080/10428194.2016.1248965.

Wagener, Rabea, Julian Seufert, Francesco Raimondi, Susanne Bens, Kortine Kleinheinz, Inga Nagel, Janine AltmüllerAltm, et al. n.d. "Brief Report LYMPHOID NEOPLASIA The Mutational Landscape of Burkitt-like Lymphoma with 11q Aberration Is Distinct from That of Burkitt Lymphoma." www.icgc.org.

Wang, Qingyu, Cooduvalli S. Shashikant, Matthew Jensen, Naomi S. Altman, and Santhosh Girirajan. 2017. "Novel Metrics to Measure Coverage in Whole Exome Sequencing Datasets Reveal Local and Global Non-Uniformity." *Scientific Reports* 7 (1). https://doi.org/10.1038/s41598-017-01005-x.

Wang, Sophia S., Wendy Cozen, James R. Cerhan, Joanne S. Colt, Lindsay M. Morton, Eric A. Engels, Scott Davis, et al. 2007. "Immune Mechanisms in Non-Hodgkin Lymphoma: Joint Effects of the TNF G308A and IL10 T3575A Polymorphisms with Non-Hodgkin Lymphoma Risk Factors." *Cancer Research* 67 (10): 5042–54. https://doi.org/10.1158/0008-5472.CAN-06-4752.

Watkins, A. James, Rifat A. Hamoudi, Naiyan Zeng, Qingguo Yan, Yuanxue Huang, Hongxiang Liu, Jianzhong Zhang, et al. 2012. "An Integrated Genomic and Expression Analysis of 7q Deletion in Splenic Marginal Zone Lymphoma." *PLoS ONE* 7 (9). https://doi.org/10.1371/journal.pone.0044997.

Swerdlow, Travis, E Brambilla, H K Hermelink, Harris C C Eds, J N Eble, G Sauter, and J I Epstein. 2008. *World Health Organization Classification of Tumours This Book and All Other Volumes of the Series Can Be Purchased : From All Countri Es*. World Health Organization.

Waterfall, Joshua J; Arons, Evgeny; Walker, Robert L; Pineda, Marbin; Roth, Laura; Killian, J Keith; Abaan, Ogan D; Davis, Sean R; Kreitman, Robert J; Meltzer, Paul S (2013). *High prevalence of MAP2K1 mutations in variant and IGHV4-34–*

expressing hairy-cell leukemias. *Nature Genetics, 46(1), 8–10.*
doi:10.1038/ng.2828

Weinberg, Olga K., Weiyun Z. Ai, M. Rajan Mariappan, Carol Shum, Ronald Levy, and Daniel A. Arber. 2007. "'Minor' BCL2 Breakpoints in Follicular Lymphoma: Frequency and Correlation with Grade and Disease Presentation in 236 Cases." *Journal of Molecular Diagnostics* 9 (4): 530–37. https://doi.org/10.2353/jmoldx.2007.070038.

Weinhold, Nils, Anders Jacobsen, Nikolaus Schultz, Chris Sander, and William Lee. 2014. "Genome-Wide Analysis of Noncoding Regulatory Mutations in Cancer." *Nature Genetics* 46 (11): 1160–65. https://doi.org/10.1038/ng.3101.

Wiestner, Adrian, Mahsa Tehrani, Michael Chiorazzi, George Wright, Federica Gibellini, Kazutaka Nakayama, Hui Liu, et al. 2007. "Point Mutations and Genomic Deletions in CCND1 Create Stable Truncated Cyclin D1 MRNAs That Are Associated with Increased Proliferation Rate and Shorter Survival" 109: 4599–4606. https://doi.org/10.1182/blood-2006-08.

Williams, Cecilia, Fredrik Pontén, Catherine Moberg, Peter Sö, Mathias Uhlé, Jan Pontén, Gisela Sitbon, and Joakim Lundeberg. 1999. "A High Frequency of Sequence Alterations Is Due to Formalin Fixation of Archival Specimens." *American Journal of Pathology*. Vol. 155.

Wu, Chenglin, Noel Fcc De Miranda, Longyun Chen, Agata M Wasik, Larry Mansouri, Wojciech Jurczak, Krystyna Galazka, et al. n.d. "Genetic Heterogeneity in Primary and Relapsed Mantle Cell Lymphomas: Impact of Recurrent CARD11 Mutations." Vol. 7. www.impactjournals.com/oncotarget.

