UNIVERSITAT DE
BARCELONA

The role of genome architecture in normal and neoplastic B cells: a multi-omics approach

Roser Vilarrasa Blasi

UNIVERSITAT DE BARCELONA

DEPARTMENT OF BASIC CLINICAL PRACTICE

FACULTY OF MEDICINE, UNIVERSITY OF BARCELONA

HOSPITAL CLÍNIC OF BARCELONA

INSTITUT D'INVESTIGACIONS BIOMÈDIQUES AUGUST PI I SUNYER (IDIBAPS)

The role of genome architecture in normal and neoplastic B cells: a multi-omics approach

Thesis presented by **Roser Vilarrasa Blasi**

to obtain the degree of Doctor in Biomedicine

PhD Program in Biomedicine

FACULTY OF MEDICINE, UNIVERSITY OF BARCELONA

Doctoral advisor Dr. **José Ignacio Martín Subero**

Tutor Prof. **Elías Campo Güerri**

Barcelona, 2020

Roser Vilarrasa Blasi   Dr. José Ignacio Martín Subero   Prof. Elías Campo Güerri

ABSTRACT

Nuclear organization and its impact on gene regulation have started to be elucidated thanks to the development of chromosomal conformation capture techniques. In this doctoral thesis, *in situ* Hi-C data and nine additional omic layers have been integrated to define and biologically characterize the dynamic changes in three-dimensional (3D) genome architecture across normal B-cell differentiation and in neoplastic cells from chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL) patients. Beyond the classical genome organization into active and inactive compartments, a third intermediate 3D compartment enriched in poised and polycomb-repressed chromatin has been identified (**Study 1**). It has been observed that during B-cell differentiation, a 28% of the 3D genome structure changed, being an extensive activation from naive to germinal center B cells and a reversal into a naive-like 3D genome upon further maturation into memory B cells the most remarkable features. In case of neoplastic B cells, both CLL and MCL displayed entity and subtype-specific alterations in chromosome organization. Those alterations comprised large chromatin blocks containing key disease-specific genes such as *EBF1* in CLL or *SOX11* in MCL. Chromosomal conformation maps were also assessed to define structural variants in neoplastic cells (**Study 2**). The t(11;14) translocation in MCL cases leading to *CCND1* deregulation was associated with a breakpoint-dependent 3D chromatin reconfiguration creating new topologically associating domain (TAD) borders. Moreover, additional rearrangements have been identified by Hi-C in MCL and have been confirmed by standard methods such as cytogenetic analyses and next generation sequencing. An integrative multi-omics approach combining 3D genome architecture with histone modifications and DNA methylation allowed the identification of candidate epigenetic drivers in MCL (**Study 3**). In aggressive, conventional MCLs with overexpression of the *SOX11* oncogene, a distant regulatory region looping to the *SOX11* promoter was uncovered promoting oncogene deregulation in a biallelic fashion. Additional experiments revealed that the PAX5 transcription factor may play a role in activating the distant *SOX11* enhancer (**Study 4**). This doctoral thesis highlights the role of the chromatin architecture as a key epigenomic player associated with the normal differentiation and neoplastic transformation of B cells.

RESUM

L'organització tridimensional de l'ADN en el nucli de la cèl·lula i el seu impacte en la regulació gènica s'està començant a entendre gràcies al desenvolupament de les tècniques de captura de la conformació cromosòmica. En aquesta tesis doctoral, s'han integrat dades d'*in situ* Hi-C i nou capes òmiques addicionals per definir i caracteritzar biològicament els canvis dinàmics en l'arquitectura tridimensional (3D) del genoma durant la diferenciació de la cèl·lula B normal i en cèl·lules neoplàsiques de pacients amb leucemia linfocítica crònica (LLC) i linfoma de cèl·lules del mantell (LCM). Més enllà de l'organització clàssica del genoma en compartiments actius i inactius, es va identificar un tercer compartiment intermedi enriquit en cromatina associada al complex repressor polycomb (**Estudi 1**). Durant la diferenciació de la cèl·lula B, es va observar que el 28% de l'estructura 3D del genoma canviava. En concret, es va detectar una activació extensa durant la transició entre les cèl·lules naive i cèl·lules del centre germinal, i aquesta activació revertia a un estat similar al de la cèl·lula naive una vegada la cèl·lula del centre germinal es diferenciava a cèl·lula memòria. En el cas de la LLC i el LCM, ambdues neoplàsies van mostrar alteracions en la organización cromosòmica tant a nivell d'entitat com a nivell de subtipus de la malaltia. Aquestes alteracions afectaven grans blocs de cromatina que contenen gens claus en la patogènesis de cada malaltia, com va ser el cas de *EBF1* en LLC o *SOX11* en LCM. Tanmateix, els mapes de conformació cromosòmica es van examinar per definir alteracions cromosòmiques en cèl·lules neoplàsiques (**Estudi 2**). La translocació t(11;14) en casos de LCM es va associar a una reestructuració de la cromatina creant un nou domini associat a la topología (TAD) depenent del punt de ruptura en el cromosoma 11 prop del locus de *CCND1*. L'ús del Hi-C va permetre també confirmar la presència d'alteracions cromosòmiques identificades mitjançant mètodes estàndards com són la citogenètica convencional i la seqüenciació de nova generació i, fins i tot, es va identificar algunes alteracions addicionals. A més, una aproximació integradora combinant l'arquitectura 3D del genoma amb la modificació de les histones i la metilació de l'ADN va permetre identificar alteracions epigenètiques amb funció patogènica en LCM (**Estudi 3**). En particular, en el LCM on es sobreexpressa l'oncogèn *SOX11*, es va identificar i estudiar una regió reguladora distant que forma un bucle tridimensional amb el promotor de *SOX11*, promovent així la desregulació de l'oncogèn de forma bial·lèlica. Finalment,

experiments addicionals van revelar que el factor de transcripció PAX5 pot jugar un paper en l'activació de la regió reguladora distant associada a *SOX11* (**Estudi 4**). En conclusió, aquesta tesis doctoral destaca el paper de l'arquitectura de la cromatina com un factor clau associat a la diferenciació normal de la cèl·lula B i la seva transformació neoplàsica.

RESUMEN

La organización tridimensional del ADN en el núcleo celular y su impacto en la regulación génica está empezando a entenderse gracias al desarrollo de las técnicas de captura de la conformación cromosómica. En esta tesis doctoral, se han integrado datos de *in situ* Hi-C y nueve capas ómicas adicionales para definir y caracterizar biológicamente los cambios dinámicos en la arquitectura tridimensional (3D) del genoma durante la diferenciación de la célula B normal y en células neoplásicas de pacientes con leucemia linfocítica crónica (LLC) y linfoma de células del manto (LCM). Más allá de la organización clásica del genoma en compartimientos activos e inactivos, se identificó un tercer compartimiento intermedio enriquecido en cromatina asociada al complejo represor polycomb (**Estudio 1**). Durante la diferenciación de la célula B, se observó que el 28% de la estructura 3D del genoma cambiaba. En concreto, se detectó una extensa activación en la transición entre las células naive y células del centro germinal, y dicha activación revertía a un estado similar al de la célula naive una vez la célula de centro germinal se diferenciaba a una célula memoria. En el caso de la LLC y el LCM, ambas neoplasias mostraron alteraciones en la organización cromosómica tanto a nivel de entidad como a nivel de los subtipos de la enfermedad. Estas alteraciones afectaban grandes bloques de cromatina que contienen genes claves en la patogénesis de cada enfermedad, como era el caso de *EBF1* en LLC y de *SOX11* en LCM. Los mapas de conformación cromosómica fueron a la vez examinados para definir alteraciones cromosómicas en las células neoplásicas (**Estudio 2**). La translocación t(11;14) en los casos de LCM se asoció a un reconfiguración de la cromatina creando un nuevo domino asociado a la topología (TAD) dependiente del punto de rotura en el cromosoma 11 cerca del locus de *CCND1*. El uso de Hi-C permitió también confirmar la presencia de alteraciones cromosómicas identificadas mediante métodos estándar como son la citogenética convencional y la secuenciación de nueva generación, e incluso identificar algunas alteraciones adicionales. Además, una aproximación integradora combinando la arquitectura 3D del genoma con las modificaciones de las histonas y la metilación del ADN posibilitó la identificación de alteraciones epigenéticas con función patogénica en el LCM (**Estudio 3**). En particular, en los LCM que sobreexpresan el oncogén *SOX11*, se identificó y estudió una región reguladora distante que formaba un bucle tridimensional con el promotor

de *SOX11*, promoviendo así la desregulación del oncogén de forma bialélica. Finalmente, experimentos adicionales revelaron que el factor de transcripción PAX5 podría jugar un papel en la activación de la región reguladora distante asociada a *SOX11* (**Estudio 4**). En conclusión, esta tesis doctoral destaca el papel de la arquitectura tridimensional de la cromatina como un factor clave asociado con la diferenciación normal de la célula B y su transformación neoplásica.

TABLE OF CONTENTS

LIST OF SELECTED ABBREVIATIONS

Ab	Antibody
AID	Activation-induced cytidine deaminase
ATAC	Assay of transposase-accessible chromatin
BCR	B-cell receptor
BM	Bone marrow
CGI	CpG island
ChIP	Chromatin immunoprecipitation
CLL	Chronic lymphocytic leukemia
cMCL	Conventional MCL
CT	Chromosome Territory
CSR	Class-switch recombination
DMR	Differentially methylated region
DNMT	DNA methyltransferase
FISH	Fluorescent *in situ* hybridization
GC	Germinal center
GCBC	Germinal center B cell
Hi-C	High-throughput chromosome conformation capture
HN	Hematological neoplasm
HSC	Hematopoietic stem cells
ICGC	International Cancer Genome Consortium
IG	Immunoglobulin
IGHV	Immunoglobulin heavy chain variable region
LAD	Lamina-associated domain
LMPP	Lymphoid-primed multipotent progenitor
MBC	Memory B cell
mCLL	CLL with mutated IGHV
MCL	Mantle cell lymphoma
NBC	Naive B cell
NGS	Next generation sequencing
nnMCL	Leukemic non-nodal MCL
PB	Peripheral blood

PC	Plasma cell
PCA	Principal component analysis
PcG	Polycomb group
PTM	Post-translation modification
RT	Room temperature
SHM	Somatic hypermutation
SV	Structural variant
T	Tonsil
TAD	Topologically associating domain
TF	Transcription factor
TFBS	Transcription factor binding site
TSS	Transcription start site
uCLL	CLL with unmutated IGHV
UTR	Untranslated region
WGBS	Whole-genome bisulfite sequencing
WHO	World health organization
3D	Three-dimensional
3C	Chromosome conformation capture
4C	Circular chromosome conformation capture
5C	Chromosome conformation capture carbon copy

INTRODUCTION

1. Epigenetics

1.1 Historical overview and definition

DNA (deoxyribonucleic acid) is the blueprint upon which life is based and transmitted. The sequence of the human genome was published in 2001 (Lander et al., 2001; Venter et al., 2001). Since then we have been able to read our own "instruction book" which arguably demarcates one of the biggest breakthroughs in the history of life sciences (Denker and de Laat, 2016). Decoding the DNA allowed to advance and speed up fields related to human development, physiology, (bio)medicine and evolution. However, we soon discovered that knowing DNA was not enough. All the cells in multicellular organisms are genetically homogenous but structurally and functionally heterogeneous (Wang and Chang, 2018), which evidences that genetics alone cannot explain phenotypic manifestations. So even though the genome encodes for potential information, the expression or silencing do not directly depend on the sequence itself, but rather on epigenetic mechanisms (Martín-Subero, 2011).

The introduction into modern biology of the word "epigenetics" is due to Conrad Hal Waddington back in 1942, emphasizing its relationship to the classical concept of "epigenesis". The term *genesis* (gr.) can be translated as origin, and *epi* as on or after. Epigenesis was coined by the physician and physiologist William Harvey around 1650 describing development as a gradual process of increasing complexity from initially homogenous material in the egg, an idea that was originally proposed by Aristotle (Deichmann, 2016). Based on epigenesis, Waddington defined epigenetics as the whole complex of developmental processes that lie between genotype and phenotype (Waddington, 1942). Or in other words, the mechanisms by which genotype give rise to phenotypes during development (Waddington, 1957). Waddington built an extensively-cited model of an epigenetic landscape, illustrating the various developmental pathways a cell might take during differentiation, attributing a major role to the genes which underlie the landscape, acting to structure it (**Figure 1**). By contrast, Arthur Riggs and colleagues tuned the epigenetics definition as the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence (Riggs et al., 1996). This new version born in 1975 from independent observations made by Riggs (Riggs, 1975) and Holliday and Pugh (Holliday and Pugh, 1975) outlined a molecular model of the switching of gene

activities, and also the heritability of gene activity or inactivity. It is based on the enzymatic methylation of cytosine in DNA, a DNA modification, suggested to play a role in switching on and off genes during development (Holliday, 2006). Adrian Bird proposed the most recent definition keeping the sense of prevailing usages but avoiding the heritability constraints. Bird took into account the biology of chromatin, including the complex language of chromatin marks, the transcriptional effects of RNA interference and, the effects of the higher-order structure of chromosomes and the nucleus, and consequently defined epigenetic events as the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states (Bird, 2007). Finally, in 2008 a consensus definition of epigenetics was established as stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence (Berger et al., 2009; Gnyszka et al., 2013). This epigenetic language comprises a serial of hierarchical layers each one influencing gene activity and thus cellular programs. Moving from a one dimension to three dimensions, the key epigenetic players are DNA methylation, chromatin marks (histone modifications or variants), nucleosome positioning, nucleosome accessibility and three-dimensional chromosome conformation. Their characterization at the whole-genome scale defines the so-called epigenome, which will be described in greater detail in the following sections.
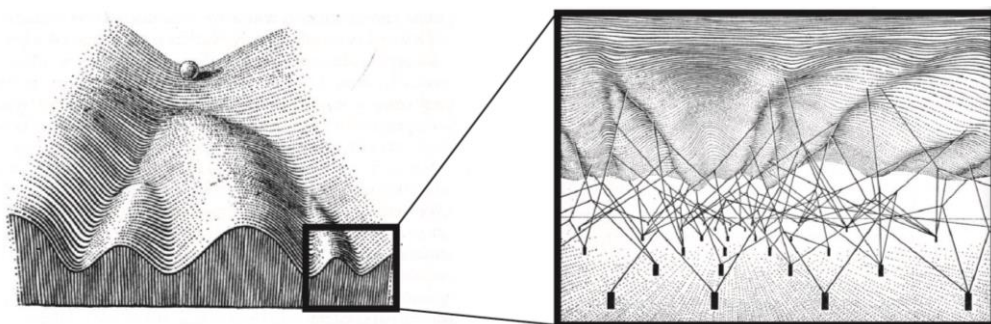
Figure 1. **Epigenetics.** The "epigenetic landscape" (*left*) and the "complex system of interactions underlying the epigenetic landscape" (*right*) proposed by Conrad Waddington. A ball rolling down on an inclined surface with a cascade of branching ridges and valleys. A metaphor for the branching pathways of cell fate determination, drawn upon the presence or absence of the underlying genes, which determine the path on a certain point of divergence. Figure adapted from Waddington, 1957.

1.2 DNA methylation

1.1.1 Molecular basis of DNA methylation

DNA methylation is one of the most intensely studied epigenetic modifications in mammals. DNA methylation is involved in orchestrating proper transcription programs as well as other biologic processes, such as X-chromosome inactivation (Mohandas et al., 1981; Wolf and Migeon, 1982), control of gene expression patterns (Bird, 2007), genomic imprinting (Li et al., 1993; Swain et al., 1987) and has effects on cellular growth and genomic stability (Bird, 2002; Takai and Jones, 2002).

At the biochemical level, DNA methylation consists on the covalent addition of a methyl group ($-CH_3$) at the 5' position of cytosine residue generating a 5-methylcytosine (5mC), generally within the context of CpG dinucleotides (**Figure 2**). Typically, CpG dinucleotides are concentrated in clusters, called "CpG islands" (CGIs) (Bird, 1986), defined as regions of more than 200 base pairs (bps) with a G+C content of at least 50% and a ratio of observed to statistically expected CpG frequencies of at least 0.6 (Deaton and Bird, 2011; Venter et al., 2001). Although the 55 million CpG dinucleotides per diploid cell only make approximately 1% of the human genome, they are enriched in promoter and first exon regions (Vinson and Chatterjee, 2012). In the human genome, nearly 60% of all human promoters contain CGIs (Saxonov et al., 2006).
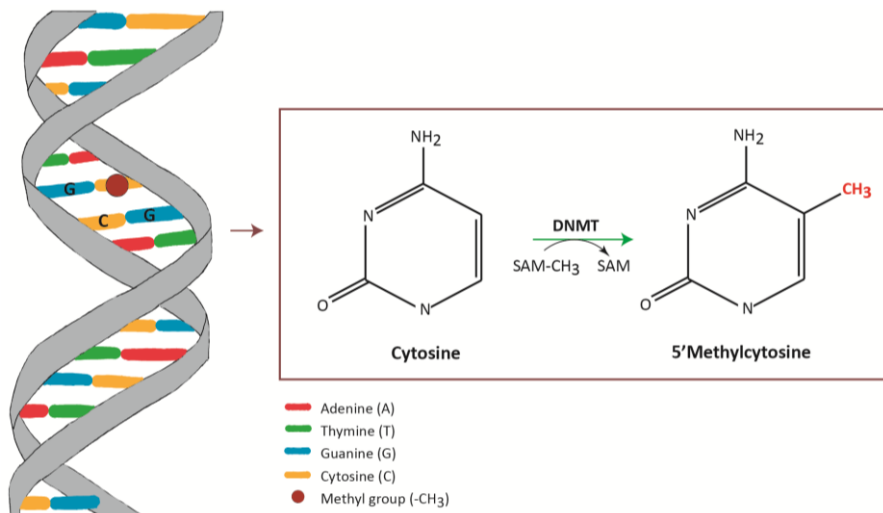
Figure 2. DNA methylation reaction. Methylation occurs by the addition of a methyl group at the 5' site of cytosine residues (represented as a red-brown sphere) when present as CG dinucleotides. Figure adapted from Barros and Offenbacher, 2009.

Cytosine methylation is mediated by DNA methyltransferases (DNMTs), which are in charge of catalyzing the transfer of a methyl group from S-adenosyl methionine (SAM) onto cytosine (**Figure 2**). Five members of the DNMT family have been identified in mammals: DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. Specifically, DNMT1, DNMT3A and DNMT3B seem to be involved in generating the global 5mC pattern of the genome. These independently encoded proteins can be overall classified as enzymes mediating *de novo* DNA methylation (DNMT3A and DNMT3B) or maintenance of DNA methylation patterns (DNMT1) (Law and Jacobsen, 2010). In contrast to DNA methylation, the exact mechanisms leading to DNA demethylation still remains controversial. It is postulated that DNA demethylation may occur passively through lack of maintenance during cell division or actively through the function of ten-eleven translocation family of proteins or activation-induced cytidine deaminase (AID) followed by base-excision repair that introduces an unmethylated cytosine (Bhutani et al., 2011).

1.1.2 Methods for DNA methylation detection

A wide range of methods have been developed to detect and quantify DNA methylation. 5mCs can be detected by three general strategies based on (i) bisulfite conversion of DNA, (ii) methyl-sensitive restriction enzymes, or (iii) immunoprecipitation (affinity enrichment) assays. These methods can be coupled with different downstream techniques and platforms (i.e. novel microarrays or next-generation sequencing) to properly measure DNA methylation, at different scale and resolution (Laird, 2010).

In particular, bisufite sequencing of genomic DNA is considered the gold standard for analyzing the methylation state of CpG sites within the genome at a single-base pair resolution (Bibikova, 2016; Eckhardt et al., 2006). Bisulfite treatment of DNA converts unmethylated cytosines into uracil, whereas methylated cytosines stay unchanged (**Figure 3**). Thus, sequencing after bisulfite treatment allows to detect and quantify DNA methylation levels in individual CpG residues (Lee et al., 2013). There are several methods implemented to analyze bisulfite-converted DNA. The CpG methylation status in a sequence up to 100bp in length can be determined by bisulfite pyrosequencing, which is a quantitative method for DNA methylation analysis (Tost and Gut, 2007). This technique can be applied to detect DNA methylation markers relevant for clinical use

serving as diagnostic or prognostic tool (Heyn and Esteller, 2012; Locke et al., 2019; Mikeska and Craig, 2014; Queirós et al., 2015).
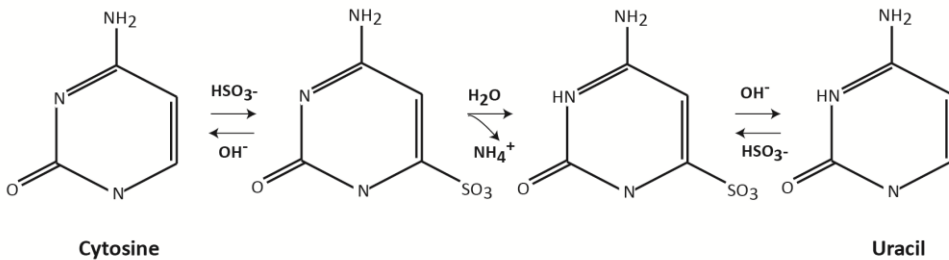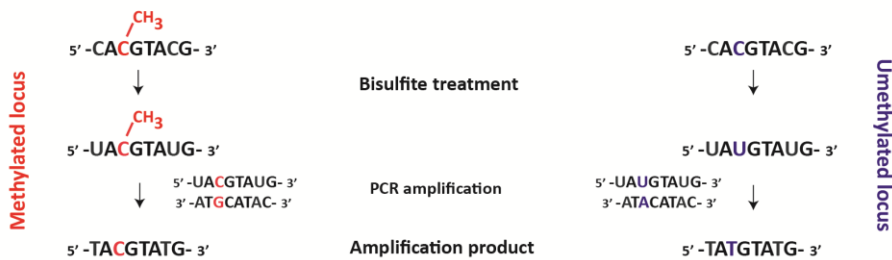
Figure 3. Bisulfite conversion reaction of the DNA. **A.** Outline of the chemical reaction that underlines the bisulfite-catalyzed conversion of cytosine to uracil. **B.** Example of DNA methylated and unmethylated locus upon bisulfite conversion.

The limitation of bisulfite pyrosequencing is that a small number of CpGs can be evaluated. Hence, to increase the number of simultaneously analyzed CpGs, microarrays platforms can be used, such as the Infinium Human Methylation 450K BeadChip (or the new version of it, the EPIC array), which allows us to quantify CpG methylation across the genome at single CpG resolution (Bibikova et al., 2011). The 450K BeadChip array is able to measure the methylation levels of 458,512 CpG sites located in 5' regions (comprising promoter regions up to 1500bp from the TSS, 5'UTR and the first exon) as well as in the gene body and 3'UTR regions of nearly all RefSeq genes, noncoding RNAs (ncRNAs), microRNAs (miRNAs), and intergenic regions. Moreover, it covers 96% of all known CGIs. Furthermore, the 450K array requires a low sample input (approx. 500ng of DNA). Thus, these advantages together with its relatively low cost, makes this array a potent tool for high-throughput DNA methylation profiling of large sample cohorts. However, this technique does not permit the characterization of the whole-genome DNA methylome at a single-bp resolution. In

order to solve this limitation, whole-genome bisulfite sequencing (WGBS), a robust and comprehensive next-generation sequencing (NGS) based technology, has been described (Lister et al., 2009). This method allows to obtain an unbiased representation of DNA methylation maps throughout the genome (over 90% of all cytosines can be measured) and constant improvements in this technology help to increase its accuracy. Even though an enormous progress has also been made toward completing whole-genome DNA methylomes, the bioinformatics techniques and the economic cost are disadvantages of WGBS. Furthermore, although bisulfite treatment is a reliable method to distinguish methylated and unmethylated cytosines, the recent discovery of 5mC derivatives, such as 5-hydroxymethylated (5hmC), call for more careful use of this technique, as it does not distinguish between 5mC and 5hmC. This implies that a proportion of genomic loci identified as methylated may actually be hydroxymethylated (Huang et al., 2010). In normal and neoplastic B cells, which are studied in this doctoral thesis, the proportion of 5hmC is relatively low, and therefore, bisulfite-based methods used in this work, i.e. bisulfite pyrosequencing, 450K arrays and WGBS represent faithful strategies to detect and measure 5mC levels.

1.3 Chromatin accessibility and histone modifications

1.3.1 Histones and nucleosomes

Each diploid human cell contains a total of 6 billion bps of DNA per cell, making up a 2 meters long string (Annunziato, 2008). Interestingly this DNA sequence fits into a nucleus of around 10 microns in diameter. This is the result of arranging DNA in a hierarchical and dynamic manner (Hergeth and Schneider, 2015).

The lowest layer of compaction of the DNA occurs through wrapping around 147bp onto octameric histone proteins called nucleosomes, forming together a structure called chromatin (Van Holde, 1988; Wolffe, 1999). Each octamer consists of two H3/H4 histone heterodimers and two H2A/H2B histone heterodimer (**Figure 4A**). In this way, nucleosomes shortens the fiber length about sevenfold (Cutter and Hayes, 2015). Even though nucleosomes likely evolved to protect and compact chromatin (Malik and Henikoff, 2003), these structures also have an impact on gene accessibility, restricting access to cellular component such as DNA-binding transcription factors and RNA polymerases (Li et al., 2007). Accordingly, three properties evolved to provide

dynamism at the nucleosomes positioning on the chromatin fiber such as: composition alteration, covalent modification and translational repositioning (Saha et al., 2006).
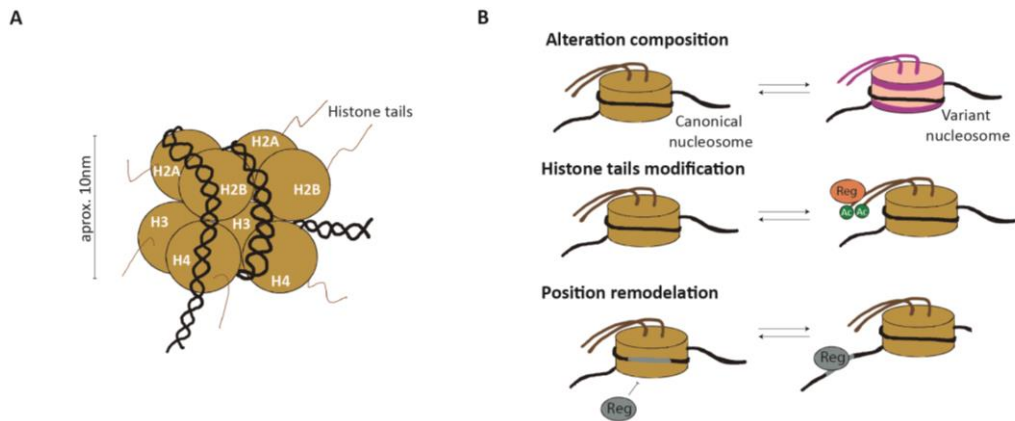
Figure 4. Nucleosome. A. Structure. **B**. Dynamic properties: (i) composition alteration, (ii) nucleosome modification, and (iii) nucleosome repositioning. Figure adapted from Saha et al., 2006.

Firstly, the canonical histones (H2A, H2B, H3, and H4), which are deposited on a replication-coupled manner to package newly replicated genome, can be replaced with histone variants that alter nucleosome structure (Weber and Henikoff, 2014), (**Figure 4B**). A second mode for dynamic chromatin can be caused by post-translational modifications (PTMs) of histones (Strahl and Allis, 2000). The N-terminal tails of histones protrude out of the nucleosomes, to which upon signals and particular conditions of the cell, effectors are recruited to induce specific transcriptional outcomes (**Figure 4B**). More than eleven types of PTMs affecting all histones are known, including among others methylation, acetylation, phosphorylation, sumoylation, and ubiquitination of various amino acid residues, which mainly affect lysine (K) and arginine (R) residues (Bannister and Kouzarides, 2011; Kouzarides, 2007). For instance, one of the first modifications to be studied was histone acetylation, which has been associated with gene expression (Grunstein, 1997). Its balance is maintained by histone acetyltransferases and histone deacetylases. Histone tails can also be methylated, a process that is catalyzed by histone methyltransferases and demethylases. Combinations of all these dynamic modifications can occur simultaneously at different sites, building up a specific "histone code" (Jenuwein and

Allis, 2001; Strahl and Allis, 2000). For instance, elevated acetylation and trimethylation of H3K4 or H3K36 are associated with active transcription. Conversely, low levels of acetylation and gain of methylation of H3K9 residues are usually linked with gene repression (Li et al., 2007). However, chromatin marks are also associated with other functions apart from transcription such as repression or activation of regulator elements (i.e. enhancers) or indicating transcription initiation or elongation. Based on this interplay of histone modifications, nine chromatin marks across nine cell types were used to define 15 chromatin states (Ernst et al., 2011), (**Figure 5**). In that study, it was observed that among different cell types, a high variation in chromatin states existed, representing the execution of necessary different gene expression programs for proper cell functioning.

A.

LAYER	FUNCTION
H3K4me3	Promoters
H3K4me1	Enhancers
H3K27ac	Active reg. element
H3K36me3	Txn. Elongation
H3K27me3	Polycomb repression
H3K9me3	Heterochromatin
Chr. accessibility	TF binding

B.

State	CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	Candidate state annotation
1	16	2	2	6	17	93	99	96	98	Active promoter
2	12	2	6	9	53	94	95	14	44	Weak promoter
3	13	72	0	9	48	78	49	1	10	Inactive/poised promoter
4	11	1	15	11	96	99	75	97	88	Strong enhancer
5	5	0	10	3	88	57	5	84	25	Strong enhancer
6	7	1	1	3	58	75	8	6	5	Weak/poised enhancer
7	2	1	2	1	56	3	0	6	2	Weak/poised enhancer
8	92	2	1	3	6	3	0	0	1	Insulator
9	5	0	43	43	37	11	2	9	4	Transcriptional transition
10	1	0	47	3	0	0	0	0	0	Transcriptional elongation
11	0	0	3	2	0	0	0	0	0	Weak transcribed
12	1	27	0	2	0	0	0	0	0	Polycomb repressed
13	0	0	0	0	0	0	0	0	0	Heterochrom;low signal
14	22	28	19	41	6	5	26	5	13	Repetitive/CNV
15	85	85	91	88	76	77	91	73	85	Repetitive/CNV

Chromatin states (y-axis label); Chromatin mark observation frequency (%) (x-axis label)

Figure 5. **Chromatin states definition. A.** Scheme of the association between epigenetic layer (histone modification or chromatin accessibility) and its attributed function. **B.** Emission table with the frequency of each chromatin marks used to define each chromatin state by a multivariate Hidden Markov model (HMM). Figure adapted from Ernst et al., 2011a.

Nucleosome positioning, the third kind of nucleosome remodeling, can be driven with specialized adenosine triphosphate (ATP)-dependent enzymes. All eukaryotes are equipped with at least five families of chromatin remodelers: SWI/SNF, ISWI, NURD/Mi-2/CHD, INO80 and SWR1. These enzymes directly interact with the DNA or chromatin structure, involving the movement of histone octamers, promoting for instance chromatin accessibility to upregulate gene expression or also leading to DNA

replication and repair (Becker and Hörz, 2002; Gangaraju and Bartholomew, 2007), (**Figure 4B**).

1.3.2 Methods for histone modification and nucleosome positioning detection

Histone modifications can be detected at a genome-wide scale by chromatin immunoprecipitation using antibodies (Abs) specific for particular PTMs followed by microarray hybridization (ChIP-Chip) (Huebert et al., 2006) or high-throughput sequencing (ChIP-seq) (Barski et al., 2007). Nowadays, the former technique has been replaced by the later due to limitations with probe design, cross-hybridization and background noise. Moreover, ChIP-seq results on profiles with higher resolution of the entire genome, and it requires less input material than ChIP-Chip. Variants of ChIP-seq have been developed to enable the analysis of small amounts of cells, such as iChIP (Lara-Astiaso et al., 2014), ChIPmentation (Schmidl et al., 2015) and the recently CUT&RUN (Skene and Henikoff, 2017).

As for the mapping of nucleosome positioning, that gives direct information about chromatin accessibility, one of the first method was micrococcal nuclease sequencing (MNAse-seq). This technique is based on an enzymatic digestion, which cleaves the naked DNA so that the undigested DNA is determined to be associated to nucleosomes. Likewise, this technique has been also adapted to capture the chromatin accessibility landscape, MACC (MNAse accessibility) (Mieczkowski et al., 2016). This layer can be also inferred with techniques such as DNAse I hypersentisitve site (DNase-seq) (Song and Crawford, 2010) and its faster and sensitive alternative named assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) (Buenrostro et al., 2013). ATAC-seq assesses regulatory landscape of chromatin on small sample size, i.e. 50,000 or less cells (Buenrostro et al., 2013; Klemm et al., 2019). Results from ChIP-seq and ATAC-seq have been used in the course of this doctoral thesis.

1.4 Nuclear folding architecture

The mammalian genome is folded at multiple levels in a non-random fashion to fit into the nucleus. Each level of the chromatin (that resembles a wool knot) highlights an important interplay between structure and function (Hansen et al., 2018; Ruiz-Velasco and Zaugg, 2017), (**Figure 6**).

The simplest chromatin structure seems to be similar to "beads on a string", with the "beads" being the individual nucleosomes and the "string" the nude DNA (Olins and Olins, 1974; Woodcock et al., 1976). However, this string needs to be further organized. Hence, the addition of the linker H1 binds to the nucleosome core particle around the DNA entry and exit sites, and wraps another 20bp, forming a structure called chromatosome (Fyodorov et al., 2018; Hergeth and Schneider, 2015). Chromatin can additionally be coiled into an even shorter, ticker fiber, called 30 nanometer (nm) fiber (Tremethick, 2007). However, studies in the field of imaging questioned the 30nm folding, highligthing the existence of a disordered and interdigitated state of compactness (Eltsov et al., 2008). For instance, at a 20nm resolution "nucleosome clutches" have been identified (Ricci et al., 2015) and even a flexible and disordered 5 to 24nm diameter chromatin chain has been recognized to be packed at different concentrations in interphase and mitotic chromosomes (Ou et al., 2017). In addition, higher-order loops and domains (topologically associating domains (TADs) and compartments) are organizing the chromatin to delineate clusters of chromosomal regions with similar biochemical and functional properties. At a largest scale, chromosomes are organized into chromosome territories (CTs) within the interphase nucleus while they become further compacted during mitosis.
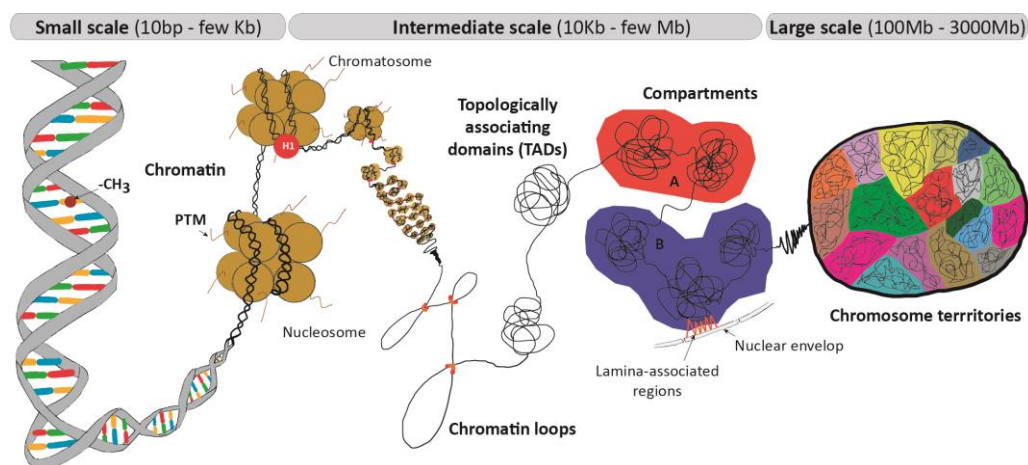
Figure 6. Genome organization. Overview of the different epigenetic layers and size scales. PTM, post-translational modifications. Figure adapted from Hansen et al., 2018.