Xiaodan Mai, PhD, MBBSa, Michael J. LaMonte, PhD, MPHa, Kathleen M. Hovey, MSa, Jo L. Freudenheim, PhDa, Christopher A. Andrews, PhDb, Robert J. Genco, DDS, PhDc, and Jean Wactawski-Wende, PhDa. 2017. "乳鼠心肌提取 HHS Public Access." *Physiology & Behavior* 176 (12): 139–48. https://doi.org/10.1038/ng.2828.High.

Xu-Monette, Zijun Y., Lin Wu, Carlo Visco, Yu Chuan Tai, Alexander Tzankov, Wei Min Liu, Santiago Montes-Moreno, et al. 2012. "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 (19): 3986–96. https://doi.org/10.1182/blood-2012-05-433334.

Xu, Chao, Jigang Zhang, Yu Ping Wang, Hong Wen Deng, and Jian Li. 2014. "Characterization of Human Chromosomal Material Exchange with Regard to the Chromosome Translocations Using Next-Generation Sequencing Data." *Genome Biology and Evolution* 6 (11): 3015–24. https://doi.org/10.1093/gbe/evu234.

Yang, Shan, Stephen E. Lincoln, Yuya Kobayashi, Keith Nykamp, Robert L. Nussbaum, and Scott Topper. 2017. "Sources of Discordance among Germ-Line Variant Classifications in ClinVar." *Genetics in Medicine* 19 (10): 1118–26. https://doi.org/10.1038/gim.2017.60.

Yang, Yibin, Arthur L. Shaffer, N. C.Tolga Emre, Michele Ceribelli, Meili Zhang, George Wright, Wenming Xiao, et al. 2012. "Exploiting Synthetic Lethality for the Therapy of ABC Diffuse Large B Cell Lymphoma." *Cancer Cell* 21 (6): 723–37. https://doi.org/10.1016/j.ccr.2012.05.024.

Yohe, Sophia L., Alexis B. Carter, John D. Pfeifer, James M. Crawford, Allison Cushman-Vokoun, Samuel Caughron, and Debra G.B. Leonard. 2015. "Standards for Clinical Grade Genomic Databases." *Archives of Pathology and Laboratory Medicine* 139 (11): 1400–1412. https://doi.org/10.5858/arpa.2014-0568-CP.

Yohe, Sophia, and Bharat Thyagarajan. 2017. "Review of Clinical Next-Generation Sequencing." *Archives of Pathology and Laboratory Medicine*. College of American Pathologists. https://doi.org/10.5858/arpa.2016-0501-RA.

Yoshida, Shoko, Yoshitaka Kaneita, Yutaka Aoki, Masao Seto, Shigeo Mori, and Masatsugu Moriyama. 1999. "Identification of Heterologous Translocation Partner Genes Fused to the BCL6 Gene in Diffuse Large And-Cell Lymphomas: 5'-RACE and LA-PCR Analyses of Biopsy Samples." *Oncogene* 18 (56): 7994–99. https://doi.org/10.1038/sj.onc.1203293.

Young, Ken H., Karen Leroy, Michael B. Møller, Gisele W.B. Colleoni, Margarita Sánchez-Beato, Fábio R. Kerbauy, Corinne Haioun, et al. 2008. "Structural Profiles of TP53 Gene Mutations Predict Clinical Outcome in Diffuse Large B-Cell Lymphoma: An International Collaborative Study." *Blood* 112 (8): 3088–98. https://doi.org/10.1182/blood-2008-01-129783.

Young, Ken H, Dennis D Weisenburger, Bhavana J Dave, Lynette Smith, Warren Sanger, Javeed Iqbal, Elias Campo, et al. 2007. "Mutations in the DNA-Binding Codons of TP53, Which Are Associated with Decreased Expression of TRAIL Receptor-2, Predict for Poor Survival in Diffuse Large B-Cell Lymphoma." https://doi.org/10.1182/blood-2007-02.

Yuan, Ying Ying, Hua Yuan Zhu, Jia Zhu Wu, Yi Xia, Jin Hua Liang, Wei Wu, Lei Cao, et al. 2019. "The Percentage of Cells with 17p Deletion and the Size of 17p Deletion Subclones Show Prognostic Significance in Chronic Lymphocytic Leukemia." *Genes Chromosomes and Cancer* 58 (1): 43–51. https://doi.org/10.1002/gcc.22692.

Zenz, Thorsten, John G. Gribben, Michael Hallek, Hartmut Döhner, Michael J. Keating, and Stephan Stilgenbauer. 2012. "Risk Categories and Refractory CLL in the Era of Chemoimmunotherapy." *Blood* 119 (18): 4101–7. https://doi.org/10.1182/blood-2011-11-312421.

Zenz, Thorsten, Daniel Mertens, Ralf Küppers, Hartmut Döhner, and Stephan Stilgenbauer. 2010. "From Pathogenesis to Treatment of Chronic Lymphocytic Leukaemia." *Nature Reviews Cancer*. https://doi.org/10.1038/nrc2764.

Zhang, Jenny, Dereje Jima, Andrea B. Moffitt, Qingquan Liu, Magdalena Czader, Eric D. Hsi, Yuri Fedoriw, et al. 2014. "The Genomic Landscape of Mantle Cell Lymphoma Is Related to the Epigenetically Determined Chromatin State of Normal B Cells." *Blood* 123 (19): 2988–96. https://ashpublications.org/blood/article-pdf/123/19/2988/1375796/2988.pdf.

Zimonjic, D B, C Keck-Waggoner, and N C Popescu. 2001. "Novel Genomic Imbalances and Chromosome Translocations Involving C-Myc Gene in Burkitt's Lymphoma." *Leukemia*. Vol. 15. www.nature.com/leu.

Zinzani, Pier Luigi. 2012. "The Many Faces of Marginal Zone Lymphoma." *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* 2012: 426–32. https://doi.org/10.1182/asheducation.v2012.1.426.3798535.

Annexes

1. Scientific communications

Gómez Llonín A, Puiggros A, Nonell L, Rodríguez J, RodríguezSantiago B, Mattlin H, Banchs MI, Salido M, Larrayoz MJ, Blanco L, Collado R, Ortega M, Rodríguez-Rivera M, Melero C, Calvo X, Arenillas L, Gimeno E, Armengol L, Ferrer A, Colomo L, Espinet B. UTILIDAD DE LAS SECUENCIAS OFF-TARGET EN LA DETECCIÓN DE ALTERACIONES DE NÚMERO DE COPIA (CNA) POR NGS. EXPERIENCIA EN UN PANEL DIRIGIDO PARA NEOPLASIAS DE CÉLULAS B MADURAS. LX Congreso Nacional de la Sociedad Española de Hematología y Hemoterapia (SEHH). Granada, 11-13 octubre 2018. Haematologica 2018; 103(s2):214, PC-208.

Andrea Gómez-Llonín, Anna Puiggros, Benjamín Rodriguez-Santiago, Lara Nonell, Jairo Rodríguez, Heidi Mattlin, Maria Isabel Banchs, Marta Salido, Sílvia Beà, Anna Enjuanes, Mª José Larrayoz, Laura Blanco, Rosa Collado, Margarita Ortega, Esmeralda de la Banda, María Rodríguez-Rivera, Carme Melero, Xavier Calvo, Leonor Arenillas, Eva Gimeno, Blanca Sánchez, Antonio Salar, Lluís Armengol, Ana Ferrer, Lluís Colomo, Blanca Espinet. DEVELOPMENT OF A TARGETED NEXT-GENERATION SEQUENCING PANEL FOR DETECTION OF TRANSLOCATIONS, CNAs AND MUTATIONS IN MATURE B CELL NEOPLASMS: COMPARISON WITH CONVENTIONAL GENETIC TECHNIQUES. NGS 2018 Conference. Barcelona Biomedical Research Park (PRBB). Barcelona, 9-10 April 2018.

Gómez Llonín A, Puiggros A, Rodríguez Santiago B, Nonell L, Mattlin H, Banchs MI, Salido M, Beà S, Enjuanes A, Larrayoz MJ, Blanco L, Collado R, Ortega M, de la Banda E, Rodríguez Rivera M, Melero C, Calvo X, Arenillas L, Gimeno E, Sánchez B, Salar A, Armengol Ll, Ferrer A, Colomo Ll, Espinet B. DESARROLLO DE UN PANEL DE SECUENCIACIÓN MASIVA PARA LA DETECCIÓN DE TRANSLOCACIONES, CNA Y MUTACIONES EN NEOPLASIAS DE CÉLULAS B MADURAS: COMPARACIÓN CON TÉCNICAS GENÉTICAS CONVENCIONALES. LIX CONGRESO NACIONAL SEHH. Málaga, 26-28 de Octubre de 2017. Haematologica 2012; CO-128.