Moreover, DNA packaging is dynamically adjustable based on cues from environmental stimuli and activated cellular pathways (Aguilar and Craighead, 2013;

Mohammad and Baylin, 2010). In this way, it is demarcated euchromatin as an open and DNA accessible chromatin for DNA-binding proteins and polymerases or close regions defining heterochromatin. Below, I will explain the different levels of genome organization in further detail and the methodologies used to study them.

1.4.1 3D genome structures within the interphase nucleus

The different hierarchical genomic scales associated with genome architecture (**Figure 6**) are not only playing a structural role to pack the DNA into the cell nucleus but also are key determinants of genome function.

1.4.1.1 Chromosome territories

At the turn of the twentieth century, the work of Carl Rabl (Rabl, 1885) and later of Theodor Boveri (Boveri, 1909) suggested that animal interphase chromosomes adopt a form of territorial organization where interchromosomal contacts are minimized. Boveri coined this organization as chromosomal territories (CT). These studies using light microscopy were contradicted when in 1950s, using electron microscopy, the nucleus was shown to be filled with intermingling chromatin fibers and loops with no sign of individual chromosomes, assuming a random organization. Finally, in 1980s, Thomas and Christoph Cremer seemed to finally resolve the debate between two theories for nucleus organization, i.e. chromosome territories versus random distribution of chromosomes (**Figure 7A**).
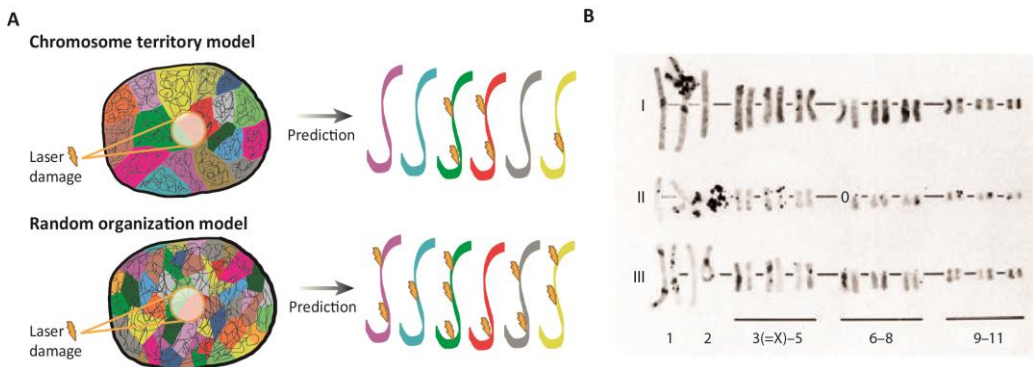
Figure 7. The discovery of chromosome territories (CT). A. Possible models of chromosome organization in the interphase nucleus and the predicted result after micro-irradiation (represented by lightnings). **B**. Three sets (I-III) of hamster chromosomes after laser damage.

The damage affected mainly chromosome 1 and 2 without a significant expansion through the rest of the chromosomes as indicated for the black grains of radioactivity. This result indirectly demonstrates the presence of CT. Figure adapted from Meaburn and Misteli, 2007.

The Cremer brothers performed a very elegant experiment in which they applied ultraviolet (UV) irradiation followed by pulse labeling on a specific site of interphase chromatin in the nucleus (**Figure 7A**). The so-induced local genome damage resulted on damage at only a subset of specific chromosomes, supporting the existence of CT (Cremer and Cremer, 2010; Cremer et al., 1982), (**Figure 7B**). Since then, numerous other microscopic studies have supported this structural organization. Nowadays, the function of CTs as first level of compartmentalization is recognized. The position of CTs has been shown to be cell type specific suggesting that the boundaries shared between a chromosome and its neighbors might be functionally relevant and contributting to particular gene expression program (Fraser et al., 2015a; Hepperger et al., 2008; Parada et al., 2004). These common regions between territories result on the establishment of interactions, a phenomenon termed as "chromosome kissing" (Cavalli, 2007) or "non-homologous chromosomal contacts" (NHCCs) (Maass et al., 2019). However, territories are dynamic in nature not only from cell to cell (Meaburn and Misteli, 2007), but also they present variations and changes upon development and differentiation (Rozwadowska et al., 2013), spermatogenesis (Foster et al., 2005) or DNA damage (Mehta et al., 2013). The tumorigenesis-related genome instability can also be associated with CT dynamics. In fact, a strong correlation between the frequency of chromosomal translocations and the spatial proximity among them has been shown (Branco and Pombo, 2006; Harewood et al., 2010), or a displacement of territories due to aneuploidy (Kemeny et al., 2018).

Apart from the boundaries between chromosomes, the relative position of chromosomes within the nucleus is also playing an important role in gene function. The intranuclear distribution is folding in a non-random radial organization, with gene-poor chromosomes oriented to the nuclear periphery and gene-rich chromosomes to the more internal nuclear regions, correlating with transcriptional activity, replication timing, and GC content (Federico et al., 2006; Goetze et al., 2007; Grasser et al., 2008; Hepperger et al., 2008; Mayer et al., 2005). However, particular exceptions to this general rule do exist, e.g. in nocturnal retina rod cells, where the heterochromatin

localizes at the center of the nucleus, and the euchromatin lines the nuclear periphery (Solovei et al., 2009).

1.4.1.2 Compartments

Each of the chromosomes can be divided into multimegabase-sized regions of similar epigenomic states which tend to contact each other, even across large distances, giving rise to the so-called compartments. In 2009, two sets of spatially segregated compartments (Lieberman-Aiden et al., 2009). These compartments were named "A" and "B" and are composed of largely "active" and "inactive" chromatin, respectively (**Figure 6**). The open A compartments include regions with high GC content that are enriched in genes, transcription activity, RNA polymerase II, high DNA accessibility, and histone modifications associated with active chromatin (H3K36me3) and poised chromatin (H3K27me3). In contrast, B compartments show higher interaction frequencies, a stronger tendency toward self-association, and high levels of the silencing H3K9me3, a heterochromatin histone mark (Lieberman-Aiden et al., 2009). The position of the compartments inside the nucleus is differential. The B or heterochromatic compartments are more positioned towards the nuclear envelop and are highly correlating with late replication timing, while the A or euchromatic compartments are located near nuclear speckles, away from the nuclear periphery, and highly correlating with early replication regions (Ryba et al., 2010; van Steensel and Belmont, 2017). The segregation of the genome into the two compartments has been further refined pointing to the existence of more compartments (Rao et al., 2014; Yaffe and Tanay, 2011). Specifically, one study segregates the genome onto three compartments, a cluster with gene-rich and active domains, another centromere-proximal cluster with non-active genome and a third centromere-distal cluster (Yaffe and Tanay, 2011). This study was followed by a higher-resolution study which revealed that functional compartmentalization of the genome does not stop at the previously described A and B compartments but that these two actually correlate with sub-compartments differing in replication timing and chromatin landscape (Rao et al., 2014). The active A compartment was associated with two sub-compartments, A1 and A2, which share highly expressed genes and active chromatin marks, such as H3K27ac and H3K36me3, slightly differ in terms of replication timing and also gene content, i.e. A2 has longer genes, lower GC content and completes replication later than A1. Loci of

the inactive B compartment may correlate to one of four sub-compartments. B1 to B3 sub-compartments are associated with different states of heterochromatin and distinct localization in the nucleus. B1 is enriched in H3K27me3, a polycomb chromatin mark, indicative of facultative heterochromatin while B2 and B3 are associated with constitutive heterochromatin. Interestingly, polycomb-associated B1 sub-compartment has been localized mainly between the two types of active sub-compartments and to lesser extent between B2 and B3 (Chen et al., 2018). B2 is enriched in pericentromeric heterochromatin and B3 is located at nuclear lamina. The B4 sub-compartment manually annotated is only present on chromosome 19 and spans 11 megabases (Mb). This special sub-compartment is enriched on KRAB-ZNF genes which exhibit a dual active and repressive chromatin pattern (Rao et al., 2014; Wang et al., 2019).

At a similar scale of tens of kilobases (Kb) to a few megabases, another organizational layer, different to compartments, has also been described, called the topologically associating domains.

1.4.1.3 Topologically Associating Domains

Topologically associating domains (TADs) are building blocks of the genome characterized by preferential 3D contacts between loci inside the same TAD as compared to loci in adjacent TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012a). Hence, fewer contacts are detected across their boundaries, which separates and delineates each TAD (**Figure 8**). Deletion of TAD boundaries results in fusion of domains and gene expression deregulation (Nora et al., 2012), (**Figure 8**). Boundaries have been found to be enriched in transcription start sites (TSSs), active transcription carrying the corresponding histone marks, housekeeping genes, tRNA genes, and short interspersed nuclear elements (SINEs) (Dixon et al., 2012; Nora et al., 2012; Phillips-Cremins et al., 2013; Sexton et al., 2012a). TAD boundaries are also enriched in binding sites for architectural proteins such as CCCTC-binding factor (CTCF) and cohesin.

CTCF is regarded as the "master weaver" of the genome. CTCF is an essential protein that is highly conserved from fly to humans with close to 100% homology between chicken, mouse, and human (Ohlsson et al., 2001; Phillips and Corces, 2009). CTCF serves as a functional regulator acting as a barrier or insulator between active and inactive chromatin, forming TAD borders, as well as preventing interactions between

enhancers and promoters (Mehra and Kalani, 2018; Merkenschlager and Nora, 2016). CTCF frequently colocalizes with cohesin, but not exclusively. Cohesin is a multisubunit protein complex composed of Smc1A, Smc3, Rad21, and Stag1/2 (SA1/2), which forms a ring-shaped ATPase in charge of topologically entrap DNA. It is implicated in sister chromatin cohesion, mitotic and meiotic chromosome segregation and DNA repair (Fraser et al., 2015a; McNairn and Gerton, 2008; Nasmyth and Haering, 2009; Peters et al., 2008). Cohesin has been described to be mainly involved in chromatin interactions within TADs, whereas CTCF is important for their spatial segregation (Zuin et al., 2014). Interestingly, although CTCF depletion clearly alters the insulation of TADs, it does not affect the genome compartmentalization into A and B compartments (Nora et al., 2017).
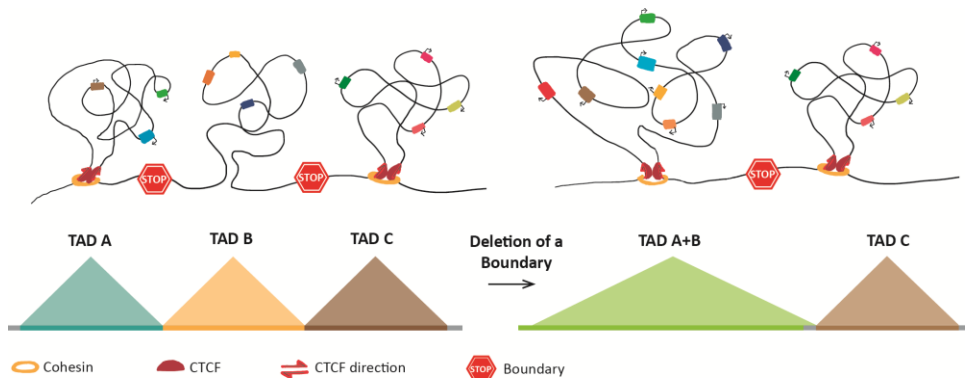
Figure 8. Topologically associating domains and their disruption. Representation of three TADs hosting several genes and regulatory regions creating each one a unique block (*left*). Deletion of a boundary between TAD A and TAD B leads to TAD fusion (*right*).

During mitosis, TADs dissolve and become re-established in G1 in the daughter cells. TADs therefore exist only during interphase. The initially identified TADs in 2012 had an average size of between 0.5Mb and 1Mb and were argued to have distributions highly conserved between cell types and species, suggesting that they are a stable feature of genome organization (Dixon et al., 2012; Sexton et al., 2012a). Some studies supporting this idea were reported upon stem cell differentiation (Dixon et al., 2015; Fraser et al., 2015b), cell reprogramming (Beagan et al., 2016; Krijger et al., 2016), or cytokine stimulation (Le Dily et al., 2014; Jin et al., 2013). TADs are further divided into submegabase-sized structures that are loosely referred to as "sub-TADs", which are

found to be more tissue specific (Phillips-Cremins et al., 2013). Thus, different length scales, TADs and sub-TADs, interact with each other to yield functional genome architectures.

TADs have also been implicated in replication timing (Pope et al., 2014), superenhancer-driven transcription (Dowen et al., 2014), double-strand break (DSB) synapsis during antibody class-switch recombination (CSR) (Dong et al., 2015; Zarrin et al., 2007) and normal limb development in human (Lupiáñez et al., 2015). They also play an important role in restricting recombination events during V(D)J recombination during B-cell maturation (Hu et al., 2015; Montefiori et al., 2016). Upon B-cell development, B cells rearrange their immunoglobulin genes to produce a specific immunoglobulin (Ig), which recognize a specific antigen (*explained in further detail in section 2.2: Normal B-cell differentiation*). Igs are composed of two light and heavy chains. The heavy chain (IGH) locus is further assembled by three segment V (variable), D (diversity) and J (joining) on a process called V(D)J recombination process. Interestingly, during this recombination, it has been described that the IGH locus forms one large compartment or TAD that is further divided into three sub-TADs, one containing the D_H and J_H segments and two dividing the V_H gene segment region (Montefiori et al., 2016), (**Figure 9**). Thus, genome architecture is critically involved in the development of the broad antibody repertoire (Jhunjhunwala et al., 2008; Johanson et al., 2019; Kumari and Sen, 2015; Lucas et al., 2014; Proudhon et al., 2015).
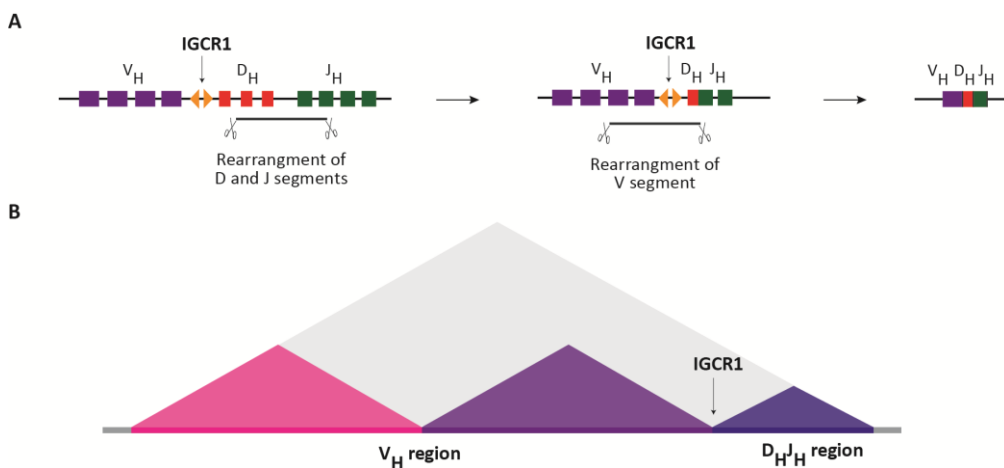
Figure 9. Overview of the immunoglobulin heavy chain (IGH). A. Diagram of the V(D)J recombination. D-J rearrangement precedes V-DJ recombination. **B.** TAD encompassing the IGH

locus and its three sub-TADs. IGCR1, intergenic control region 1. Figures adapted from Johanson et al., 2019; Kenter and Feeney, 2019.

In spite of the lines of evidence summarized above, the biological significance and robustness of TADs, and even their mere existence, are still debated because their detection is sensitive to the resolution of the Hi-C data and the algorithm employed (Rada-Iglesias et al., 2018).

1.4.1.4 Chromatin loops

Chromatin fibers can fold into "loops", bringing together in the 3D space two genomic loci that can be linearly distant. These long-distance interactions are the finest structures observed at sub-megabase scale.

Back in 1878, Walter Flemming first evidenced chromatin looping by reporting the existence of "strange and delicate structures" in the nucleus of amphibian oocytes (Callan, 1986). However, it was J. Ruckert, who concluded that those observations were looped chromosomes, calling them "lampbrush chromosomes" for their resemblance to the bristled brushed then used to clean the soot of oil-burning lamps. Nevertheless, 50 years later the work of Joseph Gall facilitated the discovery of DNA loops in human cells (Gall, 1956). Since then, many studies have been carried out leading to a division of long-distance interactions into (i) architectural or structural loops, which are long-distance interactions involved in the arrangement of stable domains, and (ii) regulatory or functional loops, implicated in regulating processes such as transcription. The regulatory loops are mainly defined as contacts between gene promoters and distant regulatory elements, particularly enhancers, leading to transcription activation (**Figure 10**). However, some studies on gene repression have shown that looping interactions can also be required to maintain silent chromatin states (Mehra and Kalani, 2018). For instance, polycomb group (PcG) of proteins maintaining looping interactions have been associated with *GATA-4* silencing in human cells (Tiwari et al., 2008). Moreover, the classification of chromatin loops can be refined when considering the time of the loop formation as (i) pre-established (permissive) or (ii) *de novo* established (instructive). Permissive loops are believed to facilitate a timely response to developmental stimuli. A proposed example is a conserved tissue-invariant configuration encompassing sonic hedgehog (*Shh*) gene, encoding a signaling protein involved in vertebrate development,

and its extensively characterized limb-bud enhancer laying nearly 1Mb away (Amano et al., 2009; Dixon et al., 2012). In the case of instructive loops, they are suggested to be particularly relevant if genes must be expressed at high levels in a particular cell type (Bouwman and de Laat, 2015; Denker and de Laat, 2016).
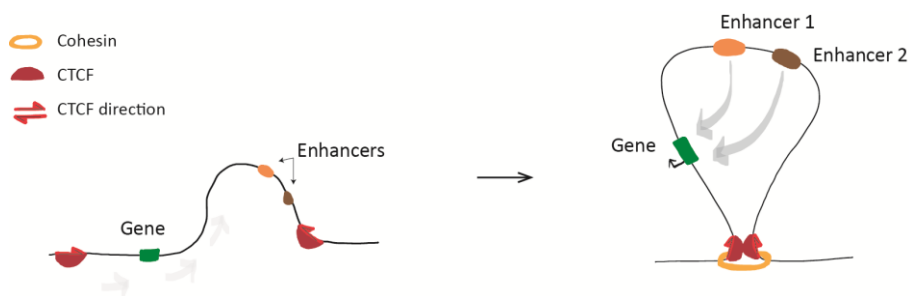
Figure 10. Chromatin loop formation. Example of chromatin loop with two distant enhancers associated with a gene.

1.4.2 Methods for detection of chromosome conformation

The 3D structure of the chromatin can be studied from two different and complementary perspectives, i.e. microscopy and genomic strategies/molecular biology techniques, known as chromosome conformation capture (3C)-based techniques. In the following sections, I will succinctly explain the methodologies most relevant for this doctoral thesis and I will clarify their advantages and disadvantages.

1.4.2.1 Visualizing Genome Organization

Subnuclear structures were first observed using light and electron microscopy-based techniques. *In situ* hybridization, and in particular fluorescence *in situ* hybridization (FISH) has been instrumental in identifying some of the key properties of genome organization. This molecular cytogenetics approach relies on the hybridization of labeled DNA probes complementary to specific regions of genomic DNA, which either are directly labeled with a fluorochrome (in the case of FISH) or indirectly by enzymatic or immunological detection (**Figure 11**). The visualization method depends on the label used, for instance, a fluorescently loci can be visualized using a fluorescence microscope (Volpi and Bridger, 2008). Diverse variations of the FISH method have been described, e.g. two-dimensional FISH (2D-FISH), 3D-FISH, and cryo-FISH, to tackle genomic structures from megabase scale (metaphase chromosome),

through submegabase (interphase chromosomes) to nucleotide scale (oligonucleotide arrays) (Speicher and Carter, 2005; Volpi and Bridger, 2008). Particularly, 2D-FISH has been applied to study nuclear location of genes or translocation from the nuclear periphery to the center and vice versa (Finlan et al., 2008). Though, for a proper colocalization of discrete genomic loci, such as promoter-enhancer interactions, the 3D reconstruction of nuclei is required. Hence, 3D FISH takes advantage of confocal microscopy or deconvolution software, with the capacity to generate image stacks through the z dimension (Speicher and Carter, 2005). Advances in the field as the combination approach of 3D FISH with live-cell imaging permitted to elucidate the chromatin topology dynamics upon cell processes. An example is the aforementioned study of the conformation changes in IGH topology during B-cell development (Jhunjhunwala et al., 2008; Lucas et al., 2014; Montefiori et al., 2016), (*see section 1.4.1.3 Topologically Associating Domains*). Finally, the cryo-FISH variation is used to determine the spatial intermingling of interphase chromosome territories (Branco and Pombo, 2006).
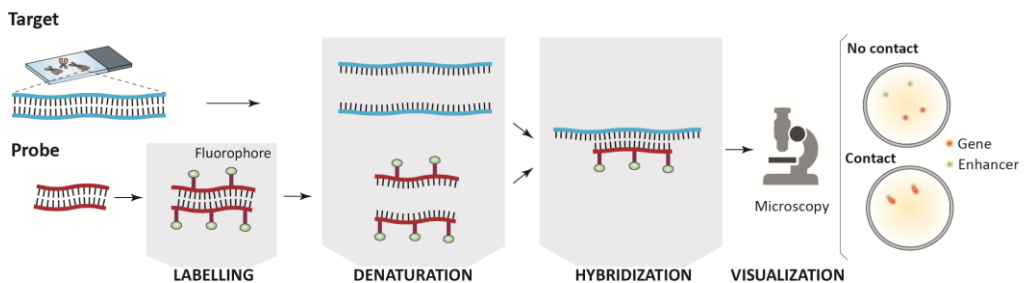
Figure 11. Principles of fluorescence *in situ* hybridization. Probe and target DNAs are submitted to a denaturation step in order to allow its hybridization. Labelling with different fluorochromes permits to easily detect 3D contacts, such as the colocalization between genes and their enhancers. Figure adapted from Speicher and Carter, 2005.

Nonetheless, some limiting factors belong to this type of methodologies. The first one is the sensitivity, defined as the capacity to capture light for a particular microscope, which is determined by the size of the probe (larger probes produce stronger signals). This brings to the second limitation that is resolution, defined as the capability to distinguish between two points along the length of a chromosome. To improve resolution, different probes have been developed such as fosmid probes,

which are used to measure large size distances between different regions, while for more close regions, oligonucleotide-base probes have been designed (Fraser et al., 2015a). Another limitation is the diffraction limit of the light microscopy, unabling to distinguish small length scales as 10-20nm. Some of these limitations are being solved by the development of super-resolution microscopy, such as structured illumination microscopy (SIM), stimulated emission depletion (STED) and photoactivation localization microscopy/stochastic optical reconstruction microscopy (PALM/STORM) (Lakadamyali and Cosma, 2015). Recent studies, which were using array-derived oligonucleotide (oligo) probes (oligopaint) with STORM have revelead cell-type-specific chromatin packaging, compartmentalization, and long-range *cis*-interactions with Kb- and nm-scale resolution (Bintu et al., 2018; Nir et al., 2018). These imaging data showed TAD-like structure in single cells (Bintu et al., 2018). Furthermore, a new approach improving resolution into tens of kilobases has been developed, named as ORCA, optical reconstruction of chromatin architecture, able to reconstruct the trajectory of a genomic region of interest (100-700Kb) (Mateo et al., 2019). Moreover, the combination with RNA labelling enables the identification of enhancer activity without prior knowledge.

Altogether, the power of FISH and other microscopy-based methods lies in the ability to perform single-cell analyses of gene positioning, usually in an affordable fashion (de Wit and de Laat, 2012a). However, they are still limited to a small number of selected loci. To address this limitation, molecular approaches have been described to characterize the physical proximity of chromatin based on the interaction frequency.

1.4.2.2 Chromosome Conformation Capture techniques

The first chromosome conformation capture (3C) methodology was developed to reflect how chromatin is organized in the nucleus. This molecular technology was based on nuclear ligation assay (Cullen et al., 1993) to analyze the frequency of contacts between selected genomic sites over the cell population at higher resolution as compared to most of the visual techniques. Since this first approach, several derived technologies have been described to quantify chromatin contacts at high-resolution, sheding light into 3D chromatin architecture at a whole-genome scale. All 3C-based protocols share five experimental steps (**Figure 12**).
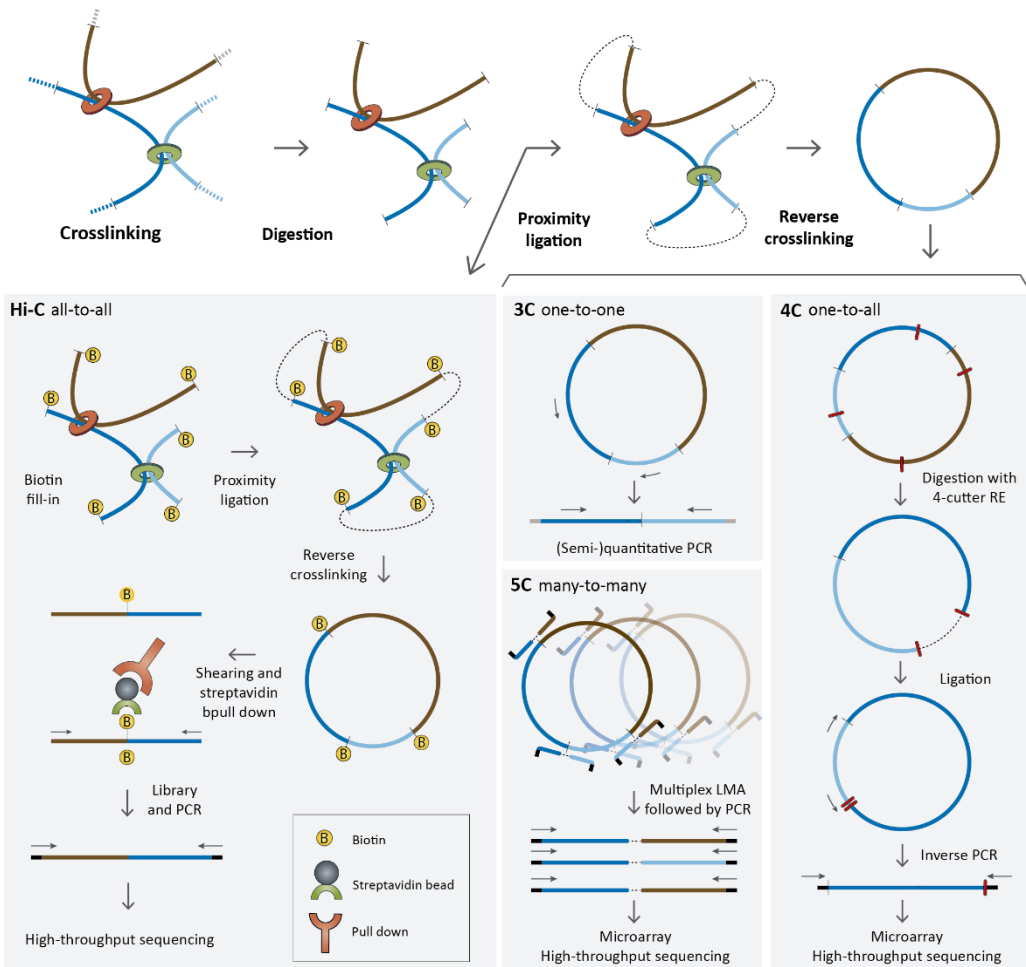
Figure 12. Overview of the 3C-derived methods. The first part (*upper* panel) shows the common steps to convert chromatin interactions into ligation products. Depending on the detection method of those ligation products, different 3C variants may be distinguish such as 4C, 5C and Hi-C. LMA, ligation-mediated amplification. Figure adapted from Bonev and Cavalli, 2016.

First, cells are crosslinked, in order to fix the chromatin proteins to their associated DNA and so maintain the chromatin structure. The most frequently used fixative is formaldehyde. Second, the crosslinked chromatin is fragmented, usually involving restriction enzymes. The type of enzyme selected defines the resolution of the 3C experiment. The enzymes recognizing 4bp sequences are going to cleave the chromatin more often resulting with smaller fragments (around 256bp) leading to higher resolution libraries, than 6-cutter which leads to 4096bp fragments. Third, DNA ends are ligated ensuring preferential ligations between contacting and crosslinked

chromatin fragments that are proximal at 3D space. Fourth, the crosslinking is reversed to obtain linear and circular DNA concatemers named as 3C templates that contain two genomic fragments of different linear locations but that colocalize at the 3D level. Finally, the templates are identified and quantified by PCR or sequencing technologies. Depending on the strategy to detect and quantify ligation junctions, different 3C-based methods can be identified, that I will briefly describe below.

In 2002, Job Dekker and colleagues described the 3C method, also known as the one-to-one approach. In this first approach, contacts were analyzed between selected pairs of sequences (i.e. targets of interest), and the frequency of ligation assessed using semiquantitative (Dekker et al., 2002a; Simonis et al., 2007a) or quantitative PCR (Hagege et al., 2007; Splinter et al., 2006; Würtele and Chartrand, 2006). Since then, 3C has been applied to numerous studies. However, for sites of interest separated by more than a few hundred kilobases, specific ligation products resulting from 3C protocol become infrequent and cannot be quantified by this method. To overcome this, 3C approach has been adapted for a genome-wide scale, resulting in unbiased methods for long-range DNA-DNA contacts (Denker and de Laat, 2016; de Wit and de Laat, 2012b).

Chromosome conformation capture-on-chip (4C) technology also known as "one-to-all" appear on scene improving the resolution of 3C "one-to-one" by identifying all contacts between a selected genomic site or viewpoint and the rest of the genome The 4C technology follows the initial steps of the prior technique. However, once the 3C templates are obtained, in 4C, a second round of digestion and ligation step results in small DNA circles, of which some contain the viewpoint plus contacting sequences. Then, a reverse PCR strategy using primers designed outwards on either side of the viewpoint fragment are employed to amplify the contacting sequences (**Figure 12**). Originally, 4C was combined with microarrays containing genome-wide probes to analyze the contacts of a selected viewpoint with all the genomic fragments that are represented on the specific array (Simonis et al., 2006). At the very same time, a slightly different protocol using custom arrays was published named as circular chromosome conformation capture (Zhao et al., 2006). Still, the scale and sensitivity of the former 4C assay was increased by combination with high throughput sequencing (4C-seq) generating high-resolution interaction profiles (van de Werken et al., 2012a). Some

variations of the protocol have been developed as the capture of specific interactions mediated by specific proteins accomplished by adding an immunoprecipitation step before the first ligation (Schoenfelder et al., 2010; Sexton et al., 2012b).

The 4C approach allows the generation of high-resolution interactions profiles, but some considerations must be taken into account. First, 4C have a lower coverage in *trans* interactions making it more challenging to find inter-chromosomal interactions. Second limitation is the bias introduced by the efficiency of the PCR amplification step due to properties of the fragments including differences in size or GC content, which leads to the limitation of quantifying contact frequencies. Recently, some tools have been described to improve the analyses although inherent biases of the approach remain (Denker and de Laat, 2016; Raviram et al., 2014; van de Werken et al., 2012b).

A step further to determine interactions between multiple selected sequences has been developed and known as the "many-to-many" approach or chromosome conformation capture carbon copy (5C) technology (Dostie et al., 2006). This methodology scales up the 3C approach by adding a mix of oligonucleotides, each of which partially overlaps a different restriction enzyme site in the genomic region of interest. The forward and reverse primers located next to each other across the 3C junction are ligated together generating a carbon copy of the ligation junction (**Figure 12**). Then, 5C libraries are analyzed by microarrays or high-throughput sequencing (Ferraiuolo et al., 2012). This variant increases in throughput, interrogating various regions as compared to 4C, and also reduces bias in PCR amplification efficiency between pairs of sites from 3C. However, it is not able to reach 4C's resolution as not every restriction enzyme-end can be targeted in the oligonucleotide design. The 5C variant has been postulated as reproducible and quantifiable technique (Ferraiuolo et al., 2012) and has paved the way onto the conformation reconstruction of entire domains or chromosomes.

Finally, to obtain a global view of chromatin structure, the Hi-C method was developed in 2009 to infer the 3D architecture of the whole genome by sequencing ligation products (Lieberman-Aiden et al., 2009). The method to create the template differs from the already presented 3C variants. Remarkably, on Hi-C a filling with biotin-labeled nucleotides of digested overhangs ensures enrichment of ligation junctions (**Figure 12**). Data is computationally processed by mapping the reads to the reference

genome and filtering the non-informative or error fragments as un-ligated, self-ligated and PCR artefacts, among others, to end up with valid reads. Then, the genome is divided into non-overlapping bins, each valid pair is assigned to a specific bin pair, and the aggregation lead to a contact matrix. In this way, the rows and columns of the contact matrix represent bins across the genome. Their sizes are based on sequence depth and library quality and are determining the resolution of the matrix. In the first application of Hi-C, around 10 million (M) paired-ends were sequenced to end up with 1Mb resolution contact matrices (Lieberman-Aiden et al., 2009). The raw matrix can be normalized by two different main models. An explicit method which considers that systemic biases as the GC content, mappability, and the frequency of restriction sites are known, whereas implicit method (or so-called balancing method) assumes that each locus throughout the genome has equal likelihood of being engaged in a 3D contact, called as equal visibility assumption (Lyu et al., 2019; Schmitt et al., 2016a). Finally, a contact heatmap can be generated and inferred (**Figure 13**).
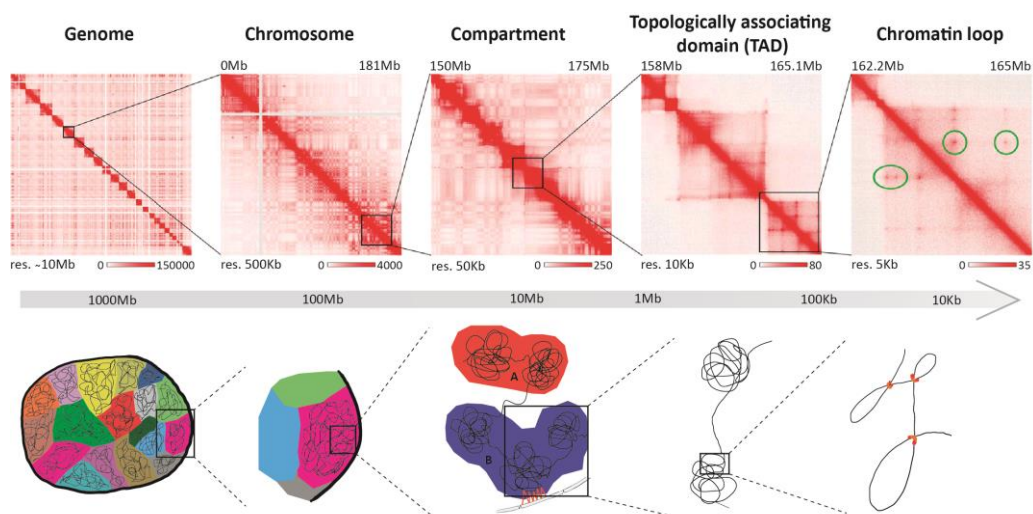
Figure 13. Hierarchical genome organization. Hi-C contact matrices for different scales: whole genome, whole chromosome, megabase and hundred kilobases (*upper* panel) and a model of genome folding at these scales (*lower* panel). At the lowest resolution, whole genome contact matrices shows that chromosomes occupy separate chromosomal territories. Moving into a megabase-level, compartments (described at the Hi-C matrices as a plaid or checkerboard-like pattern) and topological domains can be distinguished, finishing with chromatin loops corresponding to individual peaks on higher resolution Hi-C contact matrices. Color code of Hi-C heatmaps indicates number of contacts between a pair of loci. Figure adapted from Szalaj and Plewczynski, 2018.

However, one disadvantage of the Hi-C technique is the high number of cells necessary to generate biologically meaningful profiles. Of mayor interest was a publication describing a combined approach of the original HiC-seq protocol with a nuclear ligation assay named *in situ* HiC-seq, a protocol to generate much denser HiC-seq maps (Rao et al., 2014). The authors reported a Hi-C map with over 5 billion distinct contacts on a resolution of 950bp. Using these maps, several domain structures, compartmentalization, and thousands of chromatin loops can be clearly discerned.

Hi-C identifies contacts genome-wide, but it cannot inform about the nature or the function of these interactions. A technique combining maps of chromatin networks associated with specific protein was developed as ChIA-PET. This technique relies on pulling down DNA sites bound via a specific factor (Fullwood et al., 2009). However, ChIA-PET requires hundreds of millions of cells per experiment and results in a small fraction of informative reads for a given sequencing depth. To overcome this limitation, HiChIP was developed as combined protocol based on *in situ* Hi-C and ChIP with sequencing library preparation by Tn5 transposase (Mumbach et al., 2016). To assess specific proteins and 3D interactions, a similar method to HiChIP is PLAC-ChIP, proximity ligation-assisted ChIP-seq, which improves the efficiency and accuracy over ChIA-PET (Fang et al., 2016). Many other modifications and techniques have been described since the first seed of the genome-wide chromosome conformation approach, including alteration of labeling of digested overhangs, restriction enzymes-independent variants or even ligation independent methods (Beagrie et al., 2017; Liang et al., 2017; Ma et al., 2015; Quinodoz et al., 2018). In addition, to analyze contacts of targeted or custom-designed regions of the genome, the capture Hi-C methodology was developed (Dryden et al., 2014). Finally, technical modifications have been performed to reach a protocol to identify long-range contacts in single cells (Nagano et al., 2013; Ramani et al., 2017; Stevens et al., 2017).

Taking all together, current chromosome conformation capture approaches have varied substantially with respect to template 3C, or the genome-wide Hi-C library preparation procedure, target selection, capture probe design and target enrichment protocol, in order to optimize the method to distinct study designs.

2. B-cell development

2.1 Hematopoiesis

The defense system of our body is designed to fight against infectious microorganisms that may trigger a disease. Our army is the immune system, an interactive network of lymphoid organs, cells, humoral factors and cytokines (Parkin and Cohen, 2001). An army equipped to destroy a broad range of microbial cells and clear both toxic and allergenic substances, providing host protection. Therefore, host defense must be able to discriminate between self and non-self as damaging self-tissue leads to autoimmune disease. Besides the defense function, the system must avoid the elimination of beneficial commensal microbes to support normal tissue and organ function. And importantly, it has to be dynamically evolving to resist pathogens, as host and pathogen are continuously co-evolving in a process called "host-pathogen arms race" (Decaestecker et al., 2007; Woolhouse et al., 2002). Even though the primarily function of the immune system is the host defense, non-defense functions have been described implicating the immune system in processes such as reproduction, embryonic development, angiogenesis and post-injury repair and regeneration (Sattler, 2017).

Immunity can be understood in military terms as an operational response from a complex brigade formed by two distinguished battalions composed from different soldiers developed upon the hematopoietic system. The entire mammalian blood system is built upon hematopoiesis, a hierarchical developmental stages with the potential to continuously generate more than ten distinct mature cell types (soldiers) (**Figure 14**). All cells are generated from a single primary source, the hematopoietic stem cell (HSC) (Chao et al., 2008). HSCs have the ability to differentiate into all functional blood cells, a feature known as multipotency capacity and are also capable to self-renew in the absence of differentiation (Chen, 2011; Dykstra and De Haan, 2008). These two properties must be balanced to maintain the HSC pool size through life. Upon differentiation, HSC initially give rise to multipotent progenitors (MPPs), which lose the self-renew capacity of the HSC but still maintain the full-lineage differentiation potential (Christensen and Weissman, 2001; Morrison and Weissman, 1994). These MPPs can differentiate into two oligopotent progenitors, the common myeloid progenitors (CMPs) and the lymphoid-primed multipotent progenitor cells

(LMPPs). CMP differentiates to megakaryocyte-erythrocyte progenitor (MEP) and granulocyte-monocyte progenitor (GMP). The former give rise to erythrocyte (EP) and megakaryocyte (MkP) progenitors. GMP differentiates to macrophage (MacP) and granulocyte (GP) progenitors, which finally differentiates to monocytes or macrophages and granulocytes (i.e. neutrophils, eosinophils or basophils), respectively. LMPPs give rise to common lymphoid progenitor (CLP) which further differentiates into T and B lymphocytes and natural killer (NK) cells (Akashi et al., 2000; Luc et al., 2008; Orkin and Zon, 2008; Zandi et al., 2010), (**Figure 14**).
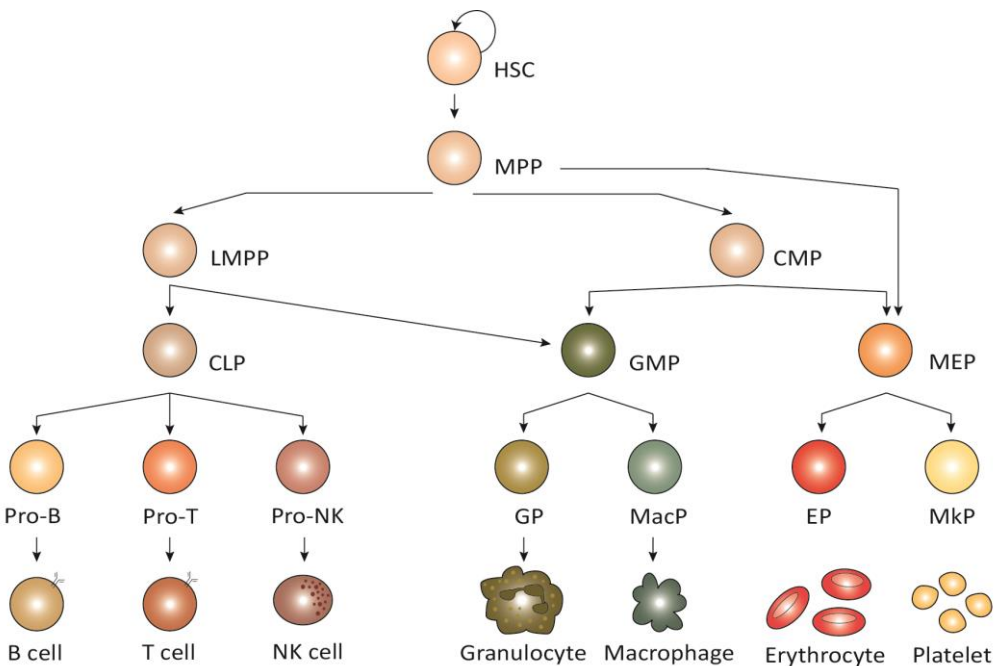
Figure 14. Hematopoiesis. Schematic representation of the hematopoiesis system differentiation. Figure adapted from Cedar and Bergman, 2011.

NK cells and the myeloid derived cells with complement (serum glycoproteins), cytokines, acute phase proteins as well as physical and chemical barriers constitute the first combat unit known as innate immune response. This operation acts rapidly after the attack because the recognition molecules used by the innate battalion are expressed broadly on a large number of aforementioned cells. For instance, neutrophils produce large quantities of reactive oxygen species that are cytotoxic to bacterial pathogens. These cells accumulate in large quantities at sites of bacterial infection and

tissue injury to phagocyte microbes and particulate antigens internally, where they can be destroyed and degraded (Witko-Sarsat et al., 2000). Like neutrophils, monocytes and macrophages are also highly phagocytic (Parkin and Cohen, 2001). This immediate but unspecific response can trigger the adaptive immune response, the second combat unit. These two battalions are determined by the speed and specificity of the battle injury or reaction. So, the adaptive immune response is a battalion composed of small number of long-lived cells with specificity for any individual pathogen, toxin, or allergen. These cells are B and T lymphocytes, both developed from CLPs within the bone marrow. B cells remain in this anatomic site for their further development while T cells migrate to the thymus. Those lymphocyte subsets can be discriminated by the expression of surface markers. Both cells accomplish a specific response mediated by the expression of antigen-specific receptors on the surface. The T-cell receptor (TCR) and B-cell receptor (BCR) are encoded by assembling of somatic rearranged germline gene elements. In the lymphoid tissue (lymph nodes, spleen, tonsils, and mucosa associated lymphoid tissue) the antigen is presented to and recognized by the antigen specific T or B cell which leads to cell priming (first recognition), activation and differentiation. Upon effector response, T cells leave the lymphoid tissue and migrate to the disease site. T cells are defined by their cell-surface expression of the TCR, a transmembrane heterodimeric protein that binds processed antigen displayed by antigen-presenting cells (APCs) (Kumar et al., 2018). Two main different functional T cells have been described. The T cytotoxic (Tc), which directly reacts on their specific antigen carrying cells, and T helper (Th), which recognizes the foreign antigen, and activates cell-mediated immune response as the B cells response (Kulinski et al., 2013; Tangye et al., 2013). A response mediated by terminally differentiated activated B cells (called plasma cells, PCs) leads to the production of specific antibodies into blood and tissue fluids directed to the infective focus. In fact, subsets of B cells have been defined that differ in the types of antigen to which they respond and in the type of antibody they produce. Those cells persist in an apparently dormant state but can re-express effector functions rapidly upon another encounter with their specific antigen (Akkaya et al., 2019; Farber et al., 2014). The responding cells must proliferate after encountering the antigen to attain sufficient numbers to mount an effective response. This battalion acts temporally after the innate response in host defence. Besides, this second battalion provides the brigade ability to manifest immune memory, permitting

it to contribute prominently to a more effective host response against specific pathogens or toxins when they are encountered for a second time, even decades after the initial sensitizing encounter (Chaplin, 2010; Parkin and Cohen, 2001).

The complex process of hematopoietic cell lineage development is constantly being redefined by the discovery of additional intermediate developmental stages, and the progress in their characterization on molecular level. Given the importance of B cells for this doctoral thesis, in the following section I will explain in detail the mechanisms underlying B-cell differentiation.

2.2 Normal B-cell differentiation

The MPPs migrate first into fetal liver and later into bone marrow, medullary cavities of the bones, where they can develop the B-cell lineage pathway (Melchers, 2015). Progenitor cells commit to the B-cell lineage to give rise the precursor B cells due to the expression of lineage specific transcription factors (TFs) such as EBF1 (Early B-cell Factor 1), PAX5 (Paired-box protein 5) and E2A (also known as TCF3) (Hagman and Lukin, 2006; Matthias and Rolink, 2005), (**Figure 15**).
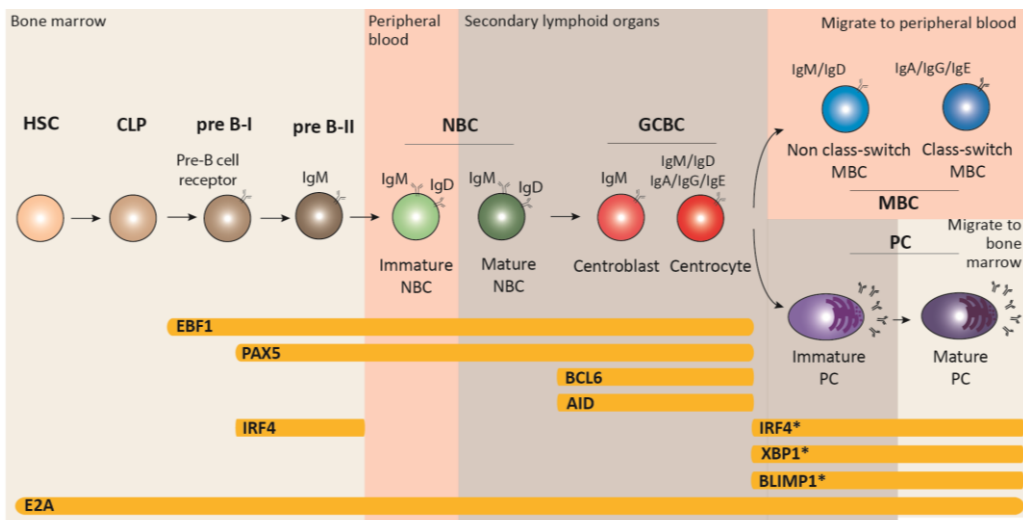
Figure 15. **Graphical scheme of B-cell development.** Representation of the different B-cell subpopulations, their location and expression patterns of associated transcription factors (TFs). The TFs marked with (*) are plasma cell specific. Figure adapted from Chaplin, 2010; Martin-Subero and Oakes, 2018; Melchers, 2015.

Those precursor B cells expressing CD19 gradually rearrange their immunoglobulin (IG) genes in order to counter infections. Ig encompase the BCR together with a signal transduction heterodimer called Igα/Igβ (CD79). In detail, Ig molecules consist of 2 identical 50-kiloDalton (kDa) heavy (IGH) chains and 2 identical 25-kDa κ (IGK) or γ (IGL) light chains (**Figure 16**). The amino terminal portions of the heavy and light chains vary in amino acid sequence from one antibody molecule to another. These variable portions are designated as V_H and V_κ or V_γ, and create the antigen-binding portion of the intact Ig molecule. Also the variable regions contain highly variable subregions forming the antigen-binding domain of the molecule. The heavy chain is composed by the rearrangement of three segments the V (variable), D (diversity) and J (joining) (**Figure 16A** and **9**), while the light chain involves tow segments V and J (**Figure 16B**). On the other end of the sequence, the carboxyl terminal is maintained constant (C) in each subclass of antibody (Chaplin, 2010). The assembling of these genetic building blocks is due to the recombination-activating genes (*RAG*). *RAG1* and *RAG2* are only expressed during times of heavy and light chain rearrangement, resulting in the assembly of the antigen-binding component of Ig (**Figure 16C**).
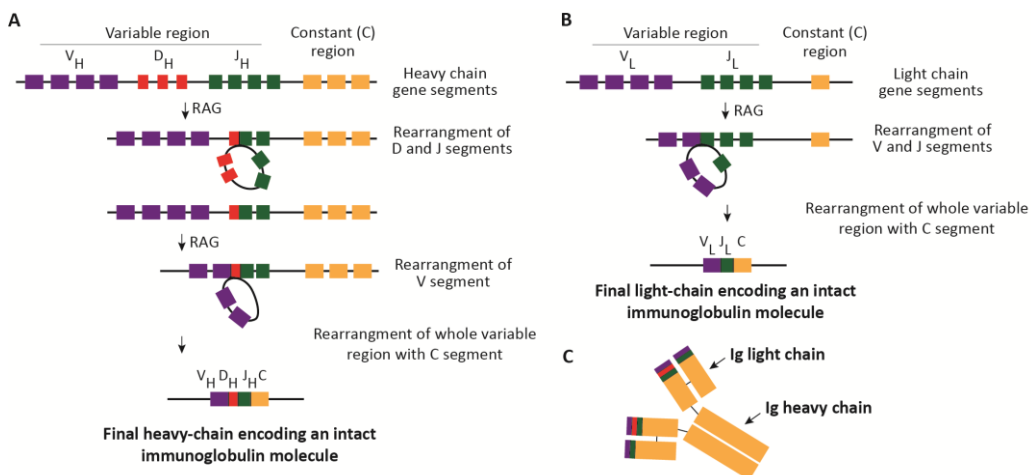
Figure 16. V(D)J recombination leads to the formation of immunoglobulins. A and **B.** Schematic representation of immunoglobulin heavy (**A**) and light (**B**) chain recombination process. The variable region is genereated by random recombination of its respective sequences. **C.** The immunoglobulin (Ig) molecule is composed of two identical light and heavy chains. RAG, Recombination Activating Genes.

Developing B cells follow a program of differential surface antigen expression and sequential heavy and light chain gene rearrangement. Three stages of early B cells can be distinguished in the bone marrow (BM). First, pro-B cells undergo the V(D)J rearrangements expressing the immunoglobulin heavy chain with a surrogated light chain giving rise to a pre-B cell receptor (pre-BCR) on the formed pre-B cells. Later, the rearranged heavy chain i.e. expressing the constant region of the µ chain pairs with an immunoglobulin light chain producing IgM, which is expressed on the cell surface and displayed as a BCR on immature B cells. However, difference in heavy chain constant region defines other isotypes (reaching the five major ones). If splicing takes places on the δ chain then IgD is generated. Each B cell expresses only one specific BCR. For instance, once the rearrangement of the H-chain is performed, a subsequent shut down of the activities and expression of the enzyme machinery is induced in a process termed allelic exclusion. In this way, the cell prevents the expression of two H-chains with two different specificities on the same cell. All the process is controlled by several checkpoints in order to discard if the individual chains or the BCR presents autoreactivity. BCRs expressing high-affinity to autoreactive are further rearranged and if in a second attempt those immature B cells still react to self-antigens they undergo apoptosis and are removed (Chaplin, 2010; Melchers, 2015). Over 85% of the newly formed immature B cells dies in BM, probably as a consequence of this autoantigen recognition (Melchers, 2015). The immature B cells mature into transitional B cells, which leave the BM and fully mature into naive B cells (NBCs) in the peripheral blood (PB). In the BM, a repertoire of immature B cells capable to recognize more than $5x10^{13}$ different antigens is formed (Pieper et al., 2013).

Circulating NBCs express IgM and IgD on their surfaces. These mature B cells circulate through the blood to reach secondary lymphoid organs such as spleen, lymph nodes, tonsils, Peyer patches, and mucosal tissues. There, two different maturation pathways can be followed related to two distinct regions of lymph nodes. One involves the marginal zone (MZ), which plays vital functions in T-cell independent humoral immune responses against blood-borne pathogens. Another involves the follicular (FO) region, where follicular B cells can capture antigen presented by follicular dendritic cells (FDCs) and T cells that are located around the B-cell zone of the developing germinal center (GC). In this context, upregulation of the BCL6 TF is essential for the GC formation. A subset of T cells (Th) presents the antigen to the GC founder B cells, which

form the dark zone (DZ) of the GC as centroblasts. These centroblasts are highly proliferative and undergo through a process called somatic hypermutation (SHM) by which the variable region of the IG genes becomes somatically mutated mediated by the action of activation-induced cytidine deaminase (*AICDA* locus encoding for the AID enzyme). This process results with clones of different high affinity antibodies for antigens (**Figure 17**). Besides, another molecular process takes places during the GC reaction at the constant-region portion of the heavy chain of IgM or IgD, resulting in the exchange to either IgA, IgG or IgE. The switch of the Ig isotype is a process called class-switch recombination (CSR) (De Silva and Klein, 2015a). CSR targets the constant-region portion of the heavy chain of IgM or IgD resulting in the exchange to either IgA, IgG or IgE. In this way, different effector functions can be propagated without changing the Ab specificity associated with the rearranged variable region. Then, centroblasts move to the adjacent region called the light zone (LZ) giving rise to centrocytes. These germinal center B cells (GCBCs) express their Ab on the cell surface and are selected based on the affinity of the Ab for the antigen, positively or negatively depending on their increased or decreased affinity of their BCR for the antigen, respectively. Thus, the centrocyte can be eliminated or rescued and sent back to the DZ to generate Ab variants with higher affinity. Finally, after some rounds of proliferation, mutation and selection, the differentiation of the centrocytes expressing high affinity antibody is initiated. These centrocytes can exit the GC as memory B cells (MBCs) or plasma cells (PCs), both crucial for mediating humoral immune response (**Figure 17**).
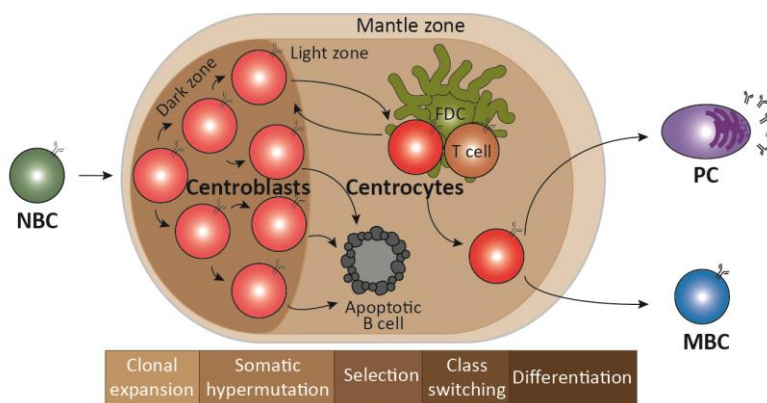
Figure 17. Representation of the germinal center (GC) reaction. A stepwise process is done in the GC to give rise to a B cell with an increased affinity for an antigen BCR. FDC, Follicular dendritic cell. Figure adapted from Küppers, 2005.

MBCs, which express CD27, (Fecteau et al., 2006), can be further divided into two main groups: non-class switch memory (ncsMBC), carrying IgG or IgM, and class-switch memory (csMBC), carrying IgG, IgA or IgE isotypes. These MBCs live for extended periods of time circulating through the PB and on a future encounter with the antigen they reactivate and proliferate, being the csMBCs able to generate high-affinity PCs and the ncsMBCs to promote secondary GC reactions (McHeyzer-Williams et al., 2012). A third memory B cell subset has been described independent of the GC as CD27-IgA+ coming from the gut. The replication of this subtype has been observed to be limited. Apart from MBCs, GCBCs may also differentiatie into PCs, upon the upregulation of TFs such as IRF4 and BLIMP1 and downregualtion of PAX5. PCs produce high amounts of antibodies and can be distinguished by the expression of CD38 and CD138 on their surface as well as by the loss common B-cell markers such as CD19 and CD20. This finally differentiated B-cell subpopulation resides in BM where it may persists for long periods of time (years), (**Figure 17**).

All these B-cell subpopulations can be easily distinguished as they present distinct phenotypic and transcriptional features, which results on their easy isolation by sorting from BM, PB or lymphoid tissues of healthy donors, using specific surface markers for each cell type, representing the perfect model to investigate the modulation of the epigenome upon maturation.

2.2.1 Epigenetic modulation during B-cell differentiation

B-cell differentiation genes and processes have been described to be epigenetically regulated (Martin-Subero and Oakes, 2018). On one side, WGBS studies have shown that the DNA methylation landscape is extensively modified from uncommitted progenitors to long-lived BM PCs affecting millions of CpGs sites, reaching 30% of the entire DNA methylome (Kulis et al., 2015; Oakes et al., 2016). A massive reconfiguration mostly targeting heterochromatin and nuclear lamina associated domains, DNA repeats and polycomb (Pc)-repressed regions. Particularly, late differentiation stages showed an extensive demethylation of heterchromatin and a methylation gain at Pc-repressed areas (Kulis et al., 2015). On the other side, an inverse correlation between the expression of TFs and the average of DNA methylation levels has been accompanying the B-cell commitment, suggesting that TFs play an important role in shaping the DNA methylome as a whole. Of interest is the enhancer

demethylation involved in leukocyte activation and B-cell signaling from uncommitted progenitors to early B cells. Those regions are binding sites of B cell-specific TFs such as EBF1, PAX5, E2F and BATF (Almamun et al., 2014; Kulis et al., 2015; Lee et al., 2012). The regulation of those TFs has been thoroughly analyzed. For example, *EBF1* was described to be regulated by ETS1, PAX5, PU.1, and the RUNX1/CBF-β complex, and epigenetic changes mediated by nucleosome remodeling and histone modifications (Choi et al., 2012; Jiang et al., 2011; Roessler et al., 2007; Seo et al., 2012). The *EBF1* locus is positioned at a heterochromatin nuclear lamina in progenitor cells whereas in pro-B cells was described to switch compartment and established new 3D interactions (Lin et al., 2012).

2.3 B-cell neoplasms

Hematological neoplasms (HN) comprise a large number of entities with different biological and clinical features that affect PB, BM and secondary lymphoid organs. The World Health Organization (WHO) classify these diseases based on all the information available on morphology, immunophenotype, genetic alterations and clinical features (Swerdlow et al., 2017). The main principle to categorize HNs is that they originate from different hematopoietic cell lineages at distinct maturation stages. Thus, neoplasms primarily are categorized according to lineage of origin i.e. myeloid, lymphoid or histocytic/dendritic. The normal counterpart is then postulated per each neoplasm. For instance, lymphoid neoplasms encompassing B-cell and T/NK-cell neoplasms arise from mature and immature B, T or NK cells. Thus, precursor B cells give rise to B lymphoblastic leukemia while mature B-cell neoplasms and Hodgkin lymphoma are developed from mature B cells. Lymphoid neoplasms are classified into two major groups, Hodgkin and non-Hodgkin lymphomas. B-cell non-Hodgkin lymphomas are branched into multiple entities with different clinical course, symptomatology, molecular features and treatment (Morton et al., 2007).

It is broadly accepted that malignant transformation of B cells is initiated by genetic abnormalities taking place during the normal B-cell differentiation process. Chromosomal translocations have been described as a result of V(D)J recombination process (Küppers and Dalla-Favera, 2001). Upon BCR formation, RAG proteins mediates DNA strand breaks within the IG loci (i.e. IGH in 14q32; immunoglobulin lambda locus, IGL in 22q11 and immunoglobulin kappa locus, IGK in 2p11). This physiological process

can lead to illegitimate rearrangements resulting in translocation of IG genes to oncogenes located in distinct regions of the genome (Willis and Dyer, 2000). Some examples involving the IGH rearrangements are the translocation linking chromosome 11q23 and 14q32, t(11;14)(q13;q32), in mantle cell lymphoma (MCL) upregulating cyclin D1 (CCND1), while on follicular lymphoma de t(14;18)(q32;q21) leads to overexpression of the anti-apoptotic *BCL-2* gene (Shaffer et al., 2002; Swerdlow et al., 2017).

Somatic hypermutation (SHM) and class switch recombination (CSR), that take place in GCBC, are further sources of B-cell malignancy development (Küppers and Dalla-Favera, 2001). The translocation t(8;14)(q24;q32) involving *c-MYC* in most of sporadic Burkitt's lymphoma (BL) cases is suggested to be derived from an error in CSR. This is demonstrated by locating the breakpoint in IGH switch regions (Dalla-Favera et al., 1983; Gelmann et al., 1983; Showe et al., 1985; Taub et al., 1982). Moreover, CSR has been also involved in insertion events. An example is the insertion of the IGH enhancer nearby the *CCND1* gene during switch recombination process, which results in cyclin D1 overexpression in rare multiple myeloma cases (Gabrea et al., 1999) or MCL cases (Fuster et al., 2019). SHM, on its part, can result in aberrant mutations of genes as *BCL6, PAX5 or MYC* in diffuse large B-cell lymphoma (DLBCL) (Migliazza et al., 1995; Pasqualucci et al., 2001).

Some of the translocations are used in the clinics for differential diagnostics of B-cell lymphomas. However, these aberrant events are not sufficient to induce lymphomagenesis. In fact, some of the translocations have been detected on small clonal expansions in healthy individual (Lecluse et al., 2009; Limpens et al., 1991; Müller et al., 1995; Roschke et al., 1997; Summers et al., 2001) and transgenic mouse models show that more than one translocation is needed for malignant transformation (Strasser et al., 1990). Therefore, it seems clear that secondary genetic abnormalities must be acquired for lymphomagenesis.

Once an initial oncogenic event takes places, B cells may be arrested at the maturation state of the oncogenic clone or continue the maturation process and become blocked in a more advanced stage (Shaffer et al., 2002). In fact, mature lymphoid neoplasms are generally categorized according to their clinical presentation and to some extent the stage of cell differentiation when this can be postuladed

(Swerdlow et al., 2017), (**Figure 18**). In this context, a deep molecular characterization by means of e.g. transcriptional profiles, has allowed us to resolve clinico-biological heterogeneity in some lymphoid tumors. For instance, two different DLBCL groups have been determined upon transcriptional similarity to GCBC or activated B cells, being the latest of worst prognosis.
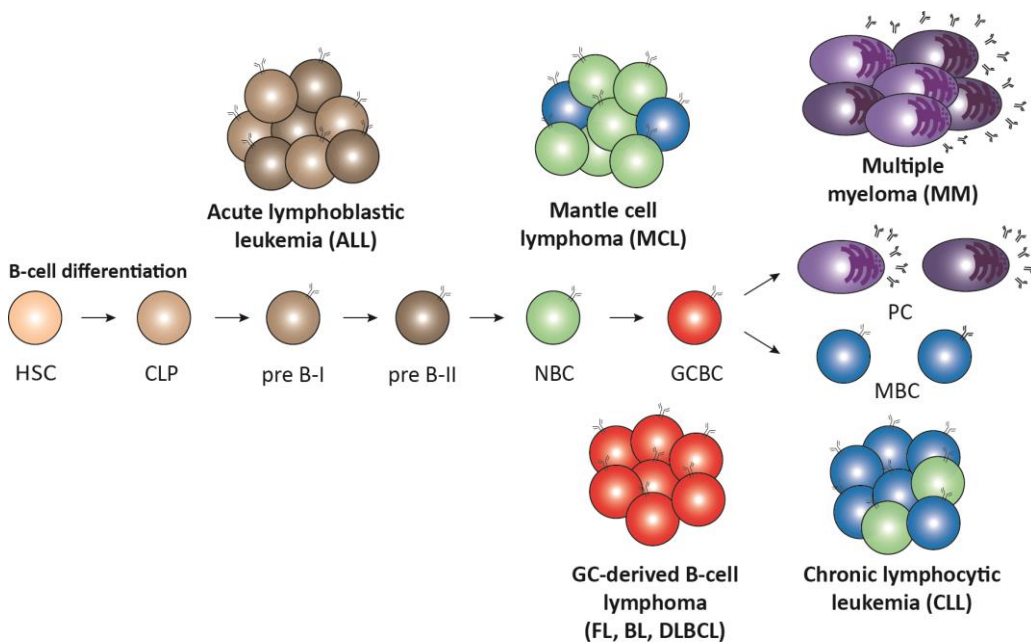
Figure 18. B-cell neoplasms. Schematic representation of the B-cell differentiation process and the derived B-cell neoplasms. In the central part of the figure B-cell lineage maturation stages are indicated, surrounded by several B-cell neoplasms derived from corresponding normal B-cell subpopulations (color code corresponds to the proposed cell of origin).

In the following sections, the two mature B-cell neoplasms studied in the course of this doctoral thesis, i.e. chronic lymphocytic leukemia and mantle cell lymphoma, will be described in detail.

2.3.1 Chronic lymphocytic leukemia

2.3.1.1 Epidemiology and clinical features

Chronic lymphocytic leukemia (CLL) is the most common leukemia of adults in western countries, accounting for an annual incidence of 5 cases per 100,000 people. The rate increases with age, increasing in individuals over 70 years old to more than 20

cases per 100,000 people per year. CLL is characterized by a proliferation and progressive accumulation of monomorphic small mature B cells coexpressing CD5 and CD23 antigens in PB, BM and secondary lymphoid tissues. A B-cell neoplasia defined by the presence of equal or greater than $5x10^9$/L monoclonal B-cells counts in PB (Swerdlow et al., 2017).

CLL is often preceded by a monoclonal B-cell lymphocytosis (MBL) which is an asymptomatic expansion of monoclonal B cells in PB. MBL is defined by the presence of less than $5x10^9$ cells/L with mainly CLL-like immunophenotype and is present in 5% of adults aged 60 years or older (Marti et al., 2005). MBL leads to leukemia requiring treatment in a 1-2% cases per year (Fazi et al., 2011; Rawstron et al., 2008). When presenting as a nodal, splenic or extramedullar tissues, CLL is diagnosed as small lymphocytic lymphoma (SLL), which is basically the same disease as CLL from the biological point of view (Hallek et al., 2008) .

CLL shows a male predominancy with a ratio of male-to-female 1.5-2:1. A familiar predisposition has been associated with 2-7 times high risk in CLL first-degree relatives (Hallek et al., 2008). CLL is a heterogeneous disease in terms of clinical behavior and outcome. Two staging methods have been described Rai and Binet systems. The Rai system is based on lymphocytosis and physical exam and divides into five stages which are classified into three upon treatment options, stage 0 is low risk, stage I and II are intermediate risk and stage III and IV are high risk (Rai et al., 1975). In the Binet staging system the number of affected lymphoid tissue groups are under consideration and is dividied into three stages: A (low risk, corresponding to Rai 0, I and II stages), B (intermediate, corresponding to Rai I and II stages) and C (worst prognosis corresponding to Rai III and IV) (Binet et al., 1981). Both methods have been used for several years and are still applied in the clinics. Nonetheless, other prognostic factors have emerged based on biological parameters. One major breakthrough was the identification of two distinguished clinico-biological groups based on the levels of somatic mutation in the immunoglobulin heavy chain variable regions (IGHV). A group of CLLs bearing high levels of IGHV somatic mutations (<98% identity with the germline), named mutated CLLs (mCLLs), having a favorable clinical outcome whereas CLLs with no or low levels of IGHV somatic mutations (i.e. ≥98% identity), which are called unmutated CLLs (uCLLs), having a worse clinical outcome (Damle et al., 1999;

Hamblin et al., 1999). In addition, several CLL patients share the same or very similar immunoglobulin sequences, a phenomenon called as "BCR stereotypy", i.e., stereotyped VH CDR3 sequences were found in one third of CLL cases. This suggest that subsets of CLL patients have a common specific antigen reactivity profiles and clinical outcome, which may direct the pathogenesis of CLL (Agathangelidis et al., 2012, 2014; Stamatopoulos et al., 2007; Swerdlow et al., 2017). Moreover, expression of certain genes has been associated with prognosis, as the expression of ZAP70, CD38 or CD49d (Cramer and Hallek, 2011; Swerdlow et al., 2017). Patients with more than 30% CD38+ cells are of worst prognosis and are seen to mainly bear unmutated IGHV (Ghia et al., 2003; Hamblin et al., 2002; Del Poeta et al., 2001). As for ZAP-70, its expression is also concordant with IGHV mutational states, being the IGHV unmutated cases those expressing more ZAP-70 (Cramer et al 2011). Prognosis can also be estimated using the serum markers β_2-microglobulin (sβ_2m), serum thymidine kinase (sTK), and soluble CD13 (sCD23) (Montserrat, 2006). Even though the simplicity and reliability of these assessments, the cut-offs may vary between laboratories. Therefore, standardization needs to be improved. Another limitation is that numerous factors have been described influencing serum markers as for the serum levels of sTK, which may be increased by vitamin B_{12} deficiency or viral infection (Cramer and Hallek, 2011; Hallek et al., 1992), restraining the broad applicability of these parameters.