PC-208

UTILIDAD DE LAS SECUENCIAS OFF-TARGET EN LA DETECCIÓN DE ALTERACIONES DE NÚMERO DE COPIA (CNA) POR NGS. EXPERIENCIA EN UN PANEL DIRIGIDO PARA NEOPLASIAS DE CÉLULAS B MADURAS

Gómez-Llonín A.[1], Puiggros A., Nonell L.[2], Rodríguez J.[3], Rodriguez-Santiago B.[4], Mattlin H.[3], Banchs Maria I.[3], Salido M., Larrayoz Mª J.[5], Blanco L.[6], Collado R.[7], Ortega M.[8], Rodríguez-Rivera M., Melero C., Calvo X., Arenillas L., Gimeno E.[9], Armengol L.[3], Ferrer A., Colomo L., Espinet B.

[1]*Laboratori de Citogenètica Molecular, Laboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar, Barcelona. Grup de Recerca Translacional en Neoplàsies Hematològiques, Cancer Research Program, IMIM-Hospital del Mar, Barcelona.Department of Medical and Health Sciences, Universitat de Barcelona, Barcelona.,* [2]*Servei d'Anàlisi de Microarrays, IMIM (Hospital del Mar Medical Research Institute), Barcelona.,* [3]*qGenomics (Quantitative Genomic Medicine Laboratories), Esplugues de Llobregat.,* [4]*Departamento de Genética, Universitat Autònoma de Barcelona, Hospital de la Santa Creu i Sant Pau, Barcelona.,* [5]*Unidad de Citogenética y de Genética Hematológica, Departamento de Genética, Universidad de Navarra, Pamplona,* [6]*Servei d'Hematologia Hospital Universitari de la Santa Creu i Sant Pau, Barcelona,* [7]*Servicio de Hematología, Consorcio Hospital General Universitario, Valencia,* [8]*Laboratorio de Citogenética, Hospital Vall d'Hebron, Barcelona,* [9]*Servei d'Hematologia Clínica, Hospital del Mar, Barcelona*

Introducción: Las alteraciones en número de copia (CNA) constituyen biomarcadores diagnósticos, pronósticos o predictivos en muchas neoplasias de células B maduras (NCBM). La mayoría de paneles de secuenciación masiva (NGS) se dirigen a la detección de mutaciones y existen pocos paneles comerciales para CNA. Estos utilizan lecturas en las regiones *on-target* y ofrecen una información sesgada a las regiones seleccionadas. La preparación de librerías de NGS por enriquecimiento genera un 40-60% de lecturas que mapean fuera de las regiones diana (secuencias *off-target*). La herramienta bioinformática CopywriteR utiliza estas secuencias para inferir los CNA y podría ser útil para obtener estos datos sin necesidad de aumentar el tamaño y coste de la secuenciación.

Objetivos: 1. Analizar la utilidad de CopywriteR en la detección de CNA diana del panel dirigido de NGS; 2. Comparar el uso de SNPs incluidos en el diseño y de secuencias *off-target* para la identificación de otros CNA; 3. Validar las CNA detectadas por Copywrite R por SNP-arrays.

Pacientes y Métodos: Hemos diseñado un panel de captura dirigido (Roche Nimblegen) para detectar CNA, mutaciones y translocaciones en NCBM. Para CNA incluye 17 regiones diana y 9111 SNPs distribuidos por el genoma. De 110 pacientes con NCBM y 15 sujetos control secuenciados con Illumina NextSeq500 (cobertura media 250X, *paired-end reads* de 150pb) se seleccionaron 37 pacientes con leucemia linfática crónica (LLC). Se alinearon las secuencias frente al genoma de referencia (hg19). Se utilizó CopywriteR para detectar CNA considerando regiones diana para LLC (13q, 12, 11q y 17p), SNPs y secuencias *off-target*. Los resultados se compararon con los obtenidos excluyendo las regiones de SNPs. Se consideraron ganancias aquellos CNA con log2ratio ≥2,2, y pér-