CLL cells can progress, increasing their size, proliferative activity and significant clinical changes, transforming into a more agressive form of large cell lymphoma named Richter syndrome (RS, also called Richter's transformation) (Rossi and Gaidano, 2009). Approximately 2-8% of patients with CLL develop DLBCL-type RS, and less than 1% develop classic Hodgkin lymphoma (Brecher and Banks, 1990; Mao et al., 2007; Timár et al., 2004). RS prognosis is extremely poor, with median survival of a year.

2.3.1.2 Cell of origin

The cellular of origin of CLL is still controversial (Chiorazzi and Ferrarini, 2011). CLL cells express CD19, CD5 and CD23 markers not found in any normal B cell described so far, which complicates the determination of the CLL cell of origin (Gaidano et al., 2012). The idea of the existence of two cells of origin for CLL was born with the description of two different CLL subtypes based on IGHV mutational status. However, gene expression analysis among the two subgroups pointed to a single cell of origin as an antigen-

experienced resembling MBCs (Klein et al., 2001; Rosenwald et al., 2001). To reach a model among both theories, it was suggested that both subtypes (mCLL and uCLL) derive from marginal zone (MZ) B cells which can either express mutated or unmmutated IGHV. Though, transcriptome analysis pointed out that uCLL derives from CD5+ B cells whereas mCLL originates from CD5+ and CD27+ post-GCBCs, being CD27 upregulated upon T-cell independent antigen stimulation (Seifert et al., 2012). Even though, nowadays the WHO postulates that the CLL normal counterpart is an antigen-experienced mature CD5+ B cell with mutated and unmutated IGHV genes (Chiorazzi et al., 2005; Swerdlow et al., 2017).

Epigenetic studies from our group identified three distinct clinico-biological CLL subgroups related to different normal B-cell counterparts (Kulis et al., 2012). Unmutated-CLLs maintain an epigenetic signature of pre-GCBCs resembling NBCs and therefore was called naive B-cell like CLL (n-CLL), whereas mCLLs carried a post-GCBCs epigenetic signature, resembling MBCs and was named as memory B-cell like CLL (m-CLL). Interestingly, a subset of CLL cases with an intermediate epigenetic profiles between NBCs and MBCs was also identified (called intermediate CLL, i-CLL, accordingly), presenting intermediate level of IGHV somatic hypermutation, although its precise cellular origin remains unknown. These three groups show a clearly distinct clinical behavior, being the intermediate group in between n-CLL and m-CLL (Kulis et al., 2012; Oakes et al., 2016). These subgroups can be easily assessed using five epigenetic biomarkers on pyrosequencing analyses (Queirós et al., 2015). This finding was corroborated by other groups (Bhoi et al., 2016; Wojdacz et al., 2019).

2.3.1.3 Genetic features

Cytogenetic abnormalities are detected in 80-90% of CLL cases (Döhner et al., 2000; Malek, 2013; Puente et al., 2015; Swerdlow et al., 2017). The most common alteration present in around 50% cases is a deletion of the long arm (q) on chromosome 13 (13q14.3) involving the deleted in leukemia *DLEU1* and *DLEU2* genes, which code for noncoding transcripts, as well as, miR-16-1 and miR-15a, which play a role in controlling the proliferation of B cells. In fact, these microRNAs downregulate genes controlling cell cycle entry (Klein et al., 2010). Trisomy of chromosome 12 (trisomy 12) or partial trisomy 12q13 is another frequent alteration present in 20% cases. Other abnormalities are deletions of chromosome 11 short arm (11q22-23) affecting *ATM* and *BIRC3* leading

to alteration of the DNA damage response and the non-canonical NF-Kβ signaling, respectively (Rose-Zerilli et al., 2014). Also, deletions in chromosome 17 (17p13) which affect the tumor suppressor *TP53* gene as well as deletions in chromosome 6 (6q21) have been described (Döhner et al., 2000; Haferlach et al., 2007; Zenz et al., 2010)..

Genome-wide studies have elucidated the mutational landscape of CLL, which is characterized by few genes mutated at moderate frequency and a larger amount of genes altered in less than 5% of the cases (Landau et al., 2013, 2015; Puente et al., 2011, 2015; Quesada et al., 2012; Wang et al., 2011). In a study with 506 CLL patients, somatic mutations detected by whole-genome (WGS) and whole-exome (WES) sequencing identified 36 genes recurrently mutated. Those affected 3% to 15% of CLL cases and included genes on cellular processes as DNA damage response (*TP53, ATM, POT1*), NOTCH1 signaling (*NOTCH1, FBXW7*), RNA maturation and export (*SF3B3, XPO1*), genome and chromatin structure (*CHD2*), NF-Kβ (*BIRC3, NFKBIE, EGR2, TRAF3, NFKB2*) or B-cell signaling (*MYD88, IKZF3, TLR2, BCOR, t(14;18)/BCL2, KRAS/NRAS*) (Puente et al., 2015)*.* Mutations could be classified into three main signatures, which are aging, AID, and noncanonical AID (nc-AID) (Alexandrov et al., 2013; Kasar et al., 2015; Puente et al., 2015). The latest are mutations caused by the processing of the AID-induced cytidine deamination and the mismatch repair pathways that recruits the error –prone DNA polymerase η. This leads to high number of mutations in mCLLs and can also be found on other lymphoid neoplasms derived from GC-experienced cells (Alexandrov et al., 2013; Kasar et al., 2015; Puente et al., 2018).

However, it seems that mutational landascape itself cannot fully explain the heterogeneous outcome of the patients, which might be additionally asociated with subclonal composition (Landau et al., 2013, 2015; Nadeu et al., 2016, 2018; Puente et al., 2015), identification of convergent mutational evolution in few patients (Jethwa et al., 2013; Ojha et al., 2015), and different patterns of clonal diversification upon disease progression (Amin et al., 2016; Landau et al., 2014; Rose-Zerilli et al., 2016; Schuh et al., 2012). Interestingly, the frequency of the mutations is increased when considering subclonal diversity. An increase in the subclonal diversity shortened the overall survival, although it was also related to the age of patients, IGHV and *TP53* status of the tumors (Nadeu et al., 2018).

2.3.1.4 Epigenetic features

CLL evolution is marked by widespread alterations in the epigenome. Looking at DNA methylation, a global DNA hypomethylation in gene bodies and enhancer sites have been described as compared to normal B cells (Kulis et al., 2012; Oakes et al., 2016). However, in general DNA methylation alterations are rather poorly associated with gene expression in CLL. In some instances, the loss of DNA methylation is associated with gene activation in CLL, although the activation of oncogenes through DNA hypomethylation is not frequent in CLL. However, DNA hypermethylation of tumor suppressor genes has been far more frequently studied (Cahill and Rosenquist, 2013). DNA methylation induces silencing of a pro-apoptotic gene, *DAPK1*, in almost all sporadic cases of CLL. Interestingly, DAPK1 downregulation through promoter methylation contributes to a heritable predisposition to CLL (Raval et al., 2007). Another example is ZAP70 a prognosticator in CLL that was shown to be differentially methylated. Low ZAP70 expression was associated with DNA methylation silencing and was found in good-prognostic mCLLs (Chantepie et al., 2010; Claus et al., 2012; Corcoran et al., 2005). In fact, over 3,265 CpGs were detected by microarray to be differentially methylated between uCLL and mCLL subgroup (Kulis et al., 2012). These affected for instance epigenetic regulators (*HDAC9*, *HDAC4* and *DNMT3B*) and pathways related to B-cell signaling (IBTK) and numerous TGF-β and NF-Kβ/TNF pathways (Cahill et al., 2013). Moreover, studies on DNA methylation in CLL showed that clonal cases maintained a low methylation heterogeneity whereas high DNA methylation changes on CLL were associated with subclonal genetic alterations, shorter time to treatment and poor prognostic (Oakes et al., 2014).

Moving to other epigenetic players, a study on WGS and WES analyses was specifically integrated with ChIP-seq and DNase-seq in normal B cells and CLL demonstrating particular recurrent mutations in a non-coding region harboring active enhancer (Puente et al., 2015). This region showed high 3D contact frequencies with the *PAX5* locus in CLL patients, a TF essential for B-cell differentiation. This finding suggested that PAX5 enhancer mutations might constitute driver events contributing to the development of CLL and pointed to the relationship between genetics and epigenetics analyses (Puente et al., 2015). Recently, the reference epigenome, consisiting of genome-wide maps of six histone marks, DNA accessibility, DNA

methylation and gene expression of seven primary CLL cases were described (Beekman et al., 2018a). Those thoroughly characterized CLL samples were analysed together with the regulatory chromatin landscape of 100 additional CLL cases, which was studied by taking into account chromatin accessibility and active regulatory regions, marked by the histone modification H3K27ac. Additionally, all these CLL cases were examined in the context of the entire mature B-cell differentiation. This data set allowed to detect roughly 500 regions that become *de novo* active specifically in CLL and were enriched for binding sites of NFAT, FOX and TCF/LEF TFs. In that study it was also observed that trisomy 12 and *MYD88* mutations show a distinct chromatin profile from those cases lacking these genetic changes (Beekman et al., 2018a). These epigenetic analyses have significantly contributed to better understand the cellular origin, pathogenesis and clinical behavior of CLL.

2.3.2 Mantle cell lymphoma

2.3.2.1 Epidemiology and clinical features

Mantle cell lymphoma (MCL) is a mature B-cell neoplasm accounting for approximately 3-10% of all non-Hodgkin lymphomas (Swerdlow et al., 2017; 1997). It is prevalent in middle-aged adults to older people with a median age of 60 years old and predominance in males with a male-to-female ratio of 2:1. Similar to CLL and other lymphoid neoplasms, cases of MCL have been described on families with first-degree relative developed the neoplasia (Argatoff et al., 1997; Bosch et al., 1998; Campo et al., 1999; Lardelli et al., 1990; Swerdlow and Williams, 2002; Swerdlow et al., 1983).

MCL has been considered one of the most aggressive lymphomas and incurable lymphoid neoplasias, with many patients following a relative rapid disease evolution with limited responses to therapeutic strategies, leading a pretty short median survival of only 3 to 5 years (Jares and Campo, 2008; Swerdlow et al., 2017). Clinical presentation of most MCL patients results from a proliferation and dissemination of mature B lymphocytes, which infiltrates lymphoid tissues leading to lymphadenopathy, but also hepatosplenomegaly, as well as, involvement of BM, extranodal sites and frequently PB (Bosch et al., 1998; Campo et al., 1999; Norton et al., 1995; Swerdlow and Williams, 2002).

Two differential cytological variants have been observed, i.e. the classical and the blastoid/pleomorphic. The classical appearance is a monotonous proliferation of small to medium cells with irregular nuclei and inconspicuous nucleoli. Otherwise, the blastoid and pleomorphic MCL variants are present in 10 to 20% of patients, and associated with cases with more aggressive clinical evolution. Both variants are frequently tetraploid and present high proliferation rates, although is higher in the blastoid variant (Jares and Campo, 2008). The immunophenotype of MCL cells is characterized by expression of CD19, CD20, CD22 and CD79a and also CD5, although some cases are negative for CD5 (Jares et al., 2007; Swerdlow et al., 2017). MCLs usually express IgM and IgD, and they are uniformly positive for BCL-2 protein (Swerdlow et al., 1993), while are negative for CD23 and for GC proteins as CD10 and BCL6 (Jares et al., 2007; Swerdlow et al., 2017).

Even though the traditional consideration of MCL as a very aggressive neoplasia, a subset of patients follow an indolent clinical course, which have been recognized in the WHO classification as leukemic non-nodal MCLs (nnMCL) (Campo and Rule, 2015; Royo et al., 2012). Those patients show a longer overall survival, of 7 to 10 years, and benefit from "watch and wait" approach as they do not need therapy for a long period of time (Royo et al., 2012). At the moment of diagnosis these cases present lymphocytosis and sometimes splenomegaly but without significant lymphadenopathy (Furtado and Rule, 2011; Swerdlow et al., 2017). Very low proliferation index, high levels of somatic mutation in the IGHV locus and few chromosomal alterations are described on nnMCL cases being stable and asymptomatic for long periods (Fernandez et al., 2010; Jares et al., 2012a; Puente et al., 2018; Royo et al., 2012). This subgroup differs from the conventional MCL (cMCL) one, which is associated with generalized lymphadenopathy, as well as, complex karyotypes, requiring treatment at diagnosis, following an aggressive evolution. Both subgroups can progress to more aggressive variants through the accumulation of secondary genetic alterations (Navarro et al., 2012; Nygren et al., 2012), and through the adquisition of tumor suppressor genes inactivating mutations, such as *TP53* or *CDKN2A* (Izban et al., 2000).

A gene expression profile study supported the overall MCL identity of these subgroups showing a cluster between nnMCL and cMCL cases differing from other neoplasias as CLL. A differentially expression analysis among these MCL subgroups

identified a small signature of 13 expressed genes highly expressed in cMCL but negative or very low in nnMCL tumors, including the expression of SOX11 in cMCL (Fernandez et al., 2010).

2.3.2.2 Role of SOX11

The main gene differentially expressed between MCL subgroups is *SOX11* (Fernandez et al., 2010). This key transcription factor has been identified as a specific marker of prognostic and diagnostic value on MCL cases (Ek et al., 2008), being highly expressed in the most aggressive behavior cases (cMCL) (Fernandez et al., 2010).

SOX11 is a member of the SRY (Sex determining Region Y)-related HMG (High Mobility Group)-box family of transcription factors characterized by containing a HMG DNA-binding domain. SOX11 together with SOX4 and SOX12 belongs to the SoxC protein family group (Wegner, 2010). SOX4 and SOX11 are essential for organogenesis, neural development and neurite growth (Dy et al., 2008; Kavyanifar et al., 2018; Penzo-Méndez, 2010). However, SOX4 has been described in T and B lymphopoiesis (Schilham et al., 1996, 1997) while *SOX11* is not expressed in normal lymphoid tissues, lymphoid progenitors or normal B cells at any maturation stage. *SOX11* has been shown to be highly expressed in different solid tumors such as glioma, specifically glioblastoma multiforme (Hide et al., 2009; Weigle et al., 2005), medulloblastoma (Lee et al., 2002) and epithelial ovarian cancer (Brennan et al., 2009), as well as hematological malignances. For instance, SOX11 apart from cMCLs has been described to be expressed in most B and T-cell lymphoblastic leukemia, some T-prolymphocytic leukemias (Dictor et al., 2009; Mozos et al., 2009), 30% Burkitt lymphoma (Dictor et al., 2009; Mozos et al., 2009; Wästerlid et al., 2017) and 50% hairy cell leukemia. SOX11 is well established as a marker that helps to distinguish the two different clinico-biological MCL subtypes. However, SOX11 cannot be employed as mere "prognostic parameter" as some exceptions have been described between SOX11 expression and adverse outcome in MCL (Beekman et al., 2018b). The first one is the presence of some MCL cases lacking SOX11 expression related with poor outcome due to additional alterations such as *TP53* (Nordström et al., 2014; Nygren et al., 2012). The mutations of this tumor suppressor are consistent with a poor evolution in SOX11-positive as well as SOX11-negative MCLs (Royo et al., 2012). Another concern is the grouping of samples, as some studies have considered as SOX11-negative expression levels below the median of all

cases, resulting on positive cases grouped in the negative ones (Kuo et al., 2015). Finally, a more technical issue against SOX11 as the "solo prognostic parameter" is the crossreactivity of some Abs with other members of the SOX family leading to immunohistochemical differences (Nakashima et al., 2014; Soldini et al., 2014). All these observations indicate that integration of different molecular, genetic and pathological aspects should be considered to properly classify MCL subgroups.

Aberrant SOX11 overexpression in MCL contribute to the blocking of the terminal B-cell differentiation process by upregulating PAX5 which represses genes involved in PC differentiation (as BLIMP1 and XBP1) (Ferrando, 2013; Vegliante et al., 2013), (**Figure 19**). Another direct target of SOX11 is *BCL6*, an essential element for development and maintenance of the follicular GC. SOX11 may block the expression of *BCL6* preventing the entrance of MCL cells in the GC (Palomero et al., 2016), (**Figure 19**). Apart from altering B-cell differentiation, SOX11 has been described to interact with tumor microenvironment by regulating the expression of platelet-derived growth factor alpha (*PDGFA*) promoting angiogenesis (Palomero et al., 2015), (**Figure 19**). Also, SOX11 through the interaction with (C-X-C motif) chemokine receptor 4 (*CXCR4*) and *PTK2*, encoding for focal adhesion kinase (FAK), regulates tumor cell migration, adhesion to stromal cells, cell proliferation and resistance to conventional drug therapies (Balsas et al., 2017), (**Figure 19**).
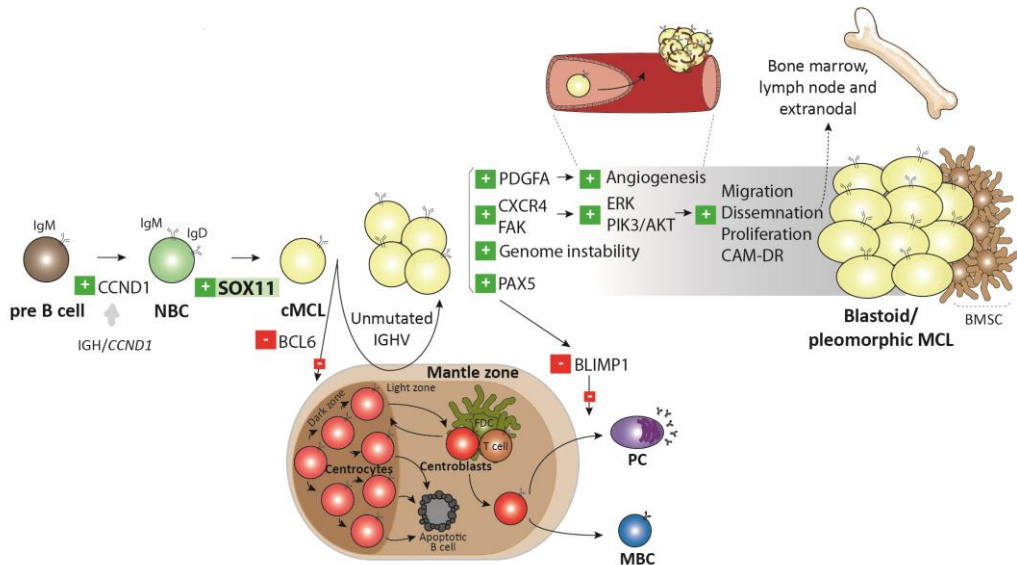
Figure 19. SOX11 oncogenic functions. Conventional SOX11-positive MCL (cMCL) cases acquire secondary genomic alteration and mutations targeting important pathways leading to high

proliferation and aggressive clinical behavior. SOX11 overexpression blocks terminal differentiation by retaining PAX5 expression and consequently BLIMP1 inactivation. Also, SOX11 repressed BCL6 transcription, blocking the entrance of the cells into the germinal center and maintaining unmutated IGHV. Furthermore, SOX11 promotes angiogenesis via PDGFA pathway activation and it also promotes cell migration and adhesion to bone marrow and lymph node by regulating CXCR4 and PTK2/FAK gene transcription. CAM-D, Cell Adhesion Mediated-Drug resistance. Figure adapted from Beekman et al., 2018b.

The mechanisms behind SOX11 expression in cMCL are unknown. No mutations or genetic alterations such as translocations, amplifications at the *SOX11* locus in MCL samples related to its expression have been described (Bea et al., 2013; Royo et al., 2011). From the epigenetic view, the promoter of *SOX11* does not present aberrant DNA methylation but instead, activating histone modifications seem to be associated with *SOX11* gene expression (Vegliante et al., 2011; Wasik et al., 2013). However, the causes behind the aberrant SOX11 activation are still broadly unknown, and this issue represents one of the topics studied in this doctoral thesis.

2.3.2.3 Cell of origin

CD5-positive NBCs have been long considered to be the normal counterparts of MCL for their expression of IgM/IgD and CD5, location in the mantle zone, and characterized to have unmutated IGHV (Jares et al., 2012a). However, a small group of around 15% of MCLs carry IGHV hypermutations, suggesting that these neoplastic cells may arise from cells that have gone through the GC maturation (Hadzidimitriou et al., 2011; Orchard et al., 2003). Interestingly, the IGHV status in MCL was reported to significantly associate with the overall survival. Hence, this molecular feature, along with the clinical and pathological traits, allows to distinguish the aforementioned two MCL subgroups and hinted the cell of origin. In that way, cMCL cases with unmutated IGHV were suggested to derive from NBCs, while nnMCL presenting mutated IGHV genes were supposed to have the post-GC origin (Navarro et al., 2012).

2.3.2.4 Genetic features

The initial oncogenic event in MCL is the translocation event t(11;14)(q13;q32) that juxtaposes the IGH at chromosome 14q32 nearby the *CCND1* gene at 11q13 (de Boer et al., 1995; Bosch et al., 1994; Campo et al., 1999; Wiestner et al., 2007). This is considered the primary oncogenic mechanism in the development of MCL, taking place

on B cells at the pre-B stage of differentiation initiated by the recombination of the V(D)J segments from the IG gene (Welzel et al., 2001). CCND1 facilitates cell cycle progression by overcoming the suppressor effect that retinoblastoma (RB) performs in the G1/S transition (Harbour and Dean, 2000). Apart from cell cycle, CCND1 participates in transcription regulation (Albero et al., 2018) as well as DNA damage response, cell migration, apoptosis escape and cell bioenergetics (Pestell, 2013). Even though the translocation is present in more than 95% of the cases (Vaandrager et al., 1996), a subset of CCND1-negative MCL have been recognized harboring high expression of *CCND2* and *CCND3* (Fu, 2005; Martín-Garcia et al., 2019; Salaverria et al., 2007). In fact chromosomal rearrangements on *CCND2* locus usually with an IG partner (as IGK or IGL) have been detected in 55% of the CCND1-negative cases (Salaverria et al., 2013). Interestingly, these cases point out the great importance of cyclin D family members on MCL lymphomagenesis.

The identification of t(11;14) in blood cells of 1-2% of healthy individuals (Hirt et al., 2004) and studies performed on transgenic mouse showed that *CCND1* overexpression alone is not able to develop lymphoma, which suggest that the complete cell transformation should be reached by the cooperation of other oncogenes as *MYC* (Lovec et al., 1994).

More than 90% of the MCL cases present at least one copy number change in addition to the primary genetic change, and the total number per case is higher than in other lymphoid neoplasias (Royo et al., 2011). MCL genome frequently suffers massive DNA rearrangements affecting one or few chromosomes, a phenomenon known as chromothripsis (Bea et al., 2013; Korbel and Campbell, 2013). Loss of material is usually related to the inactivation of tumor suppressor genes, while amplified regions may lead to oncogene overexpression. The regions frequently affected by losses and the gene targets are 1p (in 29-52% of cases), 6p (*TNFAIP3*, in 23-38%), 9q21 (*CDKN2A*, in 18-31%), 11q (*ATM*, in 21-59%), 13q (in 22-55%), 13q (in 43-52%), and 17p (*TP53*, in 21-45%), as well as gains on 3q (in 31-50% of cases), 7p (in 16-34%), and 8q (*MYC*, in 16-36%) (Beà et al., 1999; Rubio-Moscardo et al., 2005; Seto et al., 1992). Trisomy 12 has been reported in 25% of cases, but usually in the context of other alterations (Cuneo et al., 1999). The majority of the chromosome abnormalities in MCL affect two common pathogenic pathways: 1) the cell cycle machinery and 2) cellular response to DNA

damage (Fernàndez et al., 2005). An interesting example of the first one could be the loss of *CDKN2A*, which encodes for two important transcripts, i.e. INK4a, a CDK4 inhibitor, and ARF, a stabilizer of p53 protein, and thus alter two major regulatory pathways, INK4a/CDK4/RB1 and ARF/MDM2/p53. These regulatory axes may be also deregulated by microdeletion of *RB1,* also found in some MCLs (Pinyol et al., 2007), as well as *TP53* mutations usually associated with deletion of the locus (17q), identified in approximately 30% of highly proliferative MCLs (Greiner et al., 1996; Hernandez et al., 1996). As for the second aforementioned pathogenic pathway altered in MCL, that is the cellular response to DNA damage, it is mainly affected by the losses of essential genes in the DNA damage response as is the ataxia-telangiectasia mutated gene (*ATM*) (Stilgenbauer et al., 1999).

Genome-wide studies by WGS and WES have revealed a heterogeneous spectrum of somatic mutations on MCL tumor (Bea et al., 2013; Zhang et al., 2014), especially in SOX11-positive cases (Bea et al., 2013). Commonly known MCL drivers (such as *ATM, CCND1* and *TP53*) , were among the most mutated genes, detected in these studies Particularly, *ATM* mutations were present in 55% of tumors expressing SOX11, while *CCND1* mutations were predominantly located in the exon 1 and were more frequent in SOX11-negative MCL. Those particular mutations appear to increase cyclin D1 stability (Mohanty et al., 2016). *TP53* (15-30%) was equally distributed on both SOX11 positive and negative MCL subgroups. Nonetheless, other genes mutated have been identified at low frequencies (<10-15%), such as *NOTCH1* and *NOTCH2, BIRC3* the chromatin modifiers: *WHSC1*, *MLL2*, and *MEF2B* (Bea et al., 2013). Overall, global studies seems to be more informative while studying MCL pathogenesis as mutations and genetic alterations target recurrent functional pathways, rather than particular genes.

2.3.2.5 Epigenetic features

The few epigenetic studies published so far in MCL have focused on DNA methylation. The first studies were mainly centered on analysis of CpG islands and promoter regions reporting a predominant DNA hypomethylation in MCL compared to NBCs (Leshchenko et al., 2010). Promoter DNA methylation was shown to be inversely correlating with gene expression as demonstrated on four hypomethylated genes (*CD37*, *HDAC1*, *NOTCH1* and *CDK5*) in MCL. Those genes were involved in pathways

controlling biological processes with oncogenic activity as cell cycle control and apoptosis and were highly expressed whereas four DNA hypermethylated genes (*CDN2B*, *MLF-1*, *PCDH8* and *HOXD8*), functioning as tumor suppressor in the MCL context, showed gene expression downregulation (Leshchenko et al., 2010). A study comparing DNA methylation of CLL and MCL profiles described a prominent DNA hypermethylation of developmental genes, as homeobox transcription factor genes (*HLXB9*, *HOXA13*) in MCL as compared to CLL, which was suggested to lead to high proliferation and shorter overall survival on MCL cases (Halldórsdóttir et al., 2012). Indeed, a signature of five genes (*SOX9*, *HOXA9*, *AHR*, *NR2F2* and *ROBO1*) hypermethylated in MCL was defined and correlated with higher proliferation, increased number of chromosomal abnormalities and an overall shorter survival of patients (Enjuanes et al., 2011). These features were further associated with a MCL subset characterized by an accumulation of CpG hypermethylation described in by genome-wide study focused on promoter methylation (Enjuanes et al., 2013). Even so, a complete characterization of the DNA methylome of MCL along with other epigenetic layers as its 3D chromatin structure remains undone, and represents one of the subjects analyzed in the course of this doctoral thesis.

AIMS

It is becoming gradually clear that gene regulation is mediated by a complex interplay among multiple epigenetic components, including DNA methylation, chromatin marks and 3D genome architecture. This last component is gaining importance in recent years, and several studies are showing its key role during cell differentiation and in human disease. In normal and neoplastic B cells, however, the 3D genome structure is widely unknown. Therefore, the global aim of this doctoral thesis is to perform an integrative analysis of the 3D genome architecture together with other molecular features in normal and neoplastic B cells to provide new biological and clinical insights.

This global aim is divided into four specific aims, which are associated with each of the four results chapters of this thesis.

1. Generate a comprehensive multi-omics portrait of the 3D genome architecture during normal B-cell differentiation and neoplastic transformation.

2. Characterize 3D genome changes associated with chromosomal rearrangements in B-cell neoplasms.

3. Identify candidate epigenetic drivers of mantle cell lymphoma by an integrative analysis of DNA methylation, histone modifications and 3D genome interactions.

4. Elucidate the potential mechanisms underlying aberrant expression of the *SOX11* oncogene in conventional mantle cell lymphoma.

MATERIALS AND METHODS

1. Sample preparation

1.1 Isolation of B-cell subpopulations

B-cell subpopulations spanning mature normal B-cell differentiation were sorted as previously described (Kulis et al., 2015). Briefly, PB B-cell subpopulations i.e. NBC and MBC were obtained from buffy coats from healthy adult male donors of age ranging between 56 to 61 years, obtained from Banc de Sang i Teixits (Catalunya, Spain). GCBC and PC were isolated from tonsils of male children undergoing tonsillectomy, ranging in age between 2 to 12 years, obtained from the Clínica Universidad de Navarra (Pamplona, Spain). Samples were density gradient centrifuged with Ficoll-Isopaque, in case of tonsil were previously minced extensively. After density centrifugation, CD19+ B cells were isolated by positive magnetic cell separation using AutoMACS system (Milteny Biotec, Auburn, CA) followed by FACS sorting on FACSAriaII (BD Biosciences), (**Figure 20** from **Study 1**). In case of chromatin preparation the crosslinking was performed before sorting. The use of the samples analyzed in the present study was approved by the ethics committee of the Hospital Clínic de Barcelona and Clínica Universidad de Navarra.

1.2 Patient samples

The reference epigenome CLL (n=7) (Beekman et al., 2018a) and MCL (n=5) patient samples were obtained from cryopreserved mononuclear cells from the Hematopathology collection registered at the Biobank (Hospital Clínic-IDIBAPS; R121004-094). All samples were >85% tumor content. Clinical and biological characteristics of the patients are shown in **Appendix Table 1**. The enrolled patients gave informed consent for scientific study following the International Cancer Genome Consortium (ICGC) guidelines and the ICGC Ethics and Policy committee (Consortium, 2010). This study was approved by the clinical research ethics committee of the Hospital Clínic of Barcelona.

Patients samples from **Study 3**, MCL (n=82) were obtained from the Tumor Bank of the Hospital Clínic in Barcelona and the Institute of Human Genetics in Kiel. In all cases, patients gave their written informed consent, and the study was approved by the clinical research ethics committee of the Hospital Clínic of Barcelona (number 2009/5069) and the internal review board of the University of Kiel (number 447/10).

Data was deposited in the European Genome-phenome Archive (EGA). MCL whole-genome bisulfite sequencing (WGBS), ChIP-seq and microarray data is under the accession numbers EGAS00001001638, EGAD00001002655 and EGAS00001001637, respectively. Normal B cells WGBS data could be accessed under accessions EGAD00001001304 and EGAS00001000272, while microarray data are available under accession EGAS00001001196.

1.3 Cell culture

The t(11;14)(q13;q32)-positive MCL cell lines Z-138, GRANTA-519 and JeKo-1 (SOX11-positive) and JVM-2 (SOX11-negative) were used as cellular models of MCLs. These cell lines were maintained in RPMI 1640 (Z-138, JeKo-1, JVM-2) or DMEM (GRANTA-159), 10% FBS and Penicilin/Streptomicin at 37°C with 5% CO_2. Cell lines tested negative for mycoplasma contamination, and their authenticity was confirmed by qCell Identity (qGenomics).

2. Cell proliferation and cell death

Measurement of cell proliferation were performed by the MTT assay, based on the reduction of 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide by proliferating cells. Cells on a 96-well plate were incubated during approximately 2 hours with 10μl of MTT. Reaction was then stopped with 100μl Isopropanol-HCl 1M (24:1). Absorbance was measured at 570nm and 655nm on the microplate reader Synergy™ HT (BioTek). Treated cells were analyzed together with untreated control cells, used as references.

Cell death was analyzed using Annexin V Apoptosis Detection Kit FITC (eBioscience) according to manufacturer's guidelines. Briefly, cells were incubated for 10 minutes at RT with Annexin incubation buffer including Annexin-FITC and propidium iodide (PI) to quantify apoptosis and necrosis, respectively. Analyses were performed in Attune Nxt Flow Cytometer (ThermoFisher).

3. 3D structure data generation

3.1 Dilution Hi-C

HiC-seq was optimized in collaboration with the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain) following the described protocol (van Berkum et al., 2010). Two MCL cell lines, JVM-2 and Z-138, and a CLL case were successfully processed. Briefly, twenty-five million cells were used per library and two libraries were performed per sample using two different six-cutter enzymes (HindIII and NcoI). The CLL case was used on two different publications (Beekman et al., 2018a; Raineri et al., 2018).

3.2 *In situ* Hi-C

In situ Hi-C was performed based on the previously described protocol (Rao et al., 2014). Two million of crosslinked cells (at 1% formaldehyde for 10 minutes on RT) per sample were used as starting material. Chromatin was digested adding 100U DpnII (New England BioLabs) on overnight incubation. After the fill-in with bio-dCTP (Life-Technologies, 19518-018), nuclei were centrifuged 5 minutes, 3000rpm at 4°C and ligation was performed for 4 hours at 16°C adding 2µl of 2000U/µl T4 DNA ligase (NEB, M0202) in total volume of 1.2mL of ligation mix (120µl of 10X T4 DNA ligase buffer; 100µl of 10% Triton X-100; 12µl of 10mg/ml BSA; 966µl of H_2O). Following ligation, nuclei were pelleted and resuspended with 400µl 1X NEBuffer2 (New England BioLabs). Then, 10µl of RNAseA (10mg/ml) was added to the nuclei and incubated during 15 minutes at 37°C while shaking (300rpm), and after that 20µl of proteinase K (10mg/mL) was added and incubated overnight at 65°C while shaking (600rpm). After reversion of the crosslinked material, DNA was extracted by phenol/chloroform/isoamyl alcohol and DNA was precipitated by adding to the upper aqueous phase: 0.1X of 3M sodium acetate pH 5.2, 2.5X of pure ethanol and 50µg/ml glycogen. Samples were mixed and incubated overnight at -80°C. Next, samples were centrifuged 30 minutes at 13,000rpm at 4°C and pellet was washed with 1mL of EtOH 70% followed by a 15 minutes centrifugation at 13,000rpm at 4°C. The supernatant was discarded and the pellet air-dried for 5 minutes and resuspended in 130µl of 1X Tris buffer (10 mM TrisHCl, pH 8.0), which, in order to be fully dissolved, was incubated at 37°C for 15 minutes. Purified DNA was sonicated using Covaris S220, and then the final volume was adjusted to 300µl

with 1X Tris buffer. Sonicated DNA was mixed with washed magnetic streptavidin T1 beads (total of 100µl 10mg/ml beads), split in two tubes (150µl each), and incubated for 30 minutes at RT under rotation. Subsequently, beads were separated on the magnet, the supernatant discarded and the DNA was washed with 400µl of BB 1X, twice. Sonicated DNA conjugated with beads was washed with 100µl of 1X T4 DNA ligase buffer, pooling the two tubes per condition. After that, beads were reclaimed in end-repair mix. Once incubated during 30 minutes at RT the beads were washed twice with 400µl of BB 1X. Then, beads were washed with 100µl of NEBuffer2 and reclaimed in A-tailing mix, incubated during 30 minutes at 37°C and washed twice with 400µl of binding buffer (BB) 1X, followed by a wash in 100µl of 1X T4 DNA ligase buffer. Afterwards, the beads were resuspended in 50µl of 1X Quick ligation buffer, 2.5µl of Illumina adaptors and 4,000U of T4 DNA ligase and incubated during 15 minutes at RT. Next, beads were washed twice with 400µl BB 1X and resuspended in 30µl of 1X Tris buffer. In the end, libraries were amplified by eight cycle of PCR using 8.3µl of beads and pooling a total of 4 PCRs per sample. The PCR products were mixed by pipetting with an equal volume of AMPure XP beads and incubated at RT for 5 minutes. Beads were washed with 700µl of EtOH 70%, without mixing, twice, and left the EtOH evaporate at RT without over-drying the beads (aprox. 4 minutes). Finally, the beads were resuspended with 30µl 1X Tris buffer, incubated during 5 minutes and supernatant containing the purified library was transferred to a new tube and stored at -20°C. DNA was quantified by Qubit dsDNA High Sensitivity Assay, the library profile was evaluated on the Bioanalayzer 2100 and the ligation was assessed. Libraries were sequenced on HiSeq 2500. **Appendix Table 2** and **3** (**Study 1**) summarizes the number of reads sequenced and quality metrics for each B-cell subpopulation replicate and B-cell neoplasm.