dida ≤1,8. Las CNA detectadas en dos o más controles fueron descartadas. Se compararon las alteraciones detectadas por NGS con los resultados de FISH y SNP-arrays (N=30, Affymetrix).
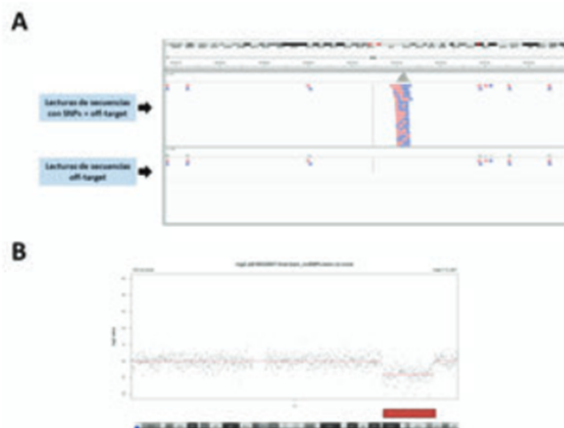

Figura 1. A) Ejemplo de la visualización de las secuencias off-target, con y sin SNPs, en IGV (Integrative Genomics Viewer); B) Representación gráfica de una deleción en 11q identificada por CopywriteR.

Resultados: No se observaron diferencias significativas en el número de CNA detectados con CopywriteR al excluir los SNPs del análisis, generando 408 CNA (Figura). La distancia media entre secuencias *off-target* fue 100Kb. En cuanto a alteraciones diana, se identificaron todas las trisomías 12 y 38/50 deleciones (76%) detectadas por FISH, siendo las discordantes presentes en <20% de células. No se identificaron falsos positivos. El tamaño de las deleciones no difirió significativamente del descrito por SNP-arrays (200Kb-56Mb) (Tabla). Se identificaron correctamente 122/144 (85%) regiones no diana descritas por SNP-arrays (200Kb-173Mb). No obstante, CopywriteR también generó 216 CNA discordantes. Para minimizar la tasa de falsos positivos se descartaron los CNA en regiones no diana de <50 secuencias *off-target* (<5Mb, límite clínicamente relevante según guías de arrays), valores de log2ratio de 1,6-2,4 y CNA en cromosomas sexuales. Esta estrategia permitió identificar 46/61 CNA conocidas (75%), reduciendo a dos los hallazgos discordantes.

Tabla 1. Comparativa de las alteraciones en regiones diana obtenidos mediante FISH y el panel de NGS.

Alteración target	Positivos FISH	Detectados NGS	% Detección NGS	Tamaño por NGS (rango)
del(11q) [ATM]	12	8	67%	14.100 - 55.900 Kb
Trisomía 12	5	5	100%	-
del(13)(q14)	21	16	76%	200 -14.100 Kb
del(17)(p13) [TP53]	17	14	82%	1.000 - 21.100 Kb

Conclusiones: El método CopywriteR és útil para detectar CNA diana por NGS, con alta especificidad y sensibilidad limitada (20%); 2. La inclusión de SNPs en los paneles no es imprescindible para el análisis de CNA en otras regiones, permitiendo reducir el coste de secuenciación; 3. Las secuencias *off-target* permiten identificar CNA pero ofrecen resolución limitada a 5Mb. Aunque pueden ser útiles para obtener información adicional de otras regiones, es imprescindible enriquecer las regiones diana en el panel para optimizar su detección.

329

NGS 2018
Next Generation Sequencing Conference 2018

CRG Centre for Genomic Regulation

iSCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Submission closes: Monday, 29 January 2018
Early registration deadline: Monday, March 12, 2018
Online registration deadline: Monday, April 2, 2018

BARCELONA, SPAIN
April 9 - 11, 2018

Mark your calendars!