3.3 4C-sequencing

4C templates were prepared as previously described (Simonis et al., 2007a; van de Werken et al., 2012b). Briefly, $1x10^7$ cells (MCL primary cases, GRANTA-519, JEKO-1, Z-138 and JVM-2) or $4x10^6$ cells (NBC and MBC) were crosslinked with 2% (MCL primary cases, GRANTA-519, JEKO-1, Z-138 and JVM-2) or 1% (NBCs and MBCs) formaldehyde. Chromatin was digested with a first restriction enzyme (NlaIII) followed by ligation. Next, chromatin was decrosslinked and DNA was digested with a second

restriction enzyme (DpnII) and ligated. PCR amplification of viewpoint regions and its ligated fragments was performed using the 4C templates. Restriction enzymes and primers used in **Study 3** are indicated in **Table 1**.

Region	Viewpoint Fragment (GRCh37)	RE1	RE2	RE1_primer	RE2_primer
SOX11 locus	chr2:5834180-5835254	NlaIII	DpnII	CCACCAAAATTTTCATCATG	TCTTCTATGCATCCGATTCT
SOX11 enhancer	chr2:6492207-6492728	NlaIII	DpnII	TCAGACTGACTTTCCTCATG	TCTTCGTGTTTAAGATTCCC

Table 1. Primers used for 4C sequencing. RE1 = first restriction enzyme, RE2 = second restriction enzyme.

3.4 Fluorescence *in situ* hybridization (FISH)

Z-138, JVM-2, JeKo-1 and GRANTA-519 MCL cell lines and the lymphoblastoid GM12878 cell line were cultured, nuclei extracted and fixed with carnoy's solution according to standard methods. Additionally, suspensions from five SOX11-positive and five SOX11-negative MCL patients and two healthy donor samples were used. All samples from primary MCLs had the CCND1/IGH fusion in more than 70% of the cells. Bacterial artificial chromosomes (BAC) (ThermoFisher) RP11-626J3 spanning the *SOX11* gene (chr2:5,782,225-5,950,258; GRCh37) and RP11-799N5 spanning the regulatory region of interest (chr2:6,334,254-6,527,002; GRCh37) were labeled by nick translation with spectrum red and spectrum green, respectively. FISH was performed as standard protocols. Clones were tested individually by hybridization on normal metaphases. The FISH images were acquired with a Leica TCS SP5 Spectra Confocal microscope (Leica TCS SP5) using a 63x 1.4 oil objective lenses. Optical sections (Z-step sizes) were captured every 0.5μm. The distances between red and green signals were measured from border to border using the Fiji-ImageJ software until fifty signals per sample were scored. On both presented images, after the brightness and contrast were adjusted and filtered (mean=1), three confocal plains were selected and a Z-projection on maximum intensity was performed.

4. Hi-C data analyses

4.1 Hi-C data pre-processing, normalization and interaction calling

The sequencing reads of Hi-C experiments were processed with TADbit (Serra et al., 2017). Briefly, sequencing reads were aligned to the reference genome (GRCh38) applying a fragment-based strategy; dependent on GEM mapper (Marco-Sola et al., 2012). The mapped reads were filtered to remove those resulting from unspecified ligations, errors or experimental artefacts. Specifically, seven different filters were applied using the default parameters in TADbit: self-circles, dangling ends, errors, extra dangling-ends, over-represented, duplicated and random breaks (Serra et al., 2017). Hi-C data were normalized using the OneD correction (Vidal et al., 2018) at 100Kb of resolution to remove known experimental biases. The significant Hi-C interactions were called with the *analyzeHiC* function of the HOMER software suite (Heinz et al., 2010), binned at 10Kb of resolution and with the default *p*-value threshold of 0.001.

4.2 Reproducibility of Hi-C replicas

The agreement between Hi-C replicates was assessed using the reproducibility score (RS) (Yan et al., 2017). A genome-wide RS was defined for each experiment as the average RS between pairs of corresponding normalized chromosome matrix (**Figure 22A** and **Figure 34A** from **Study 1**). Then, the matrix representing all the genome-wide RSs was analyzed using a hierarchical clustering algorithm with the Ward's agglomeration method using *hclust* function from R stats package.

4.3 Definition of sub-nuclear genome compartmentalization

The segmentation of the genome into compartments was determined as previously described (Lieberman-Aiden et al., 2009). In short, normalized chromosome-wide interaction matrices at 100Kb resolution were transformed into Pearson correlation matrices. These correlation matrices were then used to perform PCA for which the first eigenvector (EV) normally delineates genome segregation. All EVs were visually inspected to ensure that the EV selected corresponded to genomic compartments and were correctly identified (Lieberman-Aiden et al., 2009). The multi-modal distribution of the EV coefficients from the B-cells dataset was modelled as a Gaussian mixture with three components (k=3). To estimate the mixture distribution

parameters, an Expectation Maximization algorithm using the *normalmixEM* function from the *mixtools* R package was applied (Benaglia et al., 2009). A Bayesian Information Criterion (BIC) was computed for the specified mixture models of clusters (from 1 to 10) using *mclustBIC* function from *mclust* package in R (Scrucca et al., 2016), (**Figure 26B** from **Study 1**). Three underlying structures were defined; an alternative compartmentalization into A-type (with the most positive EV values), B-type (with the most negative EV values) and I-type (an intermediate-valued region with a distinct distribution) compartments. Two intersection values (IV1, IV2) were defined at the intersection points between two components (**Figure 26C** from **Study 1**).

4.4 Characterizing compartment types in B cells by integrating nine omic layers

Given a set of peaks as previous defined by Beekman and colleagues (Beekman et al., 2018a) from nine different omic layers including six histone marks (H3K4me3, H3K4me1, H3K27ac, H3K36me3, H3K9me3, H3K27me3), gene accessibility (ATAC-seq), gene expression (RNA-seq) and DNA methylation (WGBS), a *bedmap* function from BEDOPS software (Neph et al., 2012) was applied to get the mean scoring peak over the 100Kb intervals genome-wide per each layer of information. Next, Pearson correlation coefficients were computed between the EV coefficients and the mean scoring value of each epigenetic mark at 100Kb intervals (**Figure 24** from **Study 1**). Finally, the mean scoring values were normalized by the total sum of the values for each mark and grouped by the three defined genomic compartments (A, I, B-type; **Figure 27A** from **Study 1**). A Wilcoxon test was used to compute the significance between all the possible pairwise comparisons of the signal distribution.

4.5 Compartment interaction score (C-Score)

The compartment score is defined as the ratio of contacts between regions within the same compartment (intra-compartment contacts) over the total chromosomal contacts per compartment (intra-compartment + inter-compartment). To compute the compartment score, all the compartments that shared the same genomic segmentation were merged.

4.6 Chromatin states enrichment in genomic compartments

The genome was segmented into 12 different chromatin states at 200bp interval as previously described (Beekman et al., 2018a). In **Study 1**, the active promoter and strong enhancer1 were merged as a unique state, giving a total of 11 chromatin states. To obtain chromatin states enrichments at the different compartments, first the genome compartmentalization was next split into 4 groups; 3 conserved groups, in which the B-cell samples shared A-type compartment (n=6,409), B-type compartment (n=6,267) or I-type compartment (n=5,467) and a dynamic group (n=7,099) of non-conserved compartmentalization among B-cell subpopulations. Each group was correlated with the defined 11 chromatin states using *foverlaps* function from *data.table* R package. The frequency of each chromatin state (corrected by the total frequency in the genome) was computed per each genomic compartment. The chromatin state score is thus the median frequency of the three replicas scaled by the columns and the rows using *scale* function from *base*R package.

4.7 Dynamics of chromatin states in compartments during B-cell maturation

To study the chromatin states dynamics upon B-cell differentiation, the B-cell differentiation axis was divided into two main branches: (i) NBC-GCBC-PC and (ii) NBC-GCBC-MBC. The 5,445 common compartments from both branches were considered for the analysis. The general modulation of chromatin structure was drawn using the *alluvial* function from *alluvial* R package.

4.8 Significant changes in compartment regions

The detection of significant changes among compartments was performed as following. Firstly, the 100Kb regions that had at least one missing value among the compared samples were removed from the analysis. Then, two different groups were defined, case and control, according to the case-control pair analyzed. The four groups analyzed were:

(I) control: all regions conserved across all B-cell samples without missing values in CLL (A-type, n=3,967, I-type, n=4,301 and B-type, n=5,226), case: all CLL regions non-conserved in B-cell samples (n=3,217).

(II) control: all regions conserved across all B-cell samples without missing values in MCL (A-type n=6,167, I-type n=5,299, B-type n=5,812), case: all MCL regions non-conserved in B-cell samples (n=4,716).

(III) control: B cell-CLL significantly changed regions (n=348) - MCL-CLL overlapping (n=31) = B cell-CLL specific regions (n=317), case: MCL regions (A-type n=97, I-type n=154, B-type n=61; total n=312).

(IV) control: B cell-MCL significantly changed regions (n=82) - MCL-CLL overlapping (n=31) = B cell-MCL specific regions (n=51), case: CLL regions (n=41).

A t-test was computed to compare each case-control pair, and the resulting p-values were adjusted using the false discovery rate (FDR) (Benjamini and Hochberg, 1995). The regions with significantly different means and fold changes were selected based on two specific thresholds: a p-adjustment value less than 0.05 and a fold change greater than 0.4.

4.9 Log-ratio of normalized interactions in the *AICDA* regulatory landscape

To represent the AICDA regulatory region, normalized Hi-C maps were analyzed at 50Kb of resolution at the specific genomic region, chr12:8,550,000-9,050,000 (GRCh38), from the four B-cell subpopulations. A logarithmic ratio of the contact maps was computed between NBC and GCBC and GCBC with PC and MBC. The result array was convolved with a 1-dimensional Gaussian filter of standard deviation (sigma) of 1.0 using and interpolated with a nearest-neighbor approach using *scipyndimage* Python package.

4.10 Integrative 3D modelling of EBF1 and structural analysis

To represent the EBF1 regulatory region, Hi-C interactions matrices from the merging of three replicas of NBC and the seven cases of CLL were used to model chr5:158,000,000-160,000,000 (GRCh38) at 5Kb of resolution. For NBC and CLL merged Hi-C interaction maps, a matrix modeling potential (MMP) score was calculated, resulting in 0.79 for NBC and 0.84 for CLL indicative of good quality Hi-C contact maps for accurate 3D reconstruction (Trussart et al., 2015). Next, this region was modelled using a restraint-based modelling approach as implemented in TADbit (Serra et al., 2017), where the experimental frequencies of interaction are transformed into a set of

spatial restraints (Baù and Marti-Renom, 2012). Briefly, each 5Kb bin of the interaction Hi-C map was represented as a spherical particle in the model, which resulted in 400 particles each of radius equal to 25nm. All the particles in the models were restrained in the space based on the frequency of the Hi-C contacts, the chain connectivity and the excluded volume. The TADbit optimal parameters (maxdist=-1.0; lowfreq=1.0; upfreq=200; and dcutoff=150) resulted in the best Spearman correlations of 0.61 (NBC) and 0.63 (CLL) between the Hi-C interaction map and the models contact map. Next, a total of 5,000 models per cell type were generated, and the top 1,000 models that best statisfied the imposed restraints were retained for the analysis. To assess the structural similarities among the 3D models, the distance root-mean-square deviations (dRMSD) value was computed for all the possible pairs of top models (1,000 in NBC and 1,000 in CLL) and a hierarchical clustering algorithm was applied on the resulting dRMSD matrix using *ward.D* method from stats package in R (**Figure 41C** from **Study 1**). The convex hull volume spanned by the 81 particles of the *EBF1* gene (chr5:158,695,000-159,000,000; GRCh38) was computed in each model using the *convexhull* function from the *scipy.spatial* Python package (**Figure 41D** from **Study 1**).

4.11 Defining *de novo* (in)active regions in sub-type specific neoplastic group

MCL and CLL patient samples were grouped according to their biological and clinical characteristics (3 nnMCL, 2 cMCL, 2 uCLL and 5 mCLL). Then, the non-assigned neoplasia compartments were removed from the analysis. A sample homogenization was applied to reduce the intra-subtype variance; the samples that presented a difference of EV smaller than 0.4 were retained (91.29% in MCL and 87.1% in CLL). Next, to study the inter-subtype variance, the mean of the EV from each subtype of B-cell malignancy was computed. Significant regions were determined if the difference between the two subtypes (cMCL vs nnMCL and uCLL vs mCLL) was equal or higher than 0.4. MCL-subtype specific regions where split into two groups according to the value of its EV coefficient. The distribution and the frequency of the significantly changed regions were studied per chromosome and compared with an indiscriminate probability. To do so, n-subsamples (n=number of target regions per group) of 100Kb size were selected from the GRCh38 genome and their frequency was calculated per chromosome (this process was randomly repeated 10,000 times). One tailed Monte-Carlo method was applied to compute *p*-values.

4.12 Defining MCL structural variants at TAD borders

TADs were called using TADbit (Serra et al., 2017) on the raw Hi-C matrices at 50Kb resolution. To assess the TAD border relationship with the structural variant break coordinates in the five different MCL cases (**Study 2**), a window of 50Kb (25Kb upstream and 25Kb downstream) from the TAD border coordinate was considered. The *findOverlaps* function from *GenomicRanges* R package (Lawrence et al., 2013) was used to annotated the structural variants that overlapped with the defined windows. Besides, to further characterize TAD borders at the *CCND1* region, TADpole (Soler-Vila et al., 2020) was applied on the 20Kb raw Hi-C interaction matrices of NBC and MBCsand the five different MCL cases.

5. 4C-sequencing data processing

Samples for 4C experiments were sequenced with the MiSeq instrument (Illumina) using 50bp single-reads, adding 5% PhiX. 4C-seq analysis was performed using the 4C-seq pipeline 4cseqpipe (http://compgenomics.weizmann.ac.il/tanay/?page_id=367). The R statistical package version 3.1.1 was used for these analyses.

6. ChIP-seq and ATAC-seq data generation and processing

6.1 Histone ChIP-seq and ATAC-seq

ChIP-seq of six different histone marks and ATAC-seq data were generated as described in (http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58) (Beekman et al., 2018a). Briefly, fastq files of ChIP-seq data were aligned to the GRCh38 reference genome using bwa 0.7.7 (Li and Durbin, 2009), PICARD (http://broadinstitute.github.io/picard/) and SAMTOOLS (Li et al., 2009), and wiggle plots were generated (using *PhantomPeakQualTools* R package) as described (http://dcc.blueprint-epigenome.eu/#/md/methods). Peaks of the histone marks were called as described in http://dcc.blueprint-epigenome.eu/#/md/methods using MACS2 (version 2.0.10.20131216) (Zhang et al., 2008b) with input control. ATAC-seq fastq files were aligned to genome build GRCh38 using bwa 0.7.7 (parameters: -q 5 –P -a 480) (Li and Durbin, 2009) and SAMTOOLS v1.3.1 (default settings) (Li et al., 2009). BAM files were sorted and duplicates were masked using PICARD tools v2.8.1 with default settings (http://broadinstitute.github.io/picard/). Finally, low quality and duplicate

reads were removed using SAMTOOLS v1.3.1 (parameters: -b -F 4 -q 5,-b, -F 1024) (Li et al., 2009). ATAC-seq peaks were determined using MACS2 (version 2.1.1.20160309, parameters: -g hs q 0.05 -f BAM –nomodel - shift -96 extsize 200 - keep -dup all) without input (Zhang et al., 2008b).

In **Study 1**, two sets of consensus peaks (chr1-22) were generated by merging the locations of the separate peaks per individual sample. One with all B-cell subpopulations (n=12 biologically independent samples for histone marks and n=15 biologically independent samples for ATAC-seq) and another with the B-cell subpopulations and B-cell neoplasia samples (CLL n=7 and MCL n=5 biologically independent samples). For the histone marks, the number of reads per sample per consensus peak was calculated using the *genomecov* function of bedtools suite (Quinlan and Hall, 2010). For ATAC-seq, the number of insertions of the TN5 transposase per sample per consensus peaks was calculated determining the estimated insertion sites (shifting the start of the first mate 4bp downstream), followed by the *genomecov* function of bedtools suite (Quinlan and Hall, 2010). Using DESeq2 R package (Love et al., 2014), counts for all consensus peaks were transformed by means of the variance stabilizing transformation (VST) with blind dispersion estimation. PCAs were generated with the *prcomp* function from the stats package in R using the VST values.

In **Study 3**, the read density of H3K27ac, H3K4me1 and H3K4me3 were visualized with seqMiner (Ye et al., 2014).

6.2 PAX5 ChIP-qPCR

ChIP for PAX5 was performed as the described RNAPII protocol (Stock et al.) with some adaptations. Briefly, chromatin of ten million cells was crosslinked, sonicated and incubated with four micrograms of PAX5 antibody (C-20, sc-1975, polyclonal goat IgG) and used as control the polyclonal goat IgG (sc-2028, normal goat IgG). Antibody complexes were recovered with 30µl of Protein G Dynabeads (ThermoFisher) through a co-incubation for 2 hours at 4^0C. Once the chromatin was washed, eluted and decrosslinked, it was finally purified using Diagenode columns (MicroChIP DiaPure columns). A sheared 1% DNA was used as input control on the qPCR validation and for analyzing sonication efficiency.

7. RNA-seq data generation and processing

Single-stranded RNA-seq data were generated as previously described (Ecker et al., 2017). Briefly, RNA was extracted using TRIZOL (Life Technologies) and libraries were prepared using TruSeq Stranded Total RNA kit with Ribo-Zero Gold (Illumina). Adapter-ligated libraries were amplified and sequenced using 100bp single-end reads. RNA-seq data aligned to the reference human genome build GRCh38 (**Appendix Table 4**). Signal files were produced and gene quantifications (gencode 22, 60,483 genes) were calculated as described (http://dcc.blueprint-epigenome.eu/#/md/methods) using the GRAPE2 pipeline with STAR-RSEM profile (adapted from the ENCODE Long RNA-Seq pipeline). The expected counts and fragments per kilobase million (FPKM) estimates were used for downstream analysis. The PCA of the RNA-seq data was generated with the *prcomp* function from the stats package in R.

8. DNA methylation data generation and processing

8.1 Whole-genome bisulfite sequencing

WGBS was generated as previously described (Kulis et al., 2015). Briefly, genomic DNA (1–2µg) was sheared and fragments of size 150-300bp were selected using AMPure XP beads (Agencourt Bioscience). Libraries were constructed using the Illumina TruSeq Sample Preparation kit (Illumina Inc.). After adaptor ligation, the DNA was treated with sodium bisulfite using the EpiTexy Bisulfite kit (Qiagen). Two rounds of bisulfite conversion were performed to ensure a conversion rate of over 99%. Enrichment for adaptor-ligated DNA was carried out through seven PCR cycles using the PfuTurboCx Hotstart DNA polymerase (Stratagene). Paired-end DNA sequencing (2x100bp) was then performed using the Illumina Hi-Seq 2000. Mapping and determination of methylation estimates were performed as described (http://dcc.blueprint-epigenome.eu/#/md/methods) using GEM3.0.

In **Study 1**, only methylation estimates of CpGs with ten or more reads per sample were used for downstream analysis. Two common sets (chr1-22) were generated one for all normal B-cell samples and another taking into account also B-cell neoplasias. In **Study 3**, a common set was generated taking into account two MCL samples and B-cell subpopulations.

8.2 DNA methylation analysis with 450K arrays

Data from 450K arrays was performed by firslty extracting genomic DNA using the QIAmp DNA kit (Qiagen) and DNA samples quality was assessed by SYBR green staining on agarose gels quantified using the Nanodrop ND-1000 spectrophotometer (Thermo scientific). The EZ DNA Methylation Kit (Zymo Research) for bisulfite conversion was used to convert 500ng genomic DNA per sample. Bisulfite-converted DNA was hybridized onto the HumanMethylation 450K BeadChip kit (Illumina) which covers 99% of RefSeq genes and 96% of CpG islands. The Infinium methylation assay was carried out as previously described (Bibikova et al., 2009, 2011).

Data from the 450K arrays were analyzed in R using the minfi package (version 1.18.2) (Aryee et al., 2014), available through the Bioconductor open source software. To exclude technical and biological biases that might produce false results in further analyses, we developed and optimized an analysis pipeline with several filters (i.e. discarding CpGs with low detection P values, sex-specific CpGs, CpGs showing individual-specific methylation and CpGs overlapping with SNPs). Taking into account the different performance of Infinium I and Infinium II assays we used the subset-quantile within array normalization (SWAN) (Maksimovic et al., 2012) that corrects for the technical differences between the Infinium I and II assay designs and produces a smoother overall beta value distribution.

8.3 Deconvolution and adjustment of DNA methylation estimates

A statistical approach in **Study 3** was tuned to subtract the DNA methylation estimates of the non-tumoral fraction in MCL samples as detailed described (Duran-Ferrer et al., 2017), by adapting the algorithm (Houseman et al., 2012; Jaffe and Irizarry, 2014) to estimate the proportion of B cells CD8$^+$ T cells, CD4$^+$ T cells, NK cells, monocytes and granulocytes in MCL samples.

8.4 Bisulfite pyrosequencing

Bisulfite pyrosequencing analysis were performed as previously described (Tost and Gut, 2007). Briefly, genomic DNA was bisulfite converted using EpiTectPlus Bisulfite Conversion Kit (Qiagen, Germany) according to manufacturer's specifications. Subsequent PCR amplification was performed using biotinylated primers.

Pyrosequencing and data analysis were performed with the pyrosequencer analyzer PyroMark Q96 (Qiagen, Germany) according to manufacturer's instructions. The CpGs analysed and the primer sequences used to study the SOX11-positive MCL enhancer region are detailed in the table below (**Table 2**).

CpG coordinate (GRCh37)	Amplification primer (fw)	Amplification primer (rv)	Sequencing primer
chr2:6477577	/5Biosg/TTTATTGTTTTATAGTAAGGGTAGAG	CTAATCAAATACTCCCTAACC	AAAAATTCTAAATAATAACTCCTAC
chr2:6477615	/5Biosg/TTTATTGTTTTATAGTAAGGGTAGAG	CTAATCAAATACTCCCTAACC	AAAAATTCTAAATAATAACTCCTAC
chr2:6484702	GAGAAGTTGGTTTTAATGAGATTTAGTAGT	/5Biosg/CAAAAAAAAACCTTTAAAAAACAATACACC	GTTAATTTAAGTGGTTTTTGTTAT
chr2:6484925	ATTGTTGGAGATATGAGAAGTGT	/5Biosg/CCAAAACCTCATCTAAACCTACTTATTTC	ATTGTATTTTTTGAAGTTTTAATT

Table 2. CpGs analysed and primers used for bisulfite sequencing to analyze the SOX11-positive MCL enhancer region (**Study 3**).

8.5 Differential DNA methylation analyses

Distinguished approaches were employed, based on the methodology of DNA methylation analyses, to define regions with differential methylation between different MCL subgroups or between MCLs and normal B-cell controls (**Study 3**).

For 450K array data, it was considered (1) an absolute difference of mean DNA methylation levels of at least 0.25 between the compared groups and (2) a false discovery rate (FDR) of less than 0.05 using a Wilcoxon test for independent samples.

Whereas WGBS analyses were performed thanks to the collaboration with the Centro Nacional de Análisi Genómico (CNAG, Barcelona, Spain). Briefly, two different strategies were used. On the one hand, differentially methylated CpGs (DMCs) were identified in a pair-wise comparison of each MCL sample versus HPCs, and the 2 MCL samples versus each other. Statistical significance difference in DNA methylation was estimated based in beta-binomial distribution using the "bdiff" algorithm (Raineri et al., 2014), and a DNA methylation difference of >0.25. Annotation of CpGs was performed using the UCSC Table Browser GRCh37/hg19 version (Karolchik et al., 2004) and considering each feature related to a gene. On the other hand, differentially methylated regions (DMRs) were calculated using a Hidden Markov Model (HMM) to

segment the methylation values in the two samples under consideration. The Markov Model has 3 states corresponding to low, intermediate and high methylation; the transition probabilities are 0.9 for staying in the same state and 0.1 to change state. The emission probabilities of the HMM are the probabilities of obtaining the observed count of non-converted and converted reads assuming an underlying methylation value of less than 0.3 ("low" state), between 0.3 and 0.7 ("intermediate" state) and higher than 0.7 ("high" state). Stretches in the genome which correspond to the first samples being in "high" state and the second sample being in "low" state (or vice versa) are candidates for being DMRs. These regions were further filtered by imposing that they should contain more than 3 CpGs. Moreover, in those regions the difference in average methylation should be larger than 0.25.

Both in the case of WGBS and 450K arrays, differentially methylated sites were classified into i) B cell-related CpGs (those whose DNA methylation level is modulated during the normal B-cell differentiation), or ii) B cell-independent CpGs (those whose DNA methylation levels do not change during normal B-cell differentiation). This classification was made based on published DNA methylation data from normal B-cell subpopulations (Kulis et al., 2015).

In the case of WGBS, DMRs were further classified based on their composition of B cell-related and B cell-independent CpGs. Three classes of DMRs were defined: i) B cell-related DMR (all CpGs within the DMR are B cell-related), ii) B cell-independent DMR (all CpGs in the DMR are B cell-independent) and iii) Mixed DMR (the DMR is formed by both B cell-related and B cell-independent CpGs).

8.6 Genomic and functional annotation of CpGs

CpGs analysed by WGBS and 450K array data were annotated using the UCSC Genome Browser database (GRCh38 but for **Study 3** GRCh37). For the location relative to a gene, the following categories were used: (i) TSS 1500 (from 201 to 1,500bp upstream of the transcriptional start site (TSS)), (ii) TSS 200 (from 1 to 200bp upstream of the TSS), (iii) 5' UTR, (iv) first exon, (v) gene body (from the first intron to the last exon), (vi) 3' UTR and (vii) intergenic regions. Owing to the presence of alternative transcription start sites and regions containing more than one gene, some of the CpGs were assigned multiple annotations. For the location relative to a CpG island (CGI), the

following groups were used: (i) within CGI, (ii) in CGI shore (0–2 kb from the CGI edge), (iii) in CGI shelf (>2 kb to 4 kb from the CGI edge) and (iv) outside CGI.

8.7 Correlation of DNA methylation with somatic hypermutation levels

In **Study 3**, for C1 and C2 MCLs separately, DNA methylation levels of each CpG and the level of somatic hypermutation (SHM) were correlated using spearman correlation tests using the rcorr function of the Hmisc R package. SHM levels ranged from 0 (no SHM, i.e. 100% germline identity) to 12.32 (high level of SHM, 87.68% identity with germline). Significant correlations between CpG methylation and SHM were detected as presenting an absolute correlation coefficient >0.4 and an FDR-value<0.05. The R statistical package version 3.1.1 was used for these analyses.

9. Gene expression analyses

9.1 Differential gene expression analyses

Differentially expressed genes based on RNA-seq were defined using the DEseq2 R package (Love et al., 2014), *nbinomWaldTest*, in all the genes. Then, the genes present on the compartments of interest were selected and Benjamini y Hochberg (BH) test (FDR<0.05) was applied. Specifically, the expression of the genes differentially expressed per each comparison of interest was assessed using the expected counts. Only genes that were expressed (FPKM median values>1) were included. The *findOverlaps* function from *GenomicRanges* R package (Lawrence et al., 2013) was used to annotated genes that overlapped with these defined regions. One tailed Monte-Carlo method was applied to evaluate the significant number of differentially expressed genes in CLL-specific compartments (this process was randomly repeated 10,000 times).

Differentially expressed genes on microarrays were normalized using the R *frma* (McCall et al., 2010) method and *limma* R package (Smyth, 2004) was used to identify differentially expressed genes with adjusted *p*-value<0.05. Standardized expression matrices were used to do the heatmaps using *pheatmap* R package.

9.2 Analysis of the proliferation signature

Gene expression data from arrays of 25 MCL patients for which DNA methylation data was available were mined in **Study 3** (Navarro et al., 2012). The expression levels of 18 genes were used to calculate the proliferation signature (Rosenwald et al., 2003), (see also **Figure 71C** from **Study 3**). To that end, the mean of the rma normalized expression values of these 18 genes was calculated per sample. To calculate the relative difference of the proliferation signature compared to the group average, the average proliferation signature value of the group was subtracted from the proliferation signature per case. This average proliferation signature was then correlated with the number of DNA methylation changes in cases from C1 and C2 MCLs.

10. Other bioinformatics analyses

10.1 Gene ontology analysis

The KEGGprofile package (Zhao et al., 2015) available through Bioconductor was used to determine the enrichment of individual ontology terms in the differentially methylated CpGs between C1 and C2 MCLs as compared to all the genes analyzed in the 450K array (**Study 3**). Significant terms were considered when adjusted *p*-value<0.05.

10.2 Analysis of superenhancers

Superenhancers in the two representative MCL cases were detected using the ROSE software in **Study 3** (Lovén et al., 2013; Whyte et al., 2013) . Thereby, the H3K27ac peaks determined by MACS2 were used as input, as well as the mapped reads of the H3K27ac ChIP-seq experiments and their respective input controls to correct for background levels. The H3K27ac peaks present at transcription start sites were removed from this analysis, by turning on the option –t in the ROSE software.

10.3 Transcription factor binding motif analyses

From targeted regions, the FASTA sequences were extracted using *getfasta* function from *bedtools* suite (Quinlan and Hall, 2010) using GRCh38 as reference assembly. An analysis of motif enrichment was done by the *AME-MEME* suite (McLeay and Bailey, 2010) using non-redundant transcription factor (TF) binding profiles of

Homo sapiens Jaspar 2018 database (Khan et al., 2018) as a reference motif database. This database contained a set of 537 DNA motifs. Maximum odd scores were used as a scoring method and one-tailed Wilcoxon rank-sum as motif enrichment test. Only TF genes that were expressed (FPKM median values>1) were included.

Besides, a search of individual motif occurrences analysis (i.e. TCF4) was done using *AME-FIMO* suite (Grant et al., 2011) library (BSgenome.Hsapiens.UCSC.hg38, masked) with a custom random model (letter frequencies: A, 0.262: C, 0.238: G, 0.238 and T, 0.262). A *p*-value<0.0001 was established as a threshold to determine significant motif occurrences.

10.4 Overall survival and prognosis analysis

Univariate and multivariate survival analyses were used to measure the impact of DNA methylation changes in the clinical behavior of MCL patients in **Study 3**.

To detect MCL groups with different clinical behavior based on their DNA methylation changes, we used the "maxstat" package from R software. This analysis allows us to detect the most suitable threshold to separate MCL groups with differences in overall survival. Overall survival Kaplan-Meier plots and long-rank tests were performed with the IBM-SPSS Statistics version 20.

A multivariate analysis was performed to determine whether the epigenetic burden is an independent predictor of prognosis, using the *coxph* function (Survival package, R software). The multivariate Cox Regression model was generated using as input variables: age (quantitative), morphology (classical/blastoid), IGHV somatic hypermutation (quantitative), presence of mutations (positive/negative, based on the analysis of the 6 most frequent drivers), SOX11 expression (positive/negative) and number of DNA methylation changes (quantitative).

11. Reverse ChIP

Reverse ChIP was performed as previously described (Belver et al., 2019; Unnikrishnan et al., 2016). Briefly, DNA bait sequences on the SOX11 superenhancer chromatin accessible peaks were generated by PCR from human genomic DNA. Two specific chromatin accessible peaks were studied and for each one biotinylated forward primer and an unmodified reverse primer were designed, **Table 3**. Then, DNA baits

were conjugated to streptavidin beads and incubated with nuclear protein extracts from the SOX11-positive MCL cell line, Z-138. Non-conjugated beads were used as negative control. SOX11-specific peak pulled down proteins were analyzed by mass spectrometry at the Proteomics Laboratory at the New York University School of Medicine.

Target	Forward primer	Reverse primer	Coordinates of the region (GRCh38)
Peak 1	TCAGCAAGCTGGCTGTAGCC	TCAGCTTCCTTAGTTTAGTTTGGA	chr2:6336962-6337881
Peak 3	AGGATGTCACAAAAAGGTGC	GAACGTCAGATCACATGTGC	chr2:6344929-6346473

Table 3. Primers used for reverse ChIP in **Study 4**.

The MS/MS spectra were searched against the Uniprot human reference proteome database using Sequest within Proteome Discoverer. A 1% false discovery rate (FDR) cut off was applied on the peptide levels using a standard target-decoy database strategy. All proteins identified with less than two unique peptides were excluded from analysis. The initial lists containing 383 proteins for peak 1 and 428 proteins for peak 3 were filtered based on the expression of the corresponding coding genes in SOX11-positive MCL cases. To do so, RNA expression data from two conventional MCL cases from the BLUEPRINT dataset (FPKM median values>1) were used, resulting in 368 proteins for peak 1 and 404 proteins for peak 3. Those proteins were further cleaned up using the contaminant repository for affinity purification, CRAPome contaminant list (CRAPome databaseH. sapiens_V 1.1matrix format with 411 experiments). The proteins present in more than 10% of the CRAPome experiments were discarded. For peak 1, 263 proteins were identified being 87 specific for that peak in comparison with peak 3, where 258 proteins were identified, from which 84 proteins were specific for this peak 3.

STUDY 1

Mapping the three-dimensional genome dynamics in normal and neoplastic human B cells

1. Introduction

Over the last decades, our understanding of higher-order chromosome organization in the eukaryotic interphase nucleus and its regulation of cell state, function, specification and fate has profoundly increased (Rowley and Corces, 2018; Szalaj and Plewczynski, 2018).

Chromatin conformation capture techniques have been used to elucidate the genome compartmentalization (Dekker et al., 2002b; Denker and de Laat, 2016). It is widely accepted that the genome is segregated into two large compartments, named A-type and B-type (Lieberman-Aiden et al., 2009), which undergo widespread remodeling during cell differentiation (Andrey and Mundlos, 2017; Dixon et al., 2015; Peric-Hupkes et al., 2010; Stadhouders et al., 2018; Szalaj and Plewczynski, 2018). However, the majority of the studies have been performed using cell lines, animal models or cultured human cells (Dixon et al., 2015; Hu et al., 2018; Johanson et al., 2018; Schmitt et al., 2016b; Stadhouders et al., 2018), and although few analyzed sorted cells from healthy human individuals (Bunting et al., 2016; Javierre et al., 2016), there is limited information regarding 3D genome dynamics across the differentiation program of a single human cell lineage (Bunting et al., 2016).