Development of a targeted next-generation sequencing panel for detection of translocations, cnas and mutations in mature b cell neoplasms: comparison with conventional genetic techniques

Andrea Gómez-Llonín[1,2,3], Anna Puiggros[1,2], Benjamín Rodríguez-Santiago[4,5], Lara Nonell[6], Jairo Rodríguez[4], Heidi Mattlin[4], Maria Isabel Banchs[4], Marta Salido[1,2], Sílvia Beà[7], Anna Enjuanes[8], Mª José Larrayoz[9], Laura Blanco[10], Rosa Collado[11], Margarita Ortega[12], Esmeralda de la Banda[13], María Rodríguez-Rivera[1,2], Carme Melero[1,2], Xavier Calvo[1,2], Leonor Arenillas[1,2], Eva Gimeno[14], Blanca Sánchez[14], Antonio Salar[14], Lluís Armengol[4], Ana Ferrer[1,2], Lluís Colomo[1,2], Blanca Espinet[1,2]

[1]Laboratori de Citogenètica Molecular, Laboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar, Barcelona, Spain. [2]Grup de Recerca Translacional en Neoplàsies Hematològiques, Cancer Research Program, IMIM-Hospital del Mar, Barcelona, Spain. [3]Department of Medical and Health Sciences, Universitat de Barcelona, Barcelona, Spain. [4]qGenomics (Quantitative Genomic Medicine Laboratories), Esplugues de Llobregat, Spain. [5]Genetics Department, Universitat Autònoma de Barcelona, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [6]Servei d'Anàlisi de Microarrays, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [7]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS); Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Barcelona, Spain. [8]Unitat de Genòmica, IDIBAPS; Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain. [9]Unidad de Citogenética y de Genética Hematológica, Departamento de Genética, Universidad de Navarra, Pamplona, Spain. [10]Servei d'Hematologia Hospital Universitari de la Santa Creu i Sant Pau, Barcelona, Spain. [11]Servicio de Hematología, Consorcio Hospital General Universitario, Valencia, Spain. [12]Laboratorio de Citogenética, Hospital Vall d'Hebron, Barcelona, Spain. [13]Servei d'Hematologia, IDIBELL-Hospital de Bellvitge, L'Hospitalet de Llobregat, Spain. [14]Servei d'Hematologia Clínica, Hospital del Mar, Barcelona, Spain.

Introduction

- Mature B cell neoplasms show characteristic genetic aberrations which constitute diagnostic, prognostic or predictive biomarkers.
- Classical techniques such as chromosome banding analysis, FISH, CGH-array, PCR and Sanger sequencing are part of the clinical practice.
- The use of massive sequencing panels (NGS) is being implemented in laboratories and may offer technical advantages and additional information.

Aim

1. To develop a targeted NGS panel for mature B-cell neoplasms in order to integrate most of the abnormalities routinely analyzed by classical techniques (gene rearrangements, copy number alterations (CNAs) and mutations) in a unique test.

2. To evaluate the effectiveness of the panel in the detection of previously known alterations.

Patients and methods

- A total of **110 patients with mature B-cell neoplasms** [37 chronic lymphocytic leukemia (CLL), 25 diffuse large B-cell lymphoma (DLBCL) ,19 splenic marginal zone lymphoma (SMZL), 11 follicular lymphoma (FL), 6 mantle cell lymphoma (MCL), 5 Burkitt lymphoma (BL), 4 mucosa-associated lymphoid tissue lymphoma (MALT-L) and 3 hairy cell leukemia (HCL)] **and 15 healthy subjects** were analyzed with a custom NGS panel targeting 11 genes involved in translocations, 35 genes frequently mutated and 17 regions with CNAs with clinical value (Figure 1).