Normal human B-cell differentiation is an ideal model to study the dynamic 3D chromatin conformation during cell maturation, as these cells show different transcriptional features and biological behaviors, and can be accurately isolated due to their distinct surface phenotypes (Kurosaki et al., 2010; Matthias and Rolink, 2005). Moreover, how the 3D genome is linked to cancer development using primary samples from patients is also widely unknown (Li et al., 2018). In this context, several types of neoplasms can originate from B cells at distinct differentiation stages (Swerdlow et al., 2017). Out of them, chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL) are derived from mature B cells and show a broad spectrum of partially overlapping biological features and clinical behaviors (Puente et al., 2018). Both diseases can be categorized according to the mutational status of the immunoglobulin heavy chain variable region (IGHV), a feature that seems to be related to the maturation stage of the cellular origin (Chiorazzi and Ferrarini, 2011). CLL cases lacking IGHV somatic hypermutation are derived from germinal center-independent B cells whereas CLL with mutated IGHV derive from germinal center-experienced B cells (Kipps et al., 2017). In CLL, this variable is strongly associated with the clinical features of the

patients, with mutated IGHV (mCLL) cases correlating with good prognosis and those lacking IGHV mutation (uCLL) with poorer clinical outcome (Kipps et al., 2017). In MCL, although two groups based on the IGHV mutational status can be recognized and partially correlate with clinical behavior, other markers such as expression of the *SOX11* oncogene are used to classify cases into clinically-aggressive conventional MCL (cMCL, SOX11-positive MCL) and clinically-indolent non-nodal leukemic MCL (nnMCL, SOX11-negative MCL) (Jares et al., 2012b; Navarro et al., 2012; Puente et al., 2018; Royo et al., 2012).

From an epigenomic perspective, previous reports have identified that B-cell maturation and neoplastic transformation to CLL or MCL entail extensive modulation of the DNA methylome and histone modifications (Beekman et al., 2018a; Kulis et al., 2012, 2015; Oakes and Martin-Subero, 2018; Oakes et al., 2016; Queirós et al., 2016). For instance, in our group, we have identified that DNA methylation levels at enhancer regions are widely modulated both during normal B-cell differentiation (Kulis et al., 2015) and in the aforementioned B-cell tumors (Beekman et al., 2018a; Kulis et al., 2012; Queirós et al., 2016). However, whether such epigenetic changes are also linked to modulation of the higher-order chromosome organization is yet unknown (Johanson et al., 2019). The aim of this part of my doctoral thesis was to untackle the 3D genome architecture of normal and neoplastic B cells, and identify the chromosome organization modulation during human B-cell maturation and neoplastic transformation.

2. Results

2.1 Sample recruitment

To decipher the 3D genome architecture of normal human B cells across their maturation program, different normal B-cell subpopulations were sorted from two distinct sources, using samples from healthy male donors, as previously described (Kulis et al., 2015) (**Figure 20A** and **20B**).

Three biological replicates of each B-cell subpopulation were processed i.e. naive B cells (NBC) and memory B cells (MBC) were sorted from peripheral blood while germinal center B cells (GCBC) and plasma cells (PC) were sorted from tonsils. MBC contained both class-switched (IgA$^+$/IgG$^+$) and non-class switched (IgM$^+$/IgD$^+$) MBC

subtypes. Each of the replicates were derived from individual male donors with the exception of plasma cells, for which two of the three replicates were derived from a pool of four different donors.
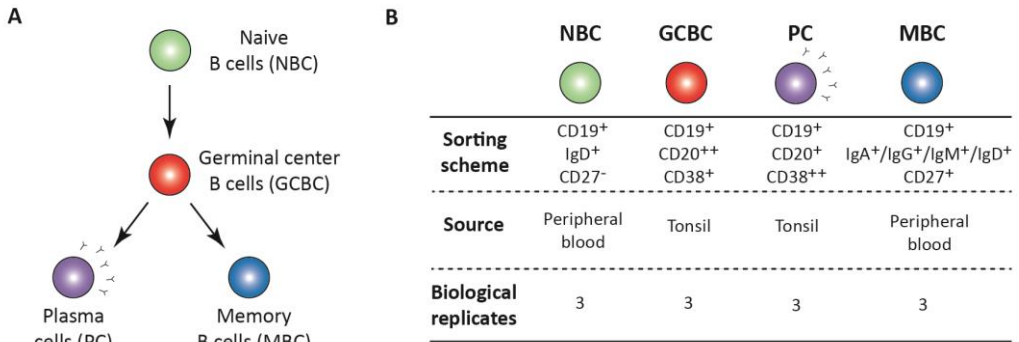
Figure 20. Study design addressing normal B-cell differentiation. A. Schematic overview of mature B-cell differentiation showing the four B-cell subpopulations considered in this study. **B**. Sample description and *in situ* Hi-C sequencing experimental design for normal B-cell differentiation subpopulations. NBC, naive B cells; GCBC, germinal center B cells; MBC, memory B cells and PC, plasma cells.

In this study, I also analyzed the 3D genome organization upon neoplastic transformation. To address this, I performed *in situ* Hi-C in thoroughly characterized tumor cells from patients with CLL (n=7) or MCL (n=5). Within each neoplasm, cases of two subtypes were included, mCLL (n=5) and uCLL (n=2) as well as cMCL (n=2) and nnMCL (n=3) (**Figure 21** and **Appendix Table 1**).
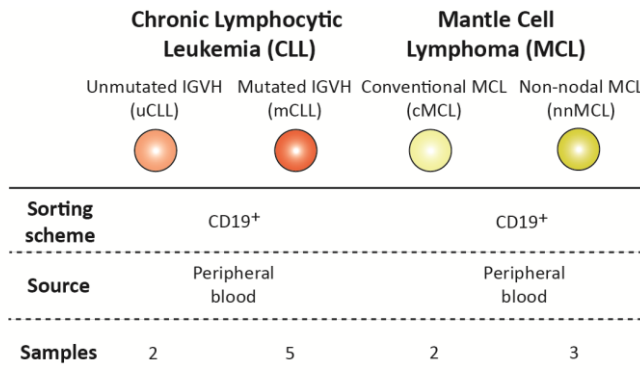
Figure 21. Study design addressing neoplastic B cells. Sample description and *in situ* Hi-C sequencing experimental design in CLL and MCL cases.

2.2 Chromosome conformation maps during normal B-cell differentiation

I used in *situ* Hi-C to generate genome-wide chromosome conformation maps spanning the B-cell lineage. I optimized this technique in our laboratory using the knowledge acquainted during the previous optimization of the original Hi-C protocol, also known as dilution Hi-C. In fact, *in situ* Hi-C is a combined approach of the original Hi-C protocol with a nuclear ligation assay, which allows to generate denser Hi-C maps (Rao et al., 2014). This protocol enables higher resolution maps and it uses much lower amount of cells. Instead of the twenty-five million cells needed for a Hi-C experiment, I could lower the amount to two million cells using *in situ* Hi-C protocol, which is the minimum number of starting material described for this technique. This was an important advantage as higher cell numbers of a single B-cell subpopulation were difficult to obtain from a single healthy donor. Once the samples were processed and sequenced, I mapped the reads and filtered those resulting from unspecified ligations, error or experimental artifacts. I ended up with around 100 million reads per B-cell replicate. Hi-C experimental quality metrics can be found at **Appendix Table 2** and **3**. I initially explored the intra- and inter-subpopulation variability and observed that the Hi-C replicate were concordant, as quantified measuring and clustering the reproducibility score (RS) (**Figure 22A** and **22B**). This metric is a measure of matrix similarity ranging between 0 (totally different matrices) and 1 (identical matrices) (Yan et al., 2017), **Appendix Figure 1**.
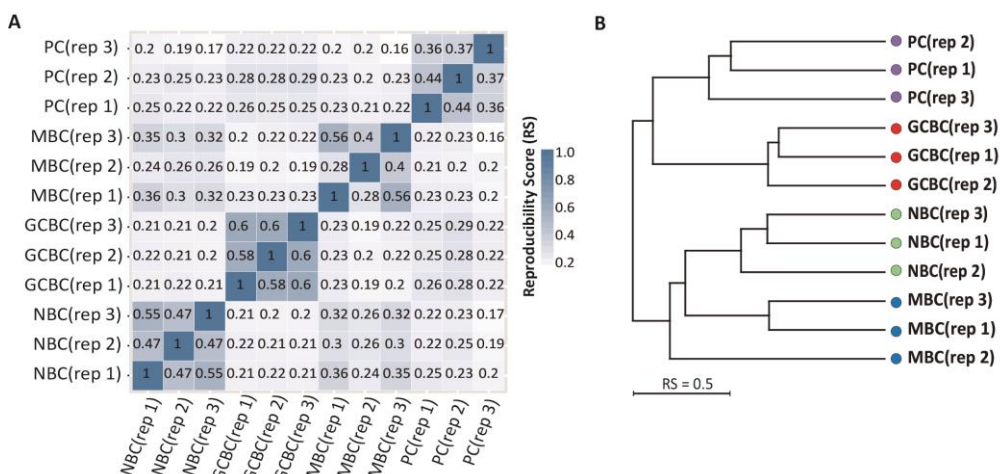
Figure 22. Similarity between normal B-cell replicates. Average genome-wide matrix (**A**) and

dendrogram (**B**) of the reproducibility score of B-cell subpopulation replicates for normalized Hi-C contact maps at 100Kb resolution.

Furthermore, the comparison of samples suggests that the overall genome architecture of NBC is more similar to MBC, and clearly different from GCBC and PC, which belong to a different cluster (**Figure 22B**). Based on these results, all biological replicates per B-cell subpopulation were merged for further analyses, resulting in interaction Hi-C maps with around 300 milionvalid reads per cell subpopulation.

2.3 Multi-omics analysis upon B-cell differentiation

Chromosomal architecture has been analyzed together with nine additional omic layers generated as part of the BLUEPRINT consortium, on the same cell types with the same sorting scheme (Adams et al., 2012; Beekman et al., 2018a). Specifically, data was obtained from chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) of six histone modification with non-overlapping functions (H3K4me3, H3K4me1, H3K27ac, H3K36me3, H3K9me3 and H3K27me3), assay of transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) assessing chromatin accessibility, whole genome bisulfite sequencing (WGBS) to assess DNA methylation and RNA-seq data assessing gene expression, **Appendix Table 4**.

The first component of the principal component analysis (PCA) of histone modifications, chromatin accessibility and gene expression reflected a resemblance between NBC and MBC separated from GCBC and PC, as previously observed when analyzing the chromatin architecture (**Figure 22B** and **23**). In contrast to other omics marks, the first component of DNA methylation data results in a division of GCBC, MBC and PC separated from the NBC (**Figure 23**). These analyses suggest fundamental differences between chromatin-based epigenetic marks, including chromosome conformation data, and DNA methylation. In fact, changes in DNA methylation linearly accumulate throughout B-cell maturation (Kulis et al., 2015; Oakes et al., 2016), which explains the clear differences between NBC and MBC in spite of their converging transcriptomes.
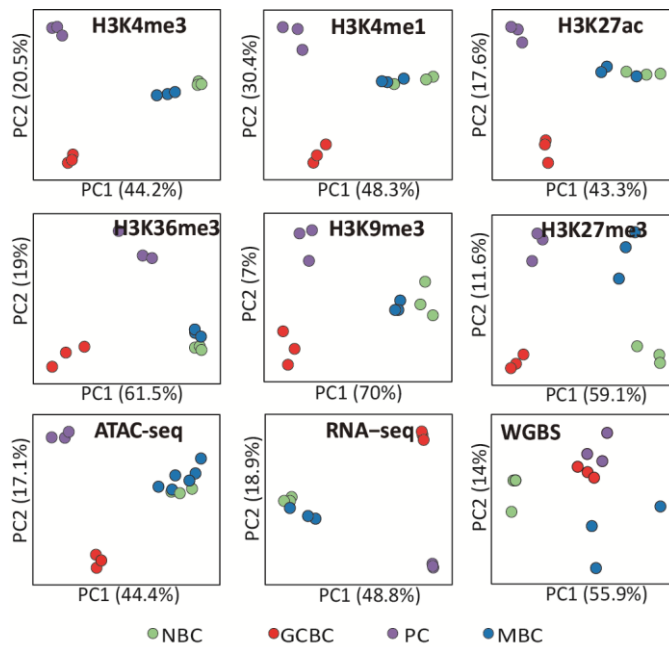
Figure 23. Multi-omics view upon B-cell differentiation. Unsupervised principal component analysis (PCA) for nine omics layers: ChIP-seq of six histone marks (H3K4me3 n=46,184 genomic regions, H3K4me1 n=44,201 genomic regions, H3K27ac n=72,222 genomic regions, H3K36me3 n=25,945 genomic regions, H3K9me3 n=40,704 genomic regions, and H3K27me3 n=20,994 genomic regions), chromatin accessibility measured by ATAC-seq (n=99,327 genomic regions), DNA methylation measured by WGBS (n=15,089,887 CpGs) and gene expression measured by RNA-seq (n=57,376 transcripts). Three independent biological replicates of NBC, GCBC, PC, and MBC were studied for all omic layers, with the exception of ATAC-seq for which six biological replicates of MBC were used.

2.4 Compartment definition in B-cell subpopulations

To study the compartmentalization of the genome during B-cell differentiation, Hi-C interactions maps at a 100Kb resolution were segmented into positive and negative eigenvalues based on the eigenvector decomposition (Imakaev et al., 2012; Lieberman-Aiden et al., 2009). To correctly assign and confirm the eigenvalues, a Pearson correlation was computed between the eigenvector coefficient and the mean scoring value of each omic layer at 100Kb intervals (**Figure 24**). I observed that the active chromatin marks i.e. H3K4me1, an enhancer mark, and chromatin accessibility were positively correlating with positive eigenvalues, which define the A-type compartment, and consequently demonstrated the activity of this compartment.
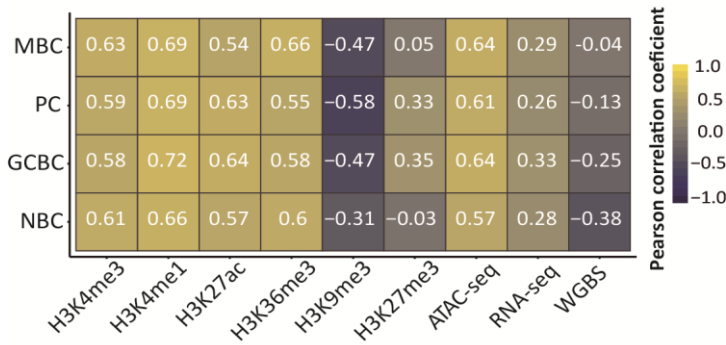
Figure 24. Correlation of compartments with different omic layers. Pearson correlation between the eigenvector coefficients, which defines 3D compartments per B-cell subpopulation, with six histone marks, chromatin accessibility (ATAC-seq), gene expression (RNA-seq) and DNA methylation (WGBS). Positive values of the eigenvector show higher correlation with H3K4me1 (enhancer mark) and chromatin accessibility.

Those two layers of information, which best correlate with A-type compartment, were used to discriminate between A-type (active) and B-type (inactive) compartments, as exemplified in chromosome 12 (**Figure 25A**). Once the eigenvector was correctly assigned, a pairwise correlation of the first eigenvector of each B-cell subpopulation showed that NBC and MBC, on the one hand, and GCBC and PC, on the other hand, have similar compartmentalization (correlation coefficient, R^2=0.95, **Figure 25B**), confirming the previous results using the RS (**Figure 22B**).
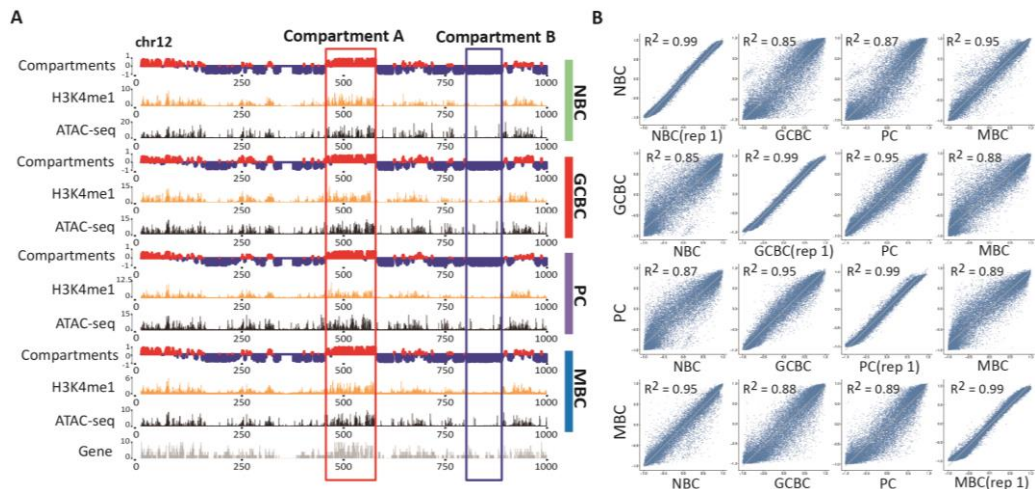
Figure 25. Genome compartmentalization. A. Example on chromosome 12 (chr12) comparing the profile of 3D data (*in situ* Hi-C), H3K4me1 ChIP-seq signal, chromatin accessibility (ATAC-seq)

and gene density. The red and blue rectangles highlight the features of A-type and B-type compartments, respectively. **B.** Genome-wide scatterplots of coefficients from the first eigenvector showing the correlation between pairs of B-cell subpopulations at 100Kb resolution. The squared correlation coefficient (R^2) is indicated.

Unexpectedly, the H3K27me3 histone mark, which is deposited by the polycomb repressive complex (Margueron and Reinberg, 2011), was neither correlated with positive nor with negative eigenvector coefficients (**Figure 24**). This observation may suggest that the histone mark H3K27me3 does not belong to the standard A-type and B-type compartments, and may be associated with a different compartment type. To shed light onto this issue, visual inspection of the first eigenvector distribution revealed a positive extreme, a negative extreme and a long intermediate valley (**Figure 26A**). Indeed, by applying the Bayesian Information Criterion, it was shown that the classification into three compartments was the best compromise between distribution fitting accuracy and minimum number of compartments (**Figure 26B**).
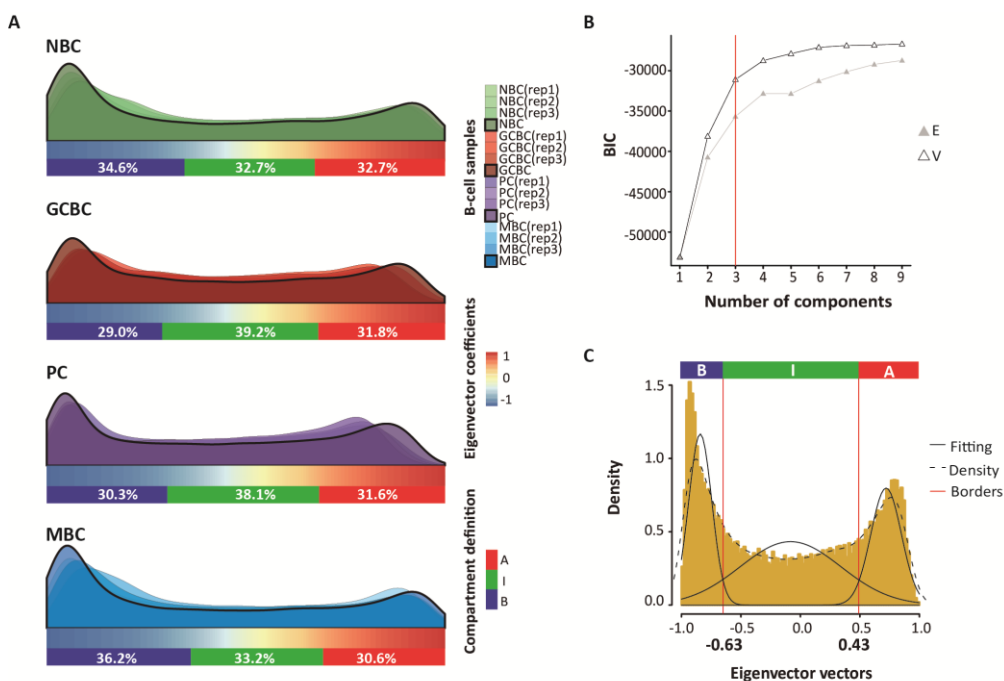
Figure 26. Identification of an intermediate compartment. A. Distribution of the first eigenvector of each B-cell subpopulation (each of the three replicates and merge). The relative abundance of A-type, B-type and intermediate (I)-type compartments per merged B-cell

subpopulations are indicated below each distribution. Compartment definition based on eigenvalue thresholds: A-type, 1 to 0.43; I-type, 0.43 to -0.63; B-type, -0.63 to -1. **B.** Bayesian Information Criterion (BIC) plot for the equal (E) and unequal (V) variance model parameterization ranged from 1 to 10 clusters. E and V are colored as grey and black, respectively. **C.** Compartment definition model. The x-axis shows the distribution of the eigenvector coefficients and the y-axis indicates the density. The fitting model proposed is highlighted using solid black line. The red lines mark the intersection points (EV1 = -0.63 and EV2 = 0.43) used to distinguish the three different compartments (A-type, I-type, B-type).

Subsequently, the eigenvector distribution was modeled to establish the thresholds segmenting the data into A-type, B-type and intermediate (I-type) compartment (**Figure 26C**). Each intersection point was defined as the mean value from all twelve B-cell subpopulation replicates, grouping the data into those three different components, composed with the following eigenvector values: A-type, between +1 and +0.63, I-type, between +0.63 and -0.43 and B-type, between -0.43 and +1.00. Each B-cell subpopulation had similar compartment composition.

Analyzing these three compartments together with other omics layers revealed the expected association of A-type compartment with active chromatin (H3K4me1, H3K4me3, H3K27ac and H3K36me3), B-type compartment with H3K9me3 (repressive mark), and a remarkable association between the I-type compartment and the presence of H3K27me3 (**Figure 27A**). Indeed, using chromHMM-based chromatin state model specific for B cells (**Appendix Figure 2**) (Beekman et al., 2018a; Ernst and Kellis, 2017), it could be observed that the regions associated with the I-type compartment were enriched for poised-promoter and polycomb-repressed chromatin states (**Figure 27B** and **27C**).
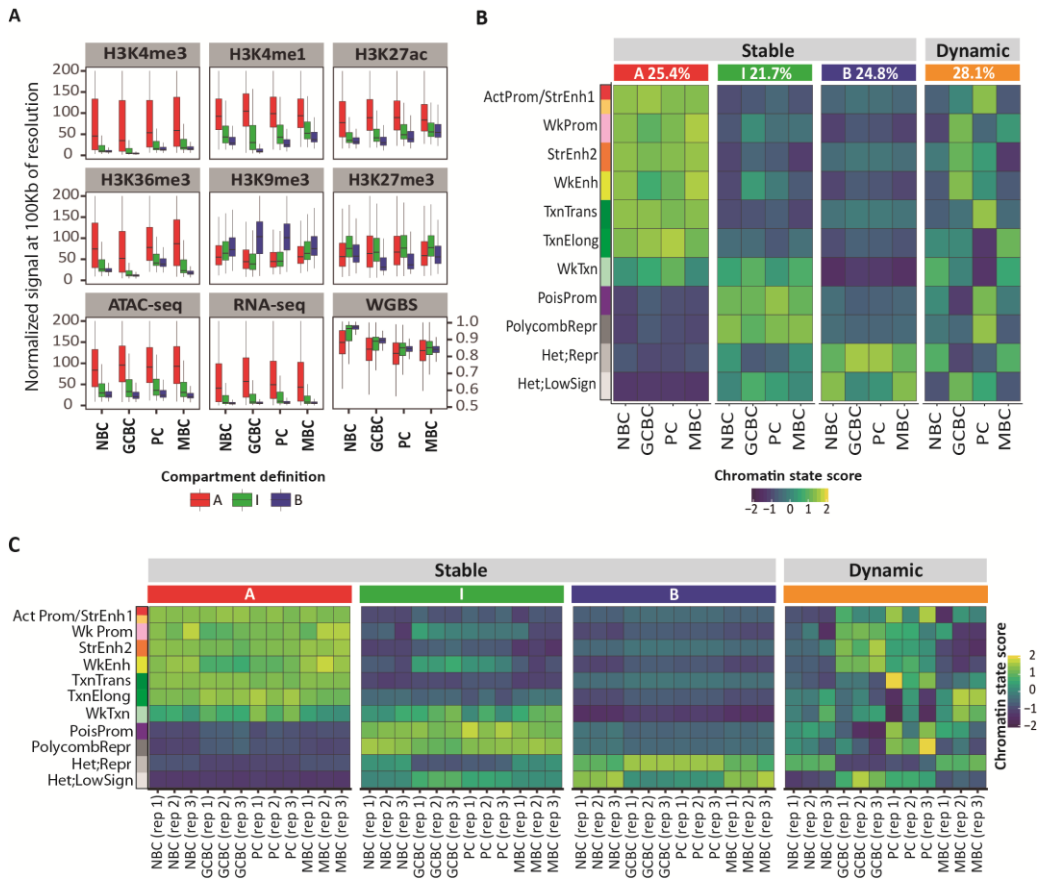
Figure 27. Functional validation of the spatial compartments defined. A. Boxplots showing the association between the three compartments (A-type, I-type and B-type) and each of the nine additional omics layers under study. **B.** Functional association of the conserved and dynamic compartments during B-cell maturation using eleven chromatin states (normalized by sample and chromatin state). Conserved compartments were segmented into A-type, I-type and B-type compartments. The percentage of each conserved or dynamic compartment is indicated for all B-cell subpopulations. ActProm-StrEnh1, Active Promoter-Strong Enhancer 1; WkProm, Weak Promoter; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; TxnTrans, Transcription Transition; TxnElong, Transcription Elongation; WkTxn, Weak Transcription; PoisProm, Poised Promoter; PolycombRepr, Polycomb-Repressed; Het;Repr, Heterochromatin-Repressed; Het;LowSign, Heterochromatin-Low Signal. **C.** Functional validation of the conserved (A-type, I-type and B-type) and dynamic compartments in all B-cell subpopulations replicates using eleven different chromatin states. The chromatin state score is normalized by sample and chromatin state.

2.5 Polycomb-associated chromatin defines an intermediate and moldable 3D genome compartment

Next, I wondered whether the compartments interactions occurring in the I-type compartment could also define its potential transitional role between A and B compartments. To do so, the compartment interactions were quantified by computing the compartment score (C-score) as the ratio of intra-compartment interactions over the total chromosomal interactions per compartment (**Figure 28A**). Interestingly, the I-type compartment was associated with lower C-score than the A-type and B-type compartments (**Figure 28B**). We further explored this phenomenon by dividing the I-type compartment into two blocks differentiating positive (IA) and negative (IB) eigenvector components (**Figure 28C**). The analysis showed that the I-type compartment, regardless being IA or IB, was consistently having lower C-score than the A or B-type compartments. This finding further supports the existence of the I-type compartment as an independent chromatin structure different from A and B-type compartments. Additionally, it suggests that the I-type compartment tends to interact not only with itself but also with A and B-type compartments, and as such it may represent an interconnected space between the fully active and inactive compartments.
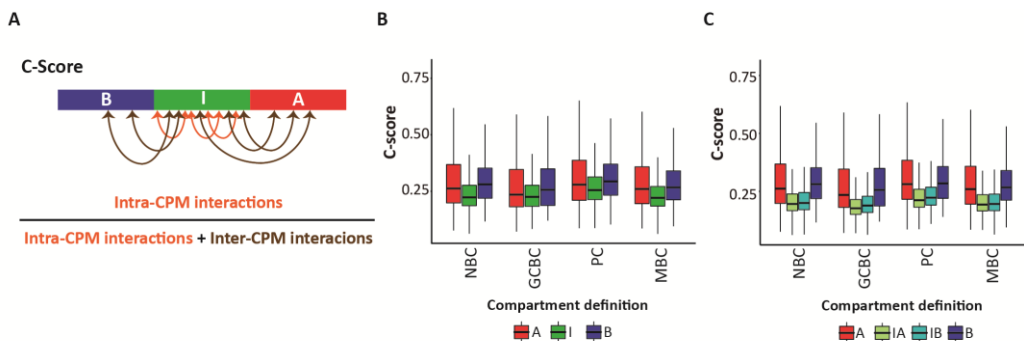
Figure 28. C-score definition and distribution acroos the compartments. A. Method defined by the ratio of contacts between regions within the same compartment (intra-compartment contacts) over the total chromosomal contacts per compartments (intra- and inter-chromosomal interactions). **B.** C-score distributions on the three defined compartments A-type, I-type and B-type. **C.** C-score distributions segmenting the I-type compartment onto positive (IA) or negative (IB) eigenvector coefficients.

To study the potential role of the I-type compartment during B-cell differentiation, I selected the poised promoters or polycomb-repressed regions within this type of compartment in NBC and studied how they change in both compartments and chromatin state upon differentiation into GCBC (**Figure 29**).
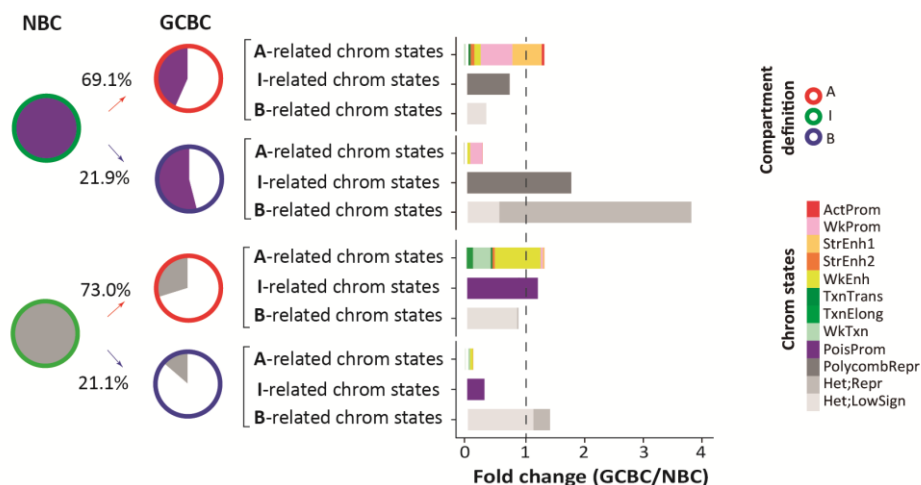
Figure 29. Intermediate compartment dynamics. Pie charts represent poised promoters (*top*, violet color) or polycomb-repressed (*bottom*, light gray color) within the I-type compartment in NBC, which shifts to A-type and B-type compartments in GCBC. Compartment are indicated by color of the circle line. The pie charts under GCBC represent the fraction that maintains the previous chromatin state (colored as previously defined) or changed chromatin states (not colored). Bar graphs represent the fold change between GCBC and NBC of each three groups of chromatin states (arranged by their relationship to the A-type, I-type and B-type compartments). Active Promoter, Weak Promoter, Strong Enhancer 1, Strong Enhancer 2, Weak Enhancer, Transcription Transition, Transcription Elongation, Weak Transcription were A-type compartment-related chromatin states. Heterochromatin-Repressed and Heterochromatin-Low signal were B-type compartment-related chromatin states. Poised Promoter or Polycomb-Repressed chromatin states were I-type compartment-related states.

The majority of I-type compartment in NBC (i.e. 69.1% of poised promoter and 73.0% of polycomb-repressed) change into A-type compartment, a consistent fraction (21.9% and 21.1%) into B-type, and only a small fraction (9% and 5.9%) maintain their intermediate definition. This finding indicates that the regions with a most prominent I-type compartment character undergo a widespread structural modulation during NBC to GCBC differentiation step. Interestingly, transitions from I-type to A-type

compartment (activation events) were paired with a reduction of poised promoters (56.7% loss) and polycomb-repressed states (70.2% loss). These reductions were associated with an increase of A-related chromatin states (1.31- or 1.33-fold change coming from poised promoter or polycomb-repressed, respectively) such as promoter, enhancer and transcription (**Figure 29**). Conversely, poised promoters and polycomb-repressed regions associated with I-type compartments in NBC that changed into B-type compartments in GCBC (inactivation events) were linked to a reduction of poised promoter (46.02% loss) and polycomb-repressed states (86.3% loss). These reductions were related to an increase of B-related chromatin states (3.81 or 1.4-fold change coming from poised promoter or polycomb-repressed, respectively) such as heterochromatin characterized by H3K9me3 (**Figure 29**).

Altogether, Hi-C analyses (using a principal component approach) uncover an intermediate transitional compartment with biological significance, enriched in poised and polycomb-repressed chromatin states, interconnected with A-type and B-type compartments, and amenable to rewire the pattern of interactions leading to active or inactive chromatin state transitions upon cell differentiation.