Figure 1. Workflow scheme.
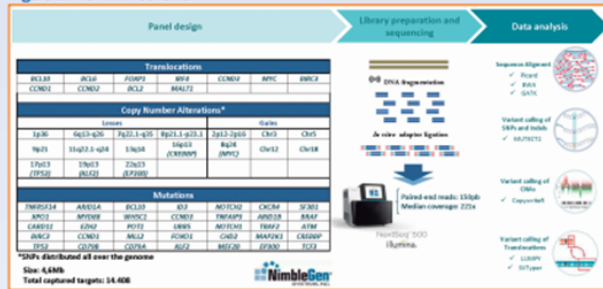
Results

- We detected 100% (37/37) of the known mutations, 80.7% (482/597) of the CNAs and 90% (45/50) of the rearrangements previously identified by conventional techniques (Figure 2).
- For the detection of mutations and CNAs, the sensitivity threshold was established at 8% and 10%, respectively.
- Regarding rearrangements:
  - The panel properly detected 83.9% (26/31) of the previously known gene partners (13/14 BCL2/IGH, 5/6 CCND1/IGH, 5/8 MYC/IGH, 1/1 MYC/IGK, 1/1 BCL6/IGH and 1/1 BCL6/6p21).
  - In the remaining 14 BCL6 and MYC rearrangements in which the partner was unknown, the partner could be identified by the panel.
  - Remarkably, 56% (14/25) of the BCL6 and MYC rearrangements involved IG genes (IGH, IGK or IGL loci).
- Results obtained after comparison with conventional techniques in CLL and DLBCL are shown in Figure 3.

Figure 2. Comparison of the results obtained by conventional techniques and the NGS panel. Only cases with previous information have been included. Patients are represented in rows and grouped by pathology. Alterations are represented in columns.
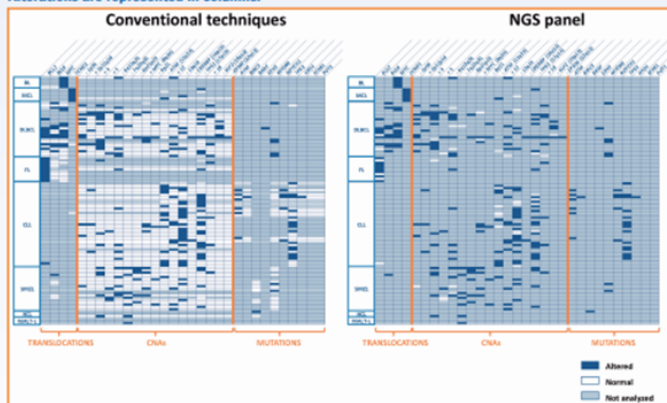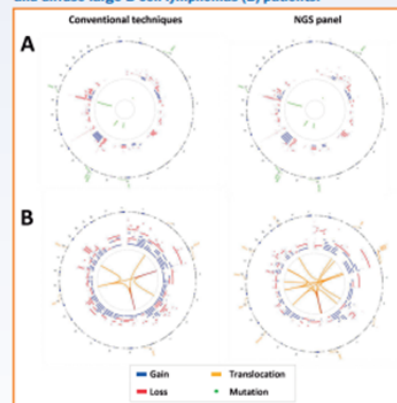

Figure 3. Circos plots of the results obtained by routine techniques and NGS panel in chronic lymphocytic leukemia (A) and diffuse large B-cell lymphomas (B) patients.

Conclusions

- The designed panel has successfully detected the alterations previously described by routine techniques.
- Next steps will include validation of certain rearrangements, discovery of new alterations not previously assessed and refine the design of the panel to improve its applicability in the clinical practice.

CONTACT
agomez.12@alumni.unav.es

CO-128

DESARROLLO DE UN PANEL DE SECUENCIACIÓN MASIVA PARA DETECCIÓN DE TRANSLOCACIONES, CNA Y MUTACIONES EN NEOPLASIAS DE CÉLULAS B MADURAS: COMPARACIÓN CON TÉCNICAS GENÉTICAS CONVENCIONALES

Gómez Llonín A.[1], Puiggros A., Rodríguez Santiago B.[2], Nonell L.[3], Mattlin H.[2], Banchs M.I.[2], Salido M., Beà S.[4], Enjuanes A.[5], Larrayoz M*J.[6], Blanco L.[7], Collado R.[8], Ortega M.[9], De la Banda E.[10], Rodríguez Rivera M.[11], Melero C.[12], Calvo X., Arenillas L., Gimeno E., Sánchez B., Salar A.[13], Armengol L.[13], Ferrer A.[13], Colomo L.[2], Espinet B.

[1]*Laboratori de Citogenètica Molecular, Laboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar, Barcelona/Grup de Recerca Translacional en Neoplàsies Hematològiques, Cancer Research Program, IMIM-Hospital del Mar, Barcelona/Department of Medical and Health Sciences, Universitat de Barcelona, Barcelona,* [2]*qGenomics (Quantitative Genomic Medicine Laboratories), Esplugues de Llobregat,* [3]*Servei d'Anàlisi de Microarrays, IMIM (Hospital del Mar Medical Research Institute), Barcelona,* [4]*Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS),* [5]*Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Barcelona,* [6]*Unitat de Genòmica, IDIBAPS,* [7]*Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain,* [8]*Unidad de Citogenética y de Genética Hematológica, Departamento de Genética, Universidad de Navarra, Pamplona,* [9]*Servei d'Hematologia Hospital Universitari de la Santa Creu i Sant Pau, Barcelona,* [10]*Servicio de Hematología, Consorcio Hospital General Universitario, Valencia,* [11]*Laboratorio de Citogenética, Hospital Vall d'Hebron, Barcelona,* [12]*Servei d'Hematologia, IDIBELL-Hospital de Bellvitge, L'Hospitalet de Llobregat,* [13]*Servei d'Hematologia Clínica, Hospital del Mar, Barcelona*

Introducción: Las neoplasias de células B maduras presentan alteraciones genéticas características que constituyen biomarcadores diagnósticos, pronósticos o predictivos de respuesta a tratamiento. Su determinación, utilizando técnicas clásicas como citogenética convencional, FISH, arrayCGH, PCR y secuenciación Sanger, forma parte de la práctica clínica asistencial. El uso de paneles de secuenciación masiva (NGS) se está implementando en los laboratorios y puede aportar ventajas técnicas e información adicional.

Objetivos: 1.Diseñar un panel dirigido de NGS para la detección de las principales translocaciones cromosómicas, alteraciones en número de copias (CNA) y mutaciones con valor diagnóstico, pronóstico y predictivo en neoplasias de células B maduras que permita integrar la mayoría de determinaciones rutinarias, así como incorporar nuevas alteraciones con relevancia clínica descritas en los últimos años; 2. Evaluar la eficacia del panel en la detección de alteraciones previamente conocidas; 3. Describir nuevas alteraciones y reordenamientos.

Pacientes y Métodos: Se analizaron 110 pacientes: 37 LLC, 25 LBDCG (6 DH y 3 TH), 19 LZME, 11 LF, 6 LCM, 5 LB, 4 MALT y 3 TL. Además, se analizaron 15 individuos control. Se extrajo el DNA de sangre periférica (n=77), médula ósea (n=2) o biopsias de tejido congelado (n=31). Se diseñó un panel de captura NimbleGen SeqCap Ez Choice (Roche Nimblegen) que cubre 11 genes implicados en translocaciones (*BCL2, BCL6, BCL10, MYC, CCND1, CCND2, CCND3, MALT1, BIRC3, FOXP1, IRF4*), 35 genes con mutaciones frecuentes y SNPs distribuidos a lo largo del genoma para identificar CNAs. Las muestras se secuenciaron utilizando Illumina NextSeq500 con una cobertura media de 250X y *paired-end reads* de 150pb. El análisis incluyó el alineamiento de las secuencias frente al genoma de referencia (hg19) (BWA, Picard, GATK), el estudio de las variantes de un solo nucleótido y pequeñas indels (GATK), translocaciones (LUMPY) y variantes de número de copia (CopywriteR).

Resultados: Se detectaron un 91% (50/55) de fusiones en regiones diana identificadas previamente mediante técnicas convencionales (14/19 *BCL2*, 12/12 *BCL6*, 6/6 *CCND1*, 1/1 *MALT1* y 17/17 *MYC*). En un 92% el panel de NGS permitió identificar correctamente la pareja de reordenamiento. Además, se identificaron 2197 CNAs; con técnicas convencionales se habían identificado 563, de las cuales el 67% se validó mediante el panel. Las alteraciones subclonales presentes en menos del 20-30% de la muestra no fueron identificadas. Finalmente, se detectaron 404 mutaciones, de las cuales 82 son patogénicas y 32 se localizan en regiones *hotspot*. Por técnicas convencionales se habían detectado 38 mutaciones, de las cuales el 79% se validó mediante el panel de NGS con un límite de sensibilidad del 10%.

Conclusiones: 1. El panel de NGS diseñado ha permitido detectar un porcentaje muy elevado de translocaciones en genes diana, además de identificar los *partners* de reordenamiento; 2. La estrategia utilizada permite identificar muchas CNAs con un límite de sensibilidad del 30%, 3. Los resultados son preliminares pero prometedores. Es necesario validar la metodología y los criterios de análisis para aumentar la sensibilidad de la técnica.