2.6 Changes in genome compartmentalization are reversible during B-cell differentiation

Next, I studied the 3D genome dynamics during B-cell differentiation. I initially observed that 71.9% of the compartments were stable across B-cell differentiation, accounting for 25.4% (n=6,409; the *n* refers to number of compartments at 100Kb resolution) of A-type, 21.7% (n=5,467) of I-type and 24.8% (n=6,267) of B-type compartment. Interestingly, the remaining 28.1% (n=7,099) of the genome dynamically changes compartment during B-cell differentiation (**Figure 30A** and **30B**). B-cell differentiation is not a linear process, NBCs differentiate into GCBCs, which then branch into long-lived MBCs or antibody-producing PCs (**Figure 20A**). Thus, I studied the 3D genome compartment dynamics along these two main differentiation paths (i.e. NBC-GCBC-PC and NBC-GCBC-MBC). Then, at each differentiation step, I classified the genome into three different dynamics: (i) compartments undergoing activation events (B-type to A-type, B-type to I-type, or I-type to A-type), (ii) compartments undergoing inactivation events (A-type to B-type, A-type to I-type, or I-type to B-type), and (iii) stable compartments (**Figure 30A** and **30B**). The NBC-GCBC-MBC differentiation path

suggests that the extensive remodeling taking place from NBC to GCBC is followed by an overall reversion of the compartmentalization in MBC, achieving a profile similar to NBC (**Figure 30A**). To assess the capacity of the genome to revert to a past 3D configuration, I analyzed the compartments in NBC as compared to those in PC and MBC. Indeed, I globally observed that 72.7% of the regions in MBC re-acquire the same compartment type as in NBC. This phenomenon was mostly related to compartments undergoing activation in GCBC, as 82.9% of them reverted to inactivation upon differentiation into MBC. This finding is in line with solid evidence showing that NBC and MBC, in spite of representing markedly different maturation B-cell stages, are transcriptionally similar (Agirre et al., 2019; Klein et al., 2003), (**Figure 23**).
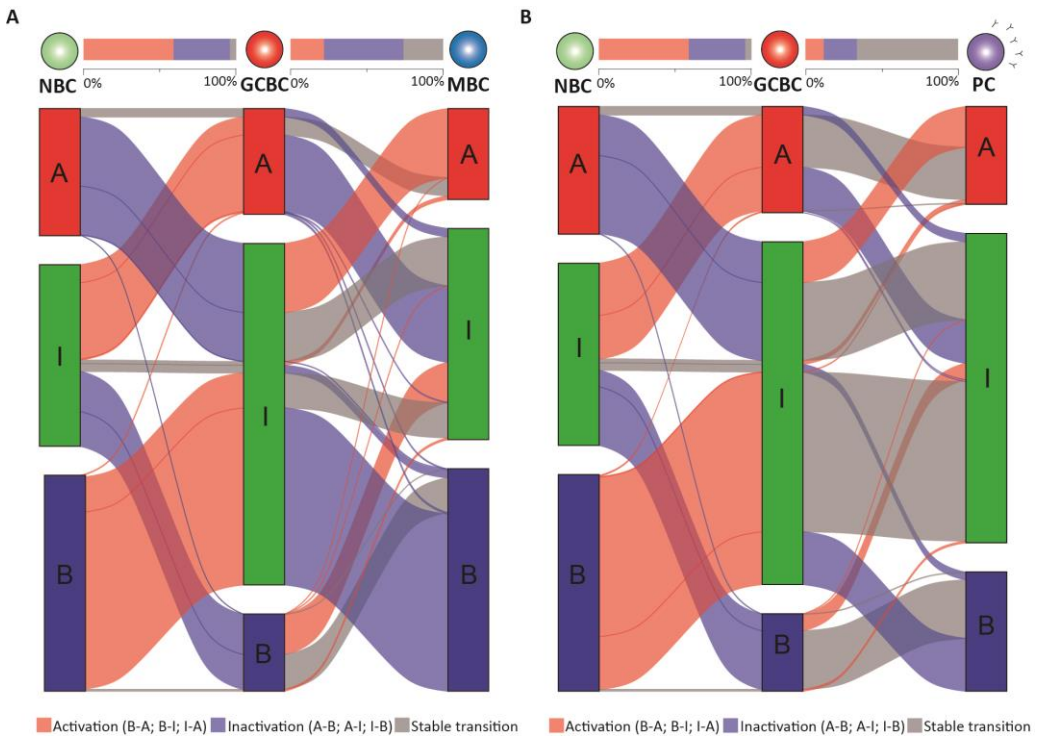
Figure 30. Structural genome modulation in the B-cell lineage. A/B. Alluvial diagrams showing the compartment dynamics in the two branches of mature B-cell differentiation: NBC-GCBC-MBC (**A**) and NBC-GCBC-PC (**B**). Activation, in red, represents changes from compartment B-type to A-type, B-type to I-type and I-type to A-type. Inactivation, in blue, represents changes from A-type to B-type, A-type to I-type and I-type to B-type compartments. The non-changed compartments are represented in gray. On the top, the bar plots between B-cell subpopulations

represent the total percentage of regions changing to active or inactive, and regions that conserve its previous compartment definition.

In the case of PC, the reversibility to a 3D pattern similar to NBC involved only 30.8% of the genome (**Figure 30B**). To determine whether this compartment reversibility was also accompanied by a functional changes, I analyzed the chromatin state dynamics within the compartments becoming uniquely active in GCBC as compared to NBC, MBC and PC (n=937). I observed that the transient compartment activation from NBC to GCBC is related to an increase of A-related chromatin states (1.36-fold change) involving active promoter, strong enhancer and transcription. Conversely, the subsequent 3D genome inactivation upon differentiation into PC and MBC was related to an increase in B-related chromatin states (1.21- and 1.15-fold change, respectively) involving mainly heterochromatin (**Figure 31A**).
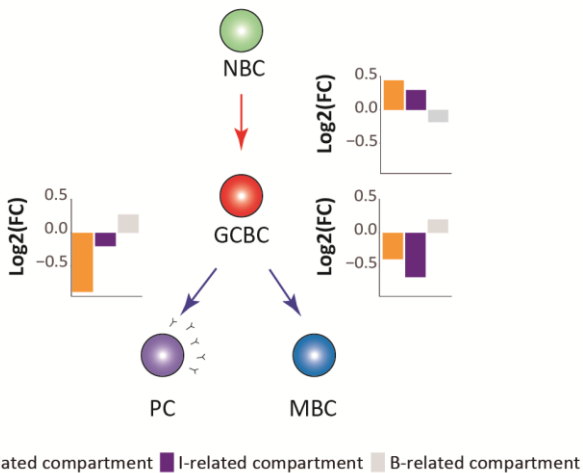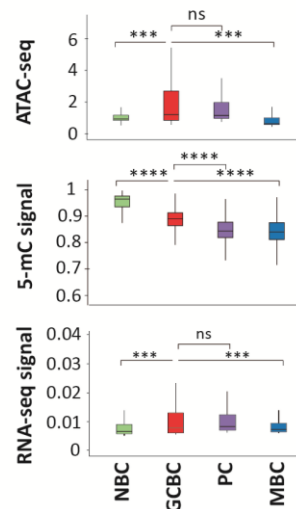
Figure 31. **Multi-omics characterization of the 937 regions (of 100Kb resolution) gaining activity exclusively in GCBC. A.** Scheme of B-cell differentiation and chromatin state dynamics, in which the barplots indicate the log2 fold change of each three groups of chromatin states (related to the A-type, I-type and B-type compartments). **B.** Boxplots of chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal) per B-cell subpopulations compared using the Wilcoxon's test. ***p-value<0.0001, ****p-value<0.00001, ns – not significant. A-type compartment-related chromatin states were: Active Promoter, Weak Promoter, Strong Enhancer 1, Strong Enhancer 2, Weak Enhancer, Transcription Transition, Transcription Elongation, Weak Transcription. B-type compartment-

related chromatin states were: Heterochromatin-Repressed and Heterochromatin-Low signal. I-type compartment-related states were: Poised Promoter or Polycomb-Repressed.

Those specific active regions in GCBC had a significant increase in chromatin accessibility and gene expression in GCBC as compared to NBC and MBC, but not in PC (**Figure 31B**). These findings suggest that structural 3D reversibility in MBC is accompanied by a functional reversibility whereas PC partially maintains gene expression levels and chromatin accessibility similar to GCBC in spite of the compartment changes. Interestingly, in contrast to chromatin-based marks, DNA methylation was overall unrelated to compartment or chromatin state dynamics of the B-cell differentiation (**Figure 31B**). This result is in line with the data already published by our group showing that DNA methylation is gradually lost during the maturation process and is marked in late-stages of B-cell differentiation (Kulis et al., 2015).

2.7 The 3D genome of GCBC undergoes extensive compartment activation

The previous analyses revealed that the NBC and GCBC transition was associated to a large structural reconfiguration of compartments involving 96.0% of all dynamic compartments (**Figure 27B**). Interestingly, taking into account the compartment changes between NBC and GCBC, I observed that 61.5% involved compartment activation (**Figure 30A** and **30B**). As the germinal center reaction is known to be mediated by specific TFs (De Silva and Klein, 2015b; Song and Matthias, 2018) and those may be involved in shaping the spatial organization of the genome (Bunting et al., 2016; Johanson et al., 2018; Stadhouders et al., 2018), I further explored the presence of TF binding motifs in accessible chromatin in the regions gaining H3K27ac within the newly activated compartments (n=937). I identified enriched motifs for several TFs, being members of the MEF2 and POU families the most significant (**Figure 32** and **Appendix Table 5A**). Remarkably, these TFs are essential for the germinal center formation (Brescia et al., 2018; Schubart et al., 2001; Wilker et al., 2008; Ying et al., 2013).
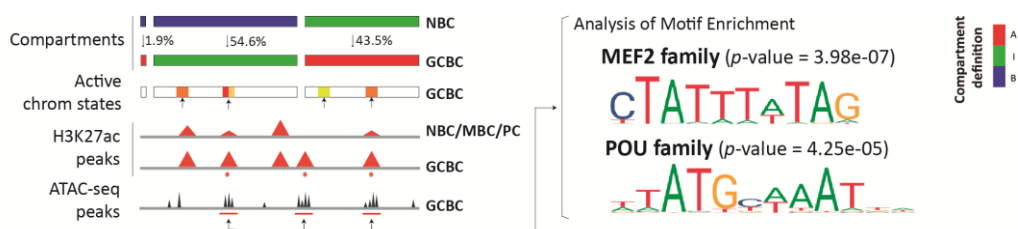
Figure 32. Enrichment analysis of transcription factor binding motifs. *Left*: Schematic representation of the analytic strategy. *Right:* Binding motifs of MEF2 and POU TF families are highly enriched in active and accessible loci in the GCBC specific regions gaining activity (n=171 independent genomic loci) versus the background (n=268 independent genomic loci). *P*-values were calculated using the AME-MEME suite. Out of the list of all enriched transcription factor binding motifs, I considered only those expressed in the three GCBC replicates.

Furthermore, the newly activated compartments hosted about 100 genes significantly upregulated in GCBC (out of the 335 genes expressed in GCBC) as compared to NBC, MBC and PC (FDR<0.05) (**Appendix Table 6**). Remarkably, among them was the Activation-Induced Cytidine Deaminase (*AICDA*) gene, which is essential for class-switch recombination and somatic hypermutation in GCBC and is specifically expressed in GCBC (de Yébenes and Ramiro, 2006). Indeed, the *AICDA* locus was globally remodeled from an inactive state in NBC to a global chromatin activation in GCBC, which included an increase in the ratio of GCBC/NBC 3D interactions, as well as increased levels of active chromatin states (that is, active promoter and enhancers as well as transcriptional elongation), open chromatin, and gene expression (**Figure 33A** and **33B**). This analysis also revealed the presence of possible upstream and downstream *AICDA*-specific enhancers that gain interactions with the gene promoter in GCBC (**Figure 33B**). Interestingly, this multilayer chromatin activation at the *AICDA* locus was reverted to the inactive ground state once GCBC differentiate into MBC or PC.
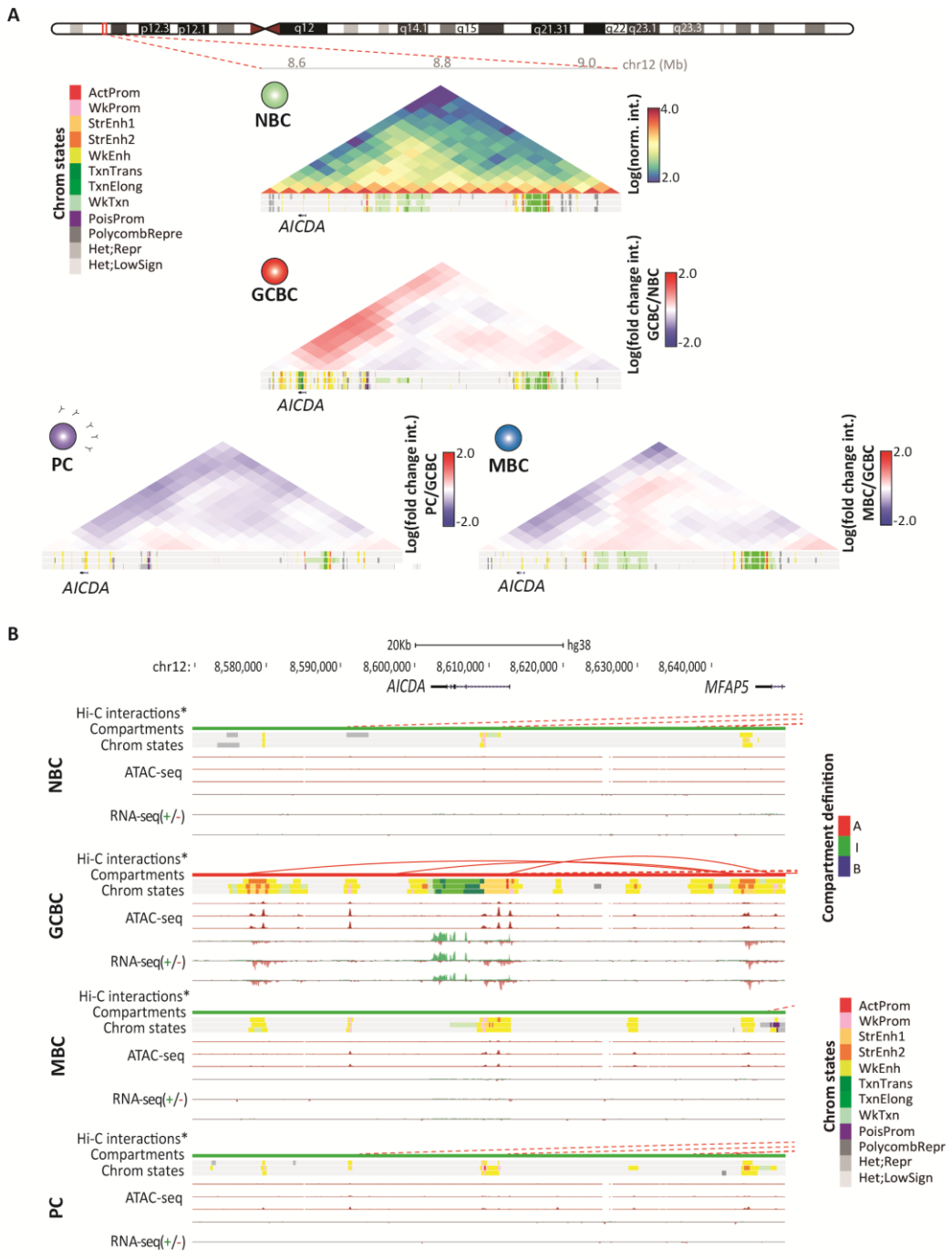
Figure 33. Chromatin organization at the *AICDA* locus. A. Normalized Hi-C contact map of the domain structure surrounding the *AICDA* gene in NBC. The log fold change interaction ratio between GCBC, MBC or PC as compared to NBC was computed. Below each interaction map,

chromatin state tracks of three biological replicates per B-cell subpopulation are shown. The coordinates of the represented region are chr12:8,550,000-9,050,000; GRCh38. **B.** Multi-layer epigenomic characterization of *AICDA* gene region (chr12:8,598,290-8,615,591; GRCh38) in four B-cell subpopulations. Arc diagrams indicate the Hi-C significant interactions (continuous red lines involve the region of interest, while dashed red lines involve other regions of chromosome 12). Below them, the compartment definition (red, compartment A-type: green, compartment I-type), chromatin states, chromatin accessibility (ATAC-seq, y-axis signal from 0 to 105) and gene expression (RNA-seq, y-axis signal from 0 to 4 for the positive strand and from 0 to -0.1 for the negative strand) is shown. Tracks of Hi-C interactions and compartment definition are based on merged replicates whereas chromatin states, chromatin accessibility and gene expression tracks of each replicate is shown separately. The coordinates of the represented region are chr12:8,570,000-8,670,000; GRCh38.

2.8 3D genome organization upon B-cell neoplasms transformation

Once defined the 3D genome changes during mature B-cell differentiation, I next focused the analysis on the 3D genome organization in neoplastic B cells. To address this, I performed *in situ* Hi-C in fully characterized tumor cells from patients with chronic lymphocytic leukemia (uCLL n=2; mCLL n=5) or mantle cell lymphoma (cMCL n=2; nnMCL n=3). Firstly, an unsupervised clustering of the RS from the entire Hi-C datasets, taking into account normal and neoplastic B cells, indicated that CLL and MCL clustered separately from each other and within a major cluster that include NBC and MBC (**Figure 34A** and **34B**). Interestingly, NBC and MBC have been described as potential cells of origin of these neoplasms (Puente et al., 2018).
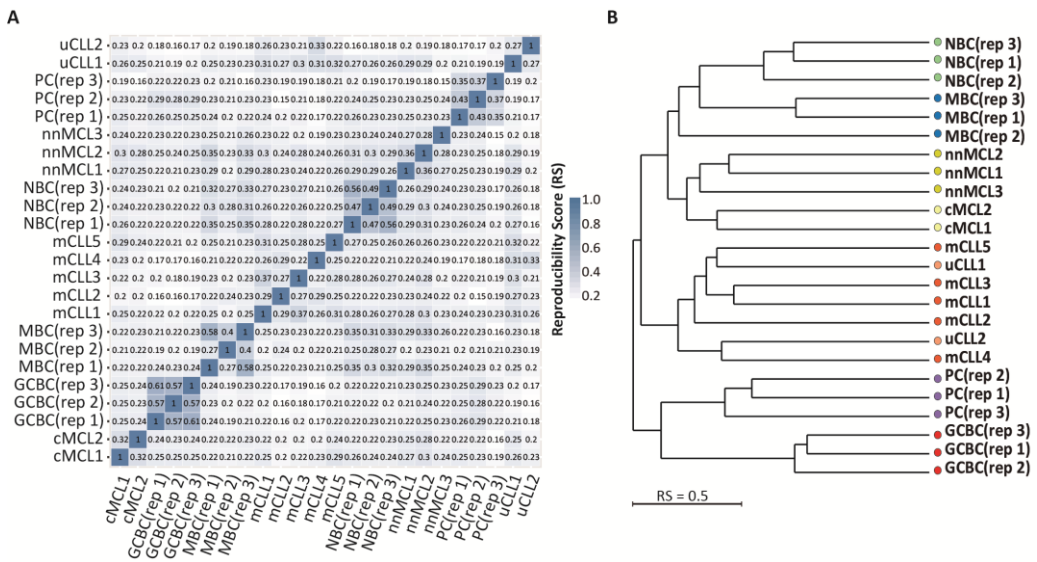
Figure 34. Similarity among B-cell neoplasia patient samples. Average genome-wide matrix (**A**) and dendrogram (**B**) of the reproducibility score for normalized Hi-C contact maps at 100Kb resolution for B-cell subpopulations replicates and samples from B-cell neoplasia patients. IGHV unmutated (u)CLL; IGHV mutated (m)CLL; conventional (c)MCL and leukemic non-nodal (nn)MCL.

Furthermore, pairwise eigenvector correlation analysis of the cancer samples suggested that the 3D genome configuration of the two clinico-biological subtypes of CLL was rather homogeneous (being the correlation among samples of the first eigenvector $R^2 > 0.91$) (**Figure 35A** and **35B**). However, this was not the case for the two MCL subtypes, as they presented a more differential chromatin conformation (being the correlation between the samples from the same subtypes of $R^2 = 0.92$-0.93 but left among subtypes, $R^2 < 0.88$) (**Figure 35C** and **35D**).
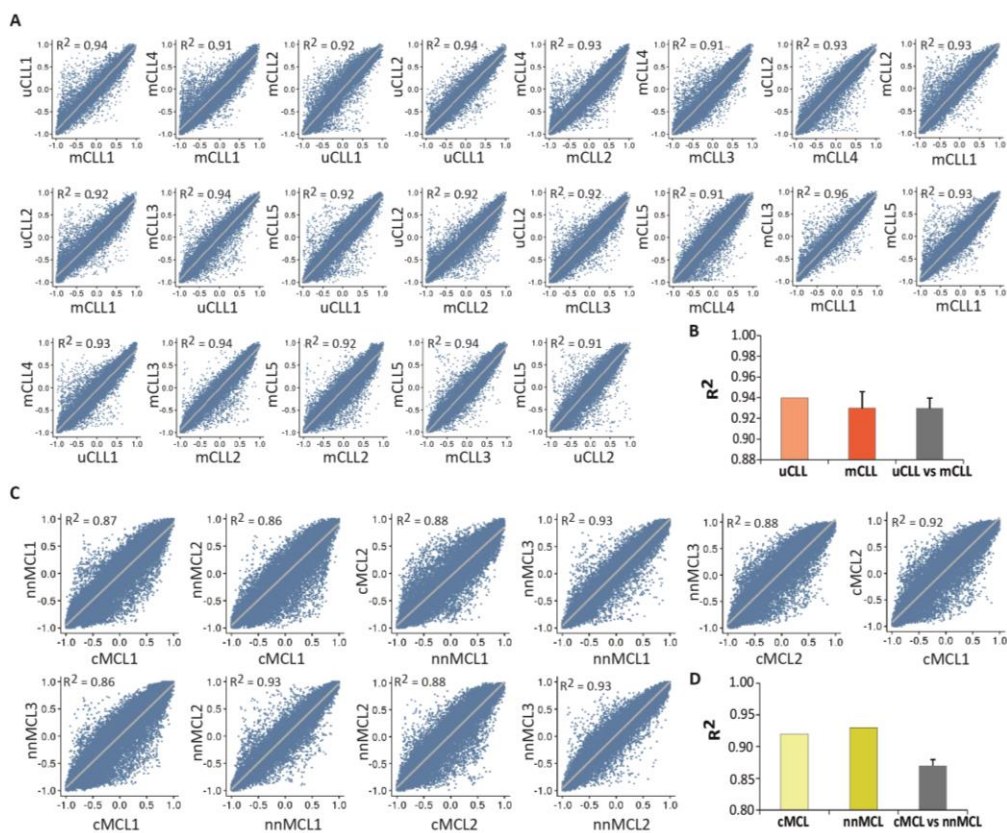
Figure 35. Eigenvector correlations between each B-cell neoplasia samples. A and C. Genome-wide scatterplots of the first eigenvector showing the correlation between pairs of each B-cell

malignancy samples at 100Kb resolution. CLL (**A**). MCL (**C**). The squared correlation coefficient (R^2) is indicated. **B and D.** Mean and standard deviation of the squared correlation coefficients calculated intra- or inter- each neoplasia subtype. CLL (**B**). MCL (**D**).

Interestingly, the described clustering patterns of 3D genome data, where both B-cell neoplasias tend to group together with NBC and MBC, was also observed in the first component of the six histone modifications, chromatin accessibility and gene expression (**Figure 36**).
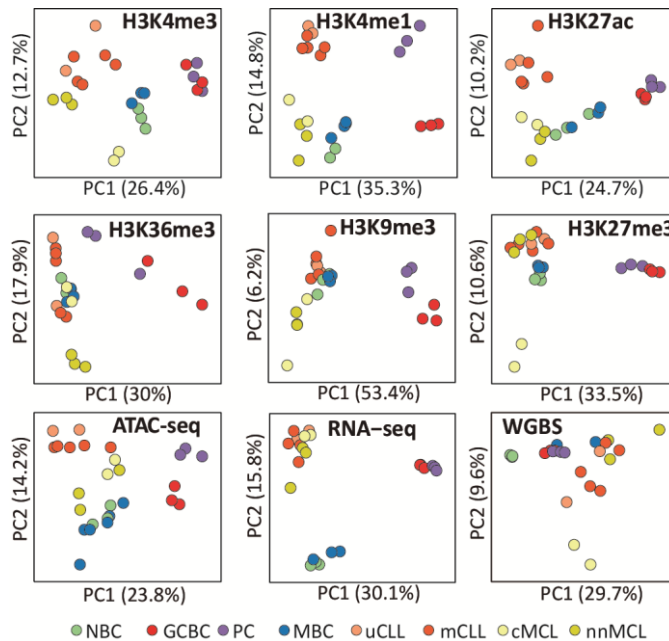
Figure 36. Multi-omics view upon B-cell differentiation and neoplastic transformation. Unsupervised PCAfor nine omic layers generated in the same patient samples as Hi-C: ChIP-seq of six histone marks (H3K4me3 n=53,241 genomic regions, H3K4me1 n=54,653 genomic regions, H3K27ac n=106,457 genomic regions, H3K36me3 n=50,530 genomic regions, H3K9me3 n=137,933 genomic regions, and H3K27me3 n=117,560 genomic regions), chromatin accessibility measured by ATAC-seq (n=140,187 genomic regions), DNA methylation measured by WGBS (n=14,088,025 CpGs) and gene expression measured by RNA-seq (n=57,376 transcripts).

2.9 3D genome changes upon neoplastic transformation

The differential 3D genome clustering of CLL and MCL samples (**Figure 34B**) suggested the presence of disease-specific changes of their 3D genome organization.

In order to study the 3D changes that take place in each B-cell malignancy I selected the conserved compartments during normal B-cell differentiation and analyzed each of the lymphoid neoplasias separately. Overall, at the qualitative level, I observed that 23.8% of the genome showed dynamic changes in CLL (**Figure 37A** *left*) and 27.3% in MCL (**Figure 37B** *left*). From those dynamic compartments, using a more stringent quantitative approach (absolute difference in the eigenvalue>0.4 and FDR<0.05), I aimed to detect significant changes associated with CLL or MCL. I ended up with a total of 348 and 82 significant compartment segments of 100Kb in CLL and MCL, respectively. The larger number of regions changing compartments in CLL correlates with the results of the Hi-C based clustering (**Figure 34B**), which indicates that MCL is in overall more similar to NBC and MBC than CLL.

In CLL, I observed that 57.5% of the changes were associated with compartment inactivation while 42.5% were activating changes (**Figure 37A** *middle* panel and **Appendix Figure 3A**). In MCL, this trend was reversed and observed that 43% of the compartment changes were inactivating and 57% were activating (**Figure 37B** *middle* panel and **Appendix Figure 3B**).

Next, I wanted to functionally characterize the most relevant dynamic changes. I looked at the regions that differentially lose the A-type compartment in normal B cells while transforming into CLL. I observed that inactivation at the 3D genome level in CLL was linked to a shift to poised promoter and polycomb-repressed chromatin states, and a significant loss of chromatin accessibility and gene expression (**Figure 37A** *right*). As in MCL a greater number of activating 3D changes were occurring, I selected the B-type compartments in normal B cells and studied activation upon MCL transformation. MCL activation at the 3D genome level was accompanied with an enrichment of active chromatin states and a significantly increase in chromatin accessibility and gene expression (**Figure 37B** *right*). Thus, the 3D genome organization changes were associated with the expected modulation of chromatin function. Overall, these results point to the presence of recurrent and specific changes in the 3D genome organization in CLL and MCL, being the former more extensively altered than the latter.
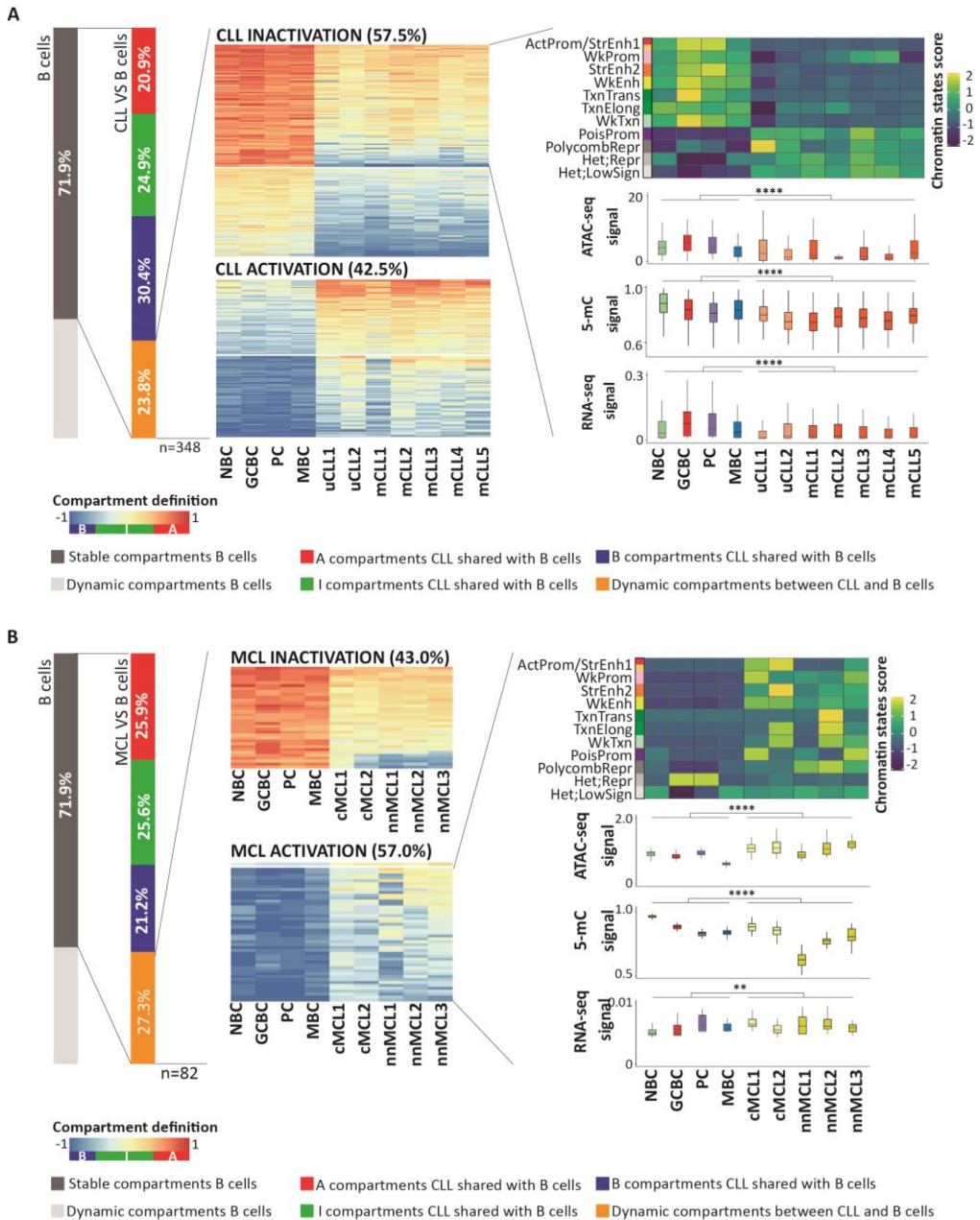
Figure 37. Characterization of the chromatin architecture of human B-cell neoplasms in the context of B cell differentiation. A. Compartment changes upon CLL transformation. *Left*: First bar graph represents the percentage of conserved and dynamic compartments during normal B-cell differentiation. Second bar graph shows the percentage of compartments stable and differential in CLL as compared to normal B cells. A total of 23.8% of the compartments change

in at least one CLL sample. *Middle*: Heatmaps showing eigenvector coefficients of the 348 compartments significantly losing (n=200) or gaining activation (n=148) between all CLL samples and normal B cells. *Right*: Multi-omics characterization of the 200 regions losing activity in CLL. Chromatin states, chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal) in CLL and normal B cells are shown. Comparisons were performed using the Wilcoxon's test. ****p-value<0.00001. **B.** Compartment changes upon MCL transformation. *Left*: First bar graph represents the percentage of conserved and dynamic compartments in B cells. Second bar graph shows the percentage of conserved compartments between B cells and MCL, being 27.3% non-conserved compartment in MCL. *Middle*: Heatmaps showing eigenvector coefficients of significant dynamic compartments (n=82) between MCL and B cells. Regions were split in two groups (MCL activation, n=35 or inactivation, n=47) according to the structural modulation of the MCL compared to B cells. *Right*: Example of the MCL activation subset (mostly those B-type compartments in B cells which significantly increase eigenvector coefficients in MCL) showing the chromatin states pattern, chromatin accessibility (ATAC-seq signal), DNA methylation (5-mC signal) and gene expression (RNA-seq signal). Comparisons were performed using the Wilcoxon's test. **p-value<0.001, ****p-value<0.00001.

2.10 Defining B-cell neoplasia specific 3D genome organization

To further characterize the 3D genome compartmentalization of neoplastic B cells, I classified the changing compartments as common (present both in CLL and MCL) or entity-specific (present either in CLL or MCL). I detected 31 compartments commonly altered in both malignancies, revealing the existence of a core of regions that distinguish normal and the neoplastic B cells under study (**Figure 38A**). Specifically, 17 inactivated and 14 activated compartments in both B-cell neoplasias as compared to normal B cells were found (**Figure 38B**). I then analyzed the genes differentially expressed at those compartments between the two neoplasias and normal B cells. Suprisingly, no differentially expressed genes in the MCL and CLL shared activated compartments were detected, while 12 genes were observed differentially downregulated in the two neoplasias shared inactived compartments as compared to normal B cells. Among them, a gene of special interest was *PDE4D*, a cyclic nucleotide phosphodiesterase isoform reported to be downregulated in CLL (Zhang et al., 2008a), (**Appendix Table 7**).
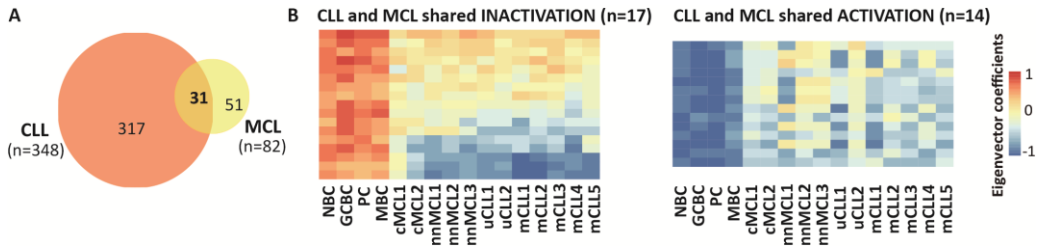
Figure 38. Common B-cell neoplasia compartments. A. Venn diagram showing the significant number of dynamic compartments in CLL and MCL as compared to normal B-cell differentiation and the regions shared between both B-cell neoplasms (n=31). **B.** Heatmaps showing eigenvector coefficients of compartments significantly losing or gaining activation between B-cell neoplasms (MCL and CLL together) and B cells.

A targeted analysis of MCL and CLL revealed the presence of 3 MCL-specific compartment changes (**Figure 39**), while 89 comparments were CLL-specific (41 and 48 inactivated and activated, respectively) (**Figure 40A**).

Figure 39. MCL-specific compartments. Heatmap showing the eigenvector coefficients of the compartments losing (*top*) or gaining (*bottom*) activation specifically in MCL. On the compartments losing activation, there are two genes differentially expressed in CLL and B cells as compared to MCL, *TXNDC5* and *BLOC1S5-TXNDC5*.

In CLL cells, I could determine 41 compartments specifically losing activation (originating from I-type, n=17 or B-type, n=24 in normal B cells). Those compartments were significantly enriched (*p*-value=0.0060) in downregulated genes (n=11, FDR<0.05) as compared to normal B cells and MCL samples, being the *Early B-cell Factor 1* (*EBF1*) a remarkable example (**Figure 40A** and **40B**).
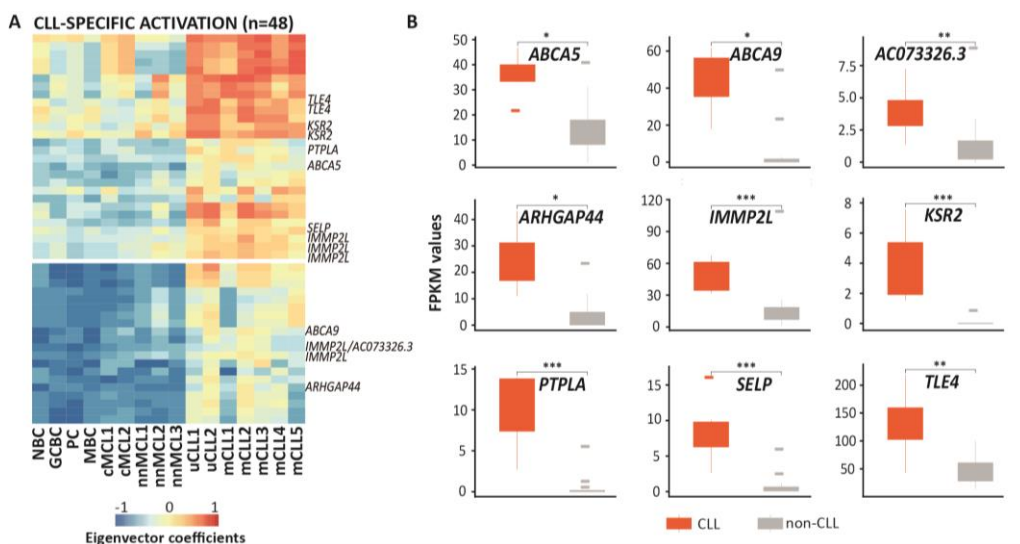
Figure 40. CLL-specific inactivated compartments. A. Heatmap showing the eigenvector coefficients of the compartments losing activation specifically in CLL (n=41). Significantly downregulated genes (FDR<0.05) associated to each compartment are shown on the right of the heatmap (*p*-value=0.0038, calculated from the total number of genes picked on 48 random compartments per 10,000 times). **B**. FPKM values of all the CLL-specific significantly downregulated genes within compartments losing activation. *adjusted *p*-value<0.05, **adjusted *p*-value<0.005, ***adjusted *p*-value<0.0005.

EBF1 downregulation has been described to be a diagnostic marker in CLL (Navarro et al., 2017), and its low expression may lead to reduced levels of numerous B-cell signaling factors contributing to the anergic signature of CLL cells (Mockridge et al., 2013; Muzio et al., 2008) and low susceptibility to host immunorecognition (Schultze et al., 1996; Seifert et al., 2012). To obtain insights into the mechanisms underlying *EBF1* silencing in CLL, I analyzed in detail a 2Mb region hosting the gene, which also contains two nearby protein coding genes, *RNF145* and *UBLCP1*, and the lncRNA *LINC02202*. I observed that a large fraction of 3D interactions involving the *EBF1* region in normal B cells were lost in CLL resulting in a change from A-type to I-type compartment and a sharp inactivation of the gene, as shown by the analysis of chromatin states (**Figure 41A**). Remarkably, in spite of the global reduction of 3D interactions, the two adjacent genes (*RNF145* and *UBLCP1*) were located in the only region (spanning 200Kb) that remained as A-type compartment in the entire 2Mb region, maintaining thus an active state.
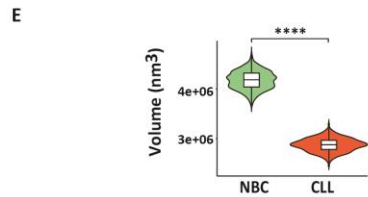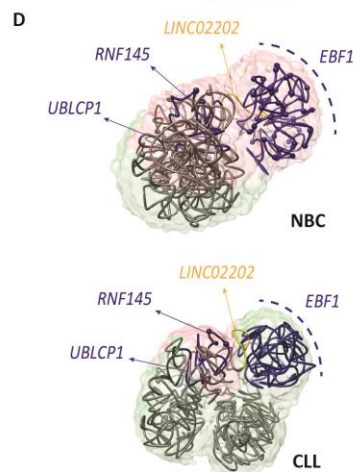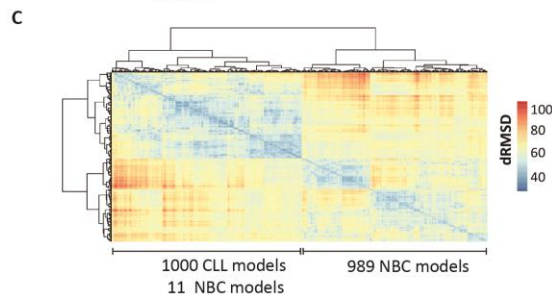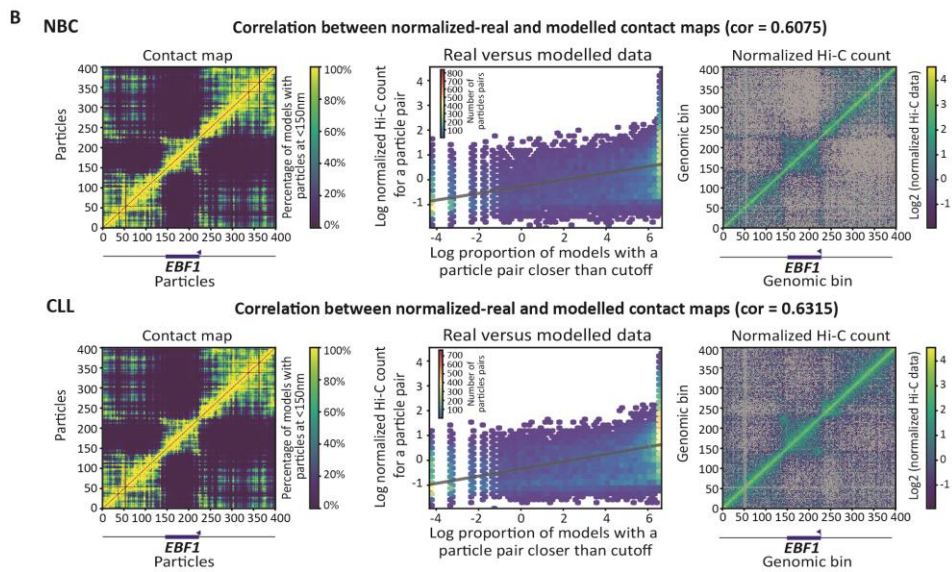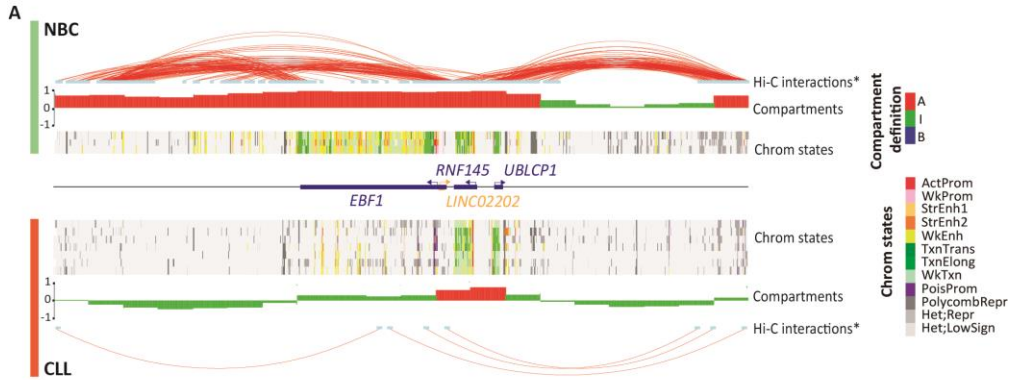
Figure 41. *EBF1* silencing in CLL is accompanied by structural changes affecting a 2Mb region.
A. Map of the *EBF1* regulatory landscape. Significant Hi-C interactions (*p*-value=0.001) and compartment type from merged NBC and a representative CLL sample, followed by chromatin state tracks from each NBC (n=3) and CLL (n=7). The coordinates of the represented region are chr5:158,000,000-160,000,000; GRCh38. **B.** Correlation between normalized Hi-C and modeled contact maps in *EBF1* regulatory landscape. *Left*: Contact map computed from the restrained-based model. *Middle*: Scatterplot of Hi-C normalized map versus modeled contact data with linear regression. *Right*: Normalized Hi-C data. *Top*: NBC. *Bottom*: CLL. The position of *EBF1* is indicated in blue at the bottom of the matrix plots. **C.** Heatmap of the hierarchical clustering of the dRMSD values computed for all the possible pairs of generated models (1,000 in NBC and 1,000 in CLL). **D.** Restraint-based model at 5Kb resolution of the 2Mb region containing *EBF1* (total 400 particles, *EBF1* locus localized from 139 to 220 particle). Data from merged NBC (*top*) and CLL (*bottom*) was used. Surface represents the ensemble of 1,000 models and is color-coded based on the compartment definition (A-type, B-type and I-type in red, blue and green, respectively). The top-scoring model is shown as trace, where protein-coding genes are colored in blue and long non-coding RNAs in yellow. Spheres represent enhancer regions. **E.** Violinplot of the convex hull volume involving the 81 particles from the *EBF1* region. Comparison was performed using Wilcoxon's test. *****p*-value=0.00001.

To obtain further insights into the *EBF1* genome structure, a model of its spatial organization in NBC and CLL was performed by using the restraint-based modeling approach implemented in TADbit (Baù and Marti-Renom, 2012; Serra et al., 2017) (**Figure 41B-41D**). The *EBF1* domain in CLL resulted in larger structural variability as compared with the models in NBC due to the depletion of interactions in neoplastic cells (**Figure 41B**). The 3D models revealed that the *EBF1* gene is located in a topological domain, isolated from the rest of the region in NBC, hosting active enhancer elements (**Figure 41D**). Remarkably, the active enhancer elements together with the interactions are lost in CLL (**Figure 41D**), resulting in more collapsed conformations (**Figure 41E**). Overall, these analyses suggest that *EBF1* silencing in CLL is linked to a compartment shift of a large genomic region leading to the abrogation of interactions and regulatory elements, which results in a significant reduction of the gene region volume.

I also detected 48 regions that changed towards more active compartment exclusively in CLL (**Figure 42A**). As expected, these regions were significantly enriched in upregulated genes (*p*-value=0.0038), harboring 9 genes with increased expression (**Figure 42B**).
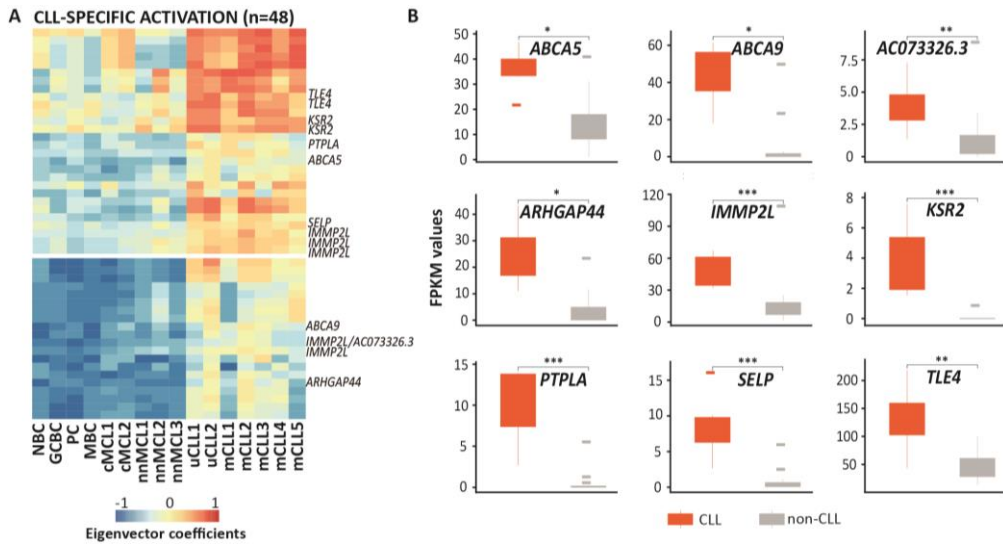
Figure 42. **CLL-specific activated compartments. A.** Heatmap showing the first eigenvector coefficients of the compartments gaining activation specifically in CLL (n=48). Significantly upregulated genes (FDR<0.05) associated to each compartment are shown on the right of the heatmap (*p*-value=0.006). **B.** FPKM values of all the CLL-specific significantly upregulated genes within compartments gaining activation. *adjusted *p*-value<0.05, **adjusted *p*-value<0.005, ***adjusted *p*-value<0.0005.

As previously shown for regions gaining activity in GCBC (**Figure 32**), I evaluated whether particular TFs were related to the CLL-specific increase in 3D interactions. Indeed, I found an enrichment in TF binding motifs of the TCF (*p*-value=0.00004) and NFAT (*p*-value=0.00647) families, which have been described to be relevant for CLL pathogenesis (Beekman et al., 2018a; Gutierrez et al., 2010; Le Roy et al., 2012) (**Figure 43A** and **Appendix Table 5B**). Of note, several motifs for the TCF4 transcription factor were found within *KSR2,* one of the nine upregulated genes in CLL-specific active compartments (**Figure 42B** and **43B**). Importantly, upregulation of this gene has a strong diagnostic value in CLL (Navarro et al., 2017), while TCF4 itself is overexpressed in CLL as compared to normal B cells (Beekman et al., 2018a), suggesting in this particular example that TCF4 overexpression may lead to aberrant binding to *KSR2* regulatory elements and a global remodeling of its 3D interactions.
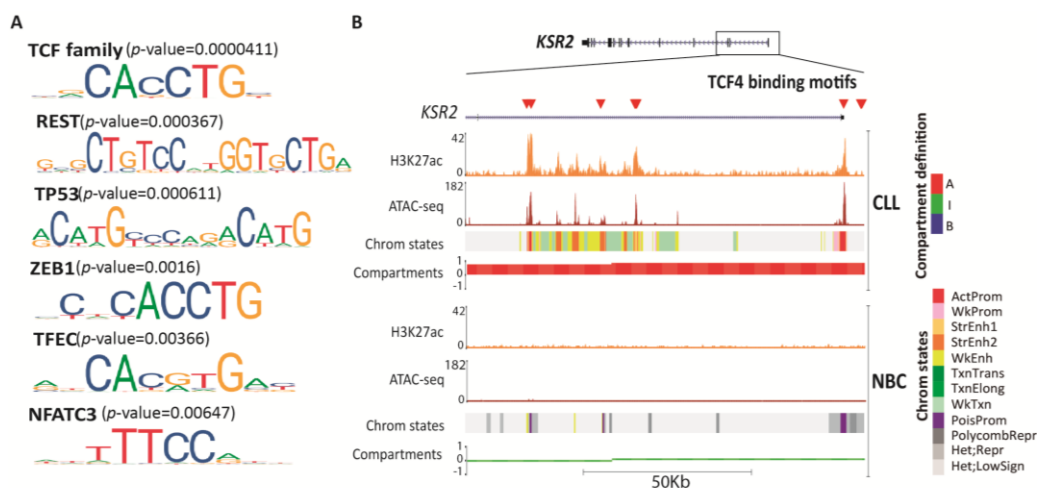
Figure 43. Transcription factors associated to CLL-specific activated compartments. A. Enrichment analysis of transcription factor binding motifs. The most significant TF binding motifs enriched in active and accessible loci within the CLL-specific regions gaining activity (n=25 independent genomic loci) versus the background (n=28 independent genomic loci) are shown. *P*-values were calculated using the AME-MEME suite. Out of the list of all enriched transcription factor binding motifs, I considered only those expressed in all CLL samples (n=7). **B.** Example of TCF4 binding motifs at the *KSR2* promoter region in CLL and NBC. The following tracks: H3K27ac signal, chromatin accessibility (ATAC-seq), chromatin states and compartment type of a representative NBC replicate and CLL sample are shown. The coordinates of the represented region are chr12:117,856,977-117,975,164; GRCh38.

2.11 Increased 3D interactions across a 6.1Mb region including the *SOX11* oncogene in aggressive MCL

In addition to entity-specific 3D genome changes, our initial analyses also suggested that different clinico-biological subtypes may have a different 3D genome organization, especially in MCL (**Figure 34B**). To identify subtype differences within each B-cell neoplasia, the regions with homogeneous compartments within each disease subtype were selected and classified as distinct if the difference between the Hi-C matrices cross-correlation eigenvalues was greater than 0.4. Applying this criterion, 47 compartment changes were defined between uCLL and mCLL, and 673 compartment changes between nnMCL and cMCL (**Figure 44A-44B**). This finding confirmed the previous analyses (**Figure 35B-E**), and indicated that the two MCL subtypes have a markedly different 3D genome organization.
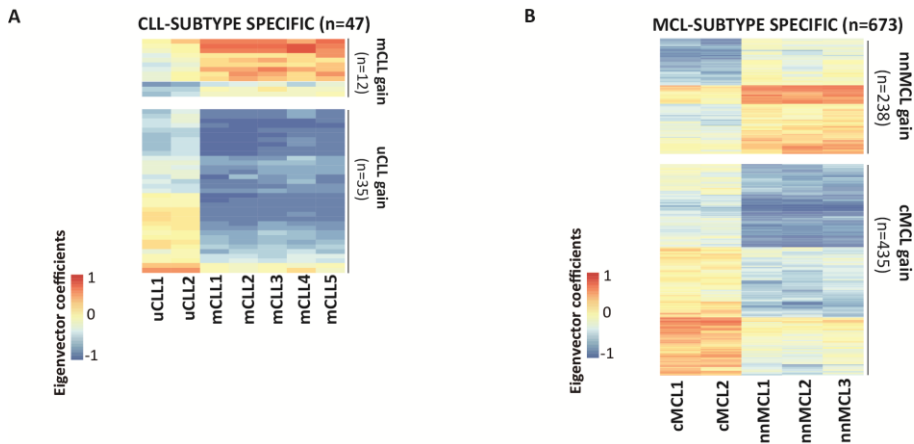
Figure 44. B-cell neoplasia subtypes specific compartments. Heatmaps showing eigenvector coefficients of compartments significantly changing in uCLL versus mCLL (n=47) (**A**), and in cMCL versus nnMCL (n=673) (**B**).

Then, I assessed the dynamic compartments between the two MCL subtypes, finding out that two thirds of the compartment changes (n=435, 64.6%) gained activity in the clinically-aggressive cMCL, and one third gained activity in nnMCL. I characterized the chromosomal distribution of these compartment shifts which, surprisingly, was significantly biased towards specific chromosomes (**Figure 45A**). In particular, those regions gaining 3D interactions in aggressive cMCL were highly enriched in chromosome 2, being 22.3% (n=97) of all 100Kb compartments located in that chromosome (**Figure 45A**). I next analyzed chromosome 2 of cMCL in detail and I observed a *de novo* gain of A-type and I-type compartments accumulated at band 2p25 as compared to both normal B cells and nnMCL (**Figure 45B**). The entire region of about 6.1Mb had a dramatic increase of interactions and active chromatin states in cMCL as compared to nnMCL (**Figure 45C** and **45D**). Most interestingly, this region contains *SOX11*, whose overexpression in cMCL represents the main molecular marker to differentiate these two MCL subtypes (Fernandez et al., 2010), and has been shown to play multiple oncogenic functions in cMCL pathogenesis (Balsas et al., 2017; Palomero et al., 2016; Vegliante et al., 2013).
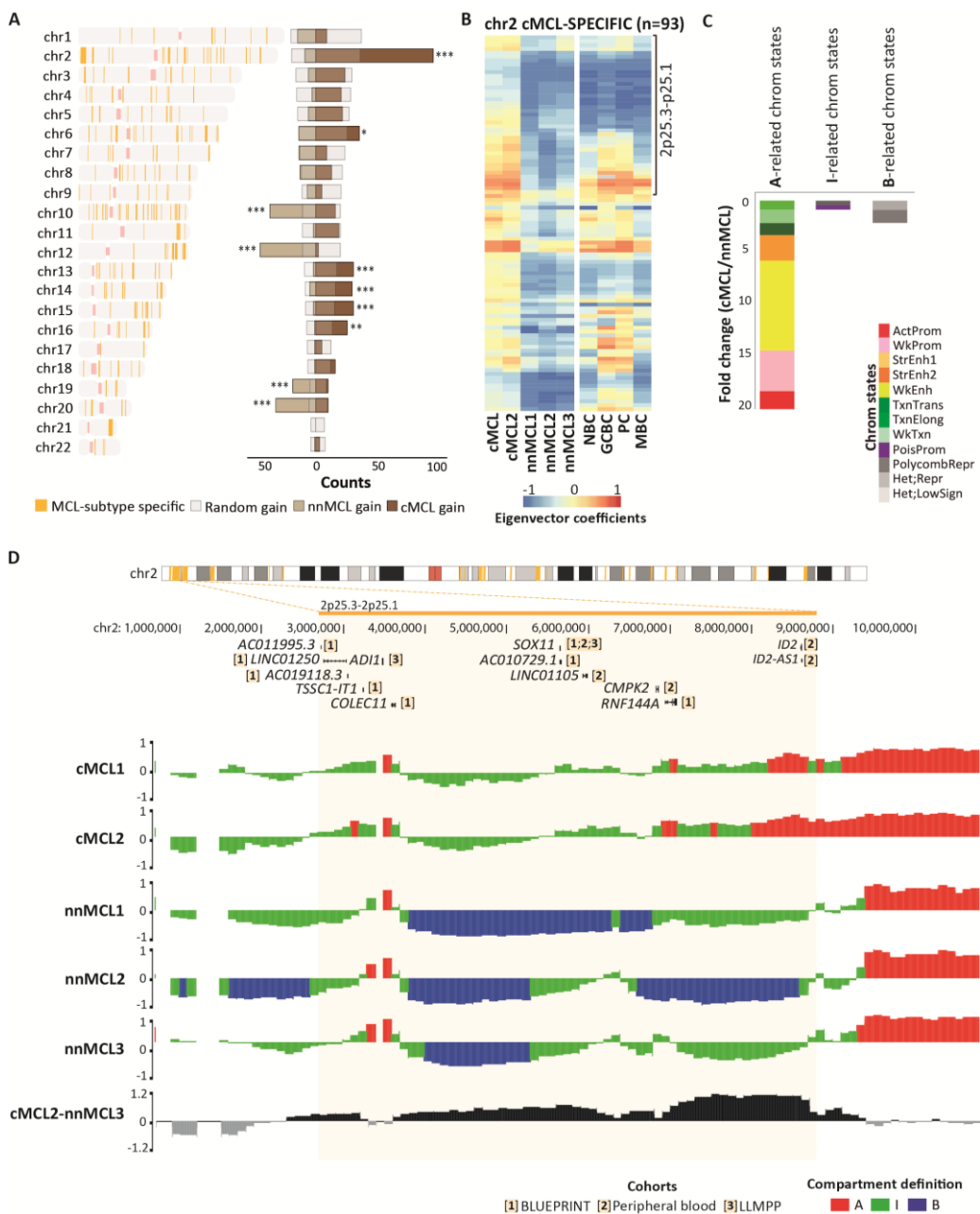
Figure 45. Long-range chromatin remodeling of a 6.1Mb involving *SOX11* in MCL. A. *Left*: Genome-wide distribution of compartments changing in MCL subtypes. The vertical orange lines point to the chromosome location of the regions. *Right*: Relative abundance of the compartments significantly gaining activity in cMCL or nnMCL as compared with a random probability. A gain in compartment activation was defined as an increase of eigenvector

coefficient of at least 0.4. *p-value<0.05, **p-value<0.005, ***p-value<0.0005. **B.** Heatmap showing eigenvector coefficients of the chromosome 2 compartments specifically gaining activation in cMCL (n=93). On the top of the heatmap, the 6.1Mb genomic block gaining activation in 2p25 is shown. **C.** Bar graphs represent the fold change between cMCL and nnMCL of each three groups of chromatin states (arranged by their relationship to the A-type, I-type and B-type compartments). Active Promoter, Weak Promoter, Strong Enhancer 1, Strong Enhancer 2, Weak Enhancer, Transcription Transition, Transcription Elongation, Weak Transcription were A-type compartment-related states. Heterochromatin-Repressed and Heterochromatin-Low signal were B-type compartment-related states. Poised Promoter or Polycomb-Repressed chromatin states were I-type compartment-related states. **D.** *Top*: Differentially expressed genes between cMCL and nnMCL in each of the three cohorts of transcriptional data of MCL patients. *Bottom*: Compartment type tracks on all the MCL samples under study and eigenvalue subtraction between representative cMCL and nnMCL samples highlighting the 6.1Mb region gaining activity in the former.

However, as *SOX11* is embedded into a large block of 6.1Mb gaining activation in cMCL, I wondered whether additional genes could also become upregulated as a consequence of the large-scale spatial organization of chromosomal band 2p25. Indeed, mining the expression data from the 5 MCL cases studied herein, as well as two additional, independent published cohorts, i.e. a series with 30 conventional and 24 leukemic non-nodal mantle cell lymphoma (GEO GSE79196) from peripheral blood (Navarro et al., 2017) and a second series from the lymphoma/leukemia molecular profiling project (LLMPP) (GEO GSE93291) (Scott et al., 2017). I observed that 13 (43%) of a total of the 30 expressed genes within the 6.1Mb region were overexpressed in cMCL as compared to nnMCL in at least one cohort (**Figure 46**). Therefore, these upregulated genes may also contribute to cMCL pathogenesis and clinical aggressiveness.

Figure 46. Cohorts used to analyse gene expression at the defined 6.1Mb region in MCL samples. A-C. Heatmaps of the differentially expressed gens in each cohort: BLUEPRINT (**A**), peripheral blood (Navarro et al., 2017) (**B**) and LLMPP (Scott et al., 2017) (**C**), with MCL samples classified as cMCL (light yellow) and nnMCL (dark yellow) subtypes. The VST values were normalized by genes.

3. Conclusions

- An integrative analysis of the Hi-C data reveals the presence of a highly-dynamic intermediate compartment enriched in poised and polycomb-repressed chromatin.

- During B-cell development, 28.1% of the compartments are dynamically modulated and mostly involve the intermediate compartment.

- The transition from NBC to GCBC is associated with widespread chromatin activation, which mostly reverts into naive state upon further maturation of GCBC to MBC.

- The 3D genome interactions are extensively modulated during normal B-cell differentiation and the genome of B-cell neoplasias acquires tumor-specific 3D genome architecture.

- The analysis of CLL and MCL neoplastic cells points both to entity and subtype-specific alterations in chromosome organization.

- Specific transcription factors for each B-cell subpopulation and neoplasia may be driving the structural organization as observed on compartment activation in GCBC and CLL.

- Large chromatin blocks containing key disease-specific genes for the pathogenesis and clinical behavior alter their 3D genome organization. These include the inactivation of a 2Mb region containing the *EBF1* gene in CLL and the activation of a 6.1Mb regions containing the *SOX11* gene in clinically aggressive MCL.

STUDY 2

Analysis of the 3D genome architecture associated with chromosomal alterations in chronic lymphocytic leukemia and mantle cell lymphoma

1. Introduction

Structural variants (SV) are defined as chromosomal rearrangements that contribute to the genetic diversity of the human genome and thus are of high relevance for cancer genetics, rare diseases and evolutionary genetics (Spielmann et al., 2018). Those alterations can be balanced, with no loss or gain of genetic material, but showing changes on a region of DNA in orientation (inversions) or chromosomal location (translocations) within or between chromosomes among individuals. In the case of unbalanced alterations a part of the genome is lost, inserted or duplicated (deletions, insertions and duplications), which are known as copy number variants (CNV) (Escaramís et al., 2015). Recent studies have shown that SVs can not only affect gene dosage but also modulate basic mechanisms of gene regulation (Klopocki et al., 2008; Lohan et al., 2014; Will et al., 2017). SVs can alter the copy number of regulatory elements or modify the 3D genome by disrupting higher-order chromatin organization such as topologically associating domains (TAD) (Northcott et al., 2017; Ordulu et al., 2016). As a result of these position effects, SVs can influence the expression of genes distant from the SV breakpoints, thereby causing disease (Lupiáñez et al., 2015).

In hematological malignancies, SVs are considered important genetic drivers of disease. Chromosomal translocations are widely studied and in several instances, constitute the primary event driving leukemo- or lymphomagenesis, and play an essential role in differential diagnosis. Classical examples are the t(9;22) in chronic myelogenous leukemia or several translocations juxtaposing the IG locus to oncogenes in B-cell malignancies. Other SVs such as deletions, duplications or inversions also play an important role in hematopoietic neoplasms, although they frequently represent secondary genetic alterations acquired in the evolution of the disease (Schütte et al., 2019). These chromosomal changes can be detected with well-established techniques such as PCR (polymerase chain reaction), FISH (fluorescence *in situ* hybridization), SNP arrays, and whole-genome sequencing (WGS) (Van Dongen et al., 2015; Jongen-Lavrencic et al., 2018; Radich, 2009; Song and Shao, 2015). Nonetheless, novel approaches to characterize structural variants in cancer are emerging, like the chromosome conformation capture methods, such as genome-wide Hi-C technique (Díaz et al., 2018; Gröschel et al., 2014; Harewood et al., 2017; Yamazaki et al., 2014).

Therefore, the aim of this chapter was to integrate the different techniques used to identify structural variants in CLL and mMCL, as well as study their role in shaping the 3D chromain structure within the nucleous.

2. Results

2.1 Chromosomal alterations as balanced and unbalanced translocations inferred by 3D maps

In the **Study 1**, I generated genome-wide chromosomal conformation maps for 5 MCL and 7 CLL cases by means of *in situ* Hi-C. On the resulting raw contact matrices at 1Mb resolution I examined the SVs per each of the studied cases. Briefly, the Hi-C contact matrices represent the intrachromosomal (*cis*) ligation events by a strong diagonal, which decreases logarithmically with genomic distance, and the interchromosomal (*trans*) interactions placed off the diagonal as blocks of strong interactions. Chromosomal aberrations leads to genome reorganization within the cell nucleous, which can be visually detected on Hi-C maps. In fact, translocations resulting in interchromosomal blocks, are easily detected chromosomal rearrangements on Hi-C heatmaps. Those translocations derived from unbalanced rearrangements seen as single blocks with strong contacts at the breakpoint or balanced which are visualized as contacts split between two blocks that produced a "butterfly" appearance (Harewood et al., 2017). To easier visualize SVs, I took advantage of the Hi-C data I generated on the B-cell subpopulations. In that way, per each contact matrix in B-cell neoplasms I could remove the interactions present on its specific cell of origin subpopulation (**Figure 47** and **48**).

I used the Hi-C data to detect and confirm the SV obtained for the very same case with other methodologies such as single nucleotide polymorphism (SNP) array, whole-genome sequencing (WGS) analyses and/or cytogenetic analyses studying the conventional karyotype. This comparison was performed using data published or available in the lab. Particularly, some MCL genomic data used in this thesis is part of a manuscript (Nadeu et al., 2020) while other data was already published (Bea et al., 2013). CLL genomic data used in this chapter have been already published (Nadeu et al., 2018; Puente et al., 2015). The conventional cytogenetics data was available in the lab.

Figure 47. MCL structural variants. *In situ* Hi-C, raw chromosome conformation capture maps at 1Mb resolution obtained subtracting the respective normal contact matrix from the

neoplastic sample. Color code indicates the difference of Hi-C interaction matrix. To each conventional MCL (cMCL) case, the naive B cell matrix was subtracted and for each leukemic non-nodal MCL (nnMCL) case the memory B cell matrix. chr, chromosome. Single nucleotide polymorphism (SNP) array and whole genome-sequencing (WGS) data is presented on a circus plot with each chromosome represented. SNP data is represented as loss (red) or gain (blue). Inner lines represents the structural variants (deletion, red; duplication, blue; inversion, green and translocation, black) obtained per WGS analysis. Cytogenetics data is the karyotype. Data available per each case is represented. Notice the order of the cases, the MCL case (cMCL2) is highlighted as contains more structural variants.

I first analyzed the presence of chromosomal translocations in the contact matrices from five MCLs. I subtracted the genome-wide chromosomal conformation interactions of NBC from the cMCL (n=2), and the interactions of MBC were removed from nnMCL(n=3) cases, as these cells have been reported as putative cells of origin of the two types of MCLs (Navarro et al., 2012; Royo et al., 2012), (**Figure 47**).

In the MCL Hi-C heatmaps, patterns compatible with the presence of the t(11;14) translocation were clearly observed in all cases. Furthermore, additional patterns related to other balanced and unbalanced translocations were observed. Next, I compared the accuracy in detecting translocations among three different methods i.e. *in situ* Hi-C, WGS and cytogenetics (**Table 4**). Also, copy numbers alterations (CNA) by SNP array were analyzed if translocations were not detected by WGS. The data available per each particular case was correlated.

Neoplasia	Case	Translocation	Methodology of detection			
			In situ Hi-C	WGS	SNP (CNA)	Cytogenetics
MCL	cMCL1	t(11;14)	Detected	Detected	Not applicable	Detected
		t(3;6)	Detected	Not detected	Pattern compatible	Not detected
	cMCL2	t(11;14)	Detected	Detected	Not applicable	Detected
		t(3;18)	Detected	Not detected	Pattern compatible	Not detected
		t(10;14)	Detected	Detected	Not applicable	Detected
		t(8;13)	Detected	Not detected	Pattern compatible *	Not detected
		t(8;19)	Not detected	Detected	Not applicable	Detected
		t(15;21)	Detected	Detected	Not applicable	Detected
		t(5;15)	Detected	Detected	Not applicable	Detected
	nnMCL1	t(11;14)	Detected	No data available	No data available	No data available
	nnMCL2	t(11;14)	Detected	Detected	Not applicable	Detected
	nnMCL3	t(11;14)	Detected	Detected	Not applicable	Detected
		t(8;17)	Detected	Not detected	Pattern compatible	Detected
CLL	uCLL1	t(3;6)	Detected	Not detected	Not applicable	Not detected
	uCLL2	t(2;18)	Detected	Not detected	Pattern compatible	Not detected
		t(4;8)	Detected	Not detected	Detected	Not detected

Legend: ▬ Detected ▬ Not detected ▬ No data available ▬ Pattern compatible with transl ▬ Not applicable

Table 4. Summary of MCL and CLL translocation, detected by different techniques.

Translocations identified in *in situ* Hi-C matrices, WGS and/or cytogenetic analyses. Copy number alterations (CNA) by SNP array were examined if translocations were not detected by WGS. Color code indicates: yellow, translocation detected; blue, translocation not detected and grey, data not available. *CNA observed in chromosome 8 but not in chromosome 13.

Interestingly, I observed that most of the translocation events could be identified by performing *in situ* Hi-C as compared to the rest of the methodologies considered. For instance, the chromosomal conformation map of cMCL2 clearly revealed the presence of a t(8;13) translocation which seemed to involve the end of the chromosomes, and therefore unable to be detected by WGS analysis. That could also be the explanation for the t(3;6) translocation detected only in the contact matrix of cMCL1 case. Intriguingly, in both cMCLs part of the long arm of chromosome 3 (3q) was visually translocated to different chromosomes, i.e. to chromosome 6 in cMCL1, t(3;6), and to chromosome 18 in cMCL2, t(3;18), (**Figure 47**). In addition, the contact matrices of both cMCLs also showed an increase of interactions at the translocated part the chromosome 3 territory. These results suggest that in both cMCL cases, the chromosomal alterations observed in chromosome 3 may also involve gain of genetic material events.

Then, I analyzed the presence of chromosomal translocations in the CLL cases. Similarly as performed for MCL cases, the interaction contact matrix of the specific normal counterpart was subtracted from each neoplastic case (**Figure 48**). From the 1Mb chromosomal contact map of uCLL (n=2), the NBCs interactions were removed, while for the mCLL (n=5), the MBCs interactions were eliminated, as putative cells of origin (Seifert et al., 2012).

I assessed the presence of translocations in the contact matrices, and I could detect one translocation in uCLL1 t(3;6) and two translocations in the uCLL2 case, t(2;18) and t(4;8). As performed with MCL, translocations detected by Hi-C were evaluated with data available on the very same case from two more methodologies, WGS and cytogenetics. However, those two translocations on uCLL2 cases were not detected neither by karyotyping nor WGS (**Table 5**). Then, I considered the SNP array data to study whether copy number changes were occurring in those chromosomes where a translocation could be defined by Hi-C data. Indeed, a gain on chromosome 2

and a loss on chromosome 18 in uCLL2 was detected by SNP array, which on the contact matrix was observed as a t(2;18) translocation.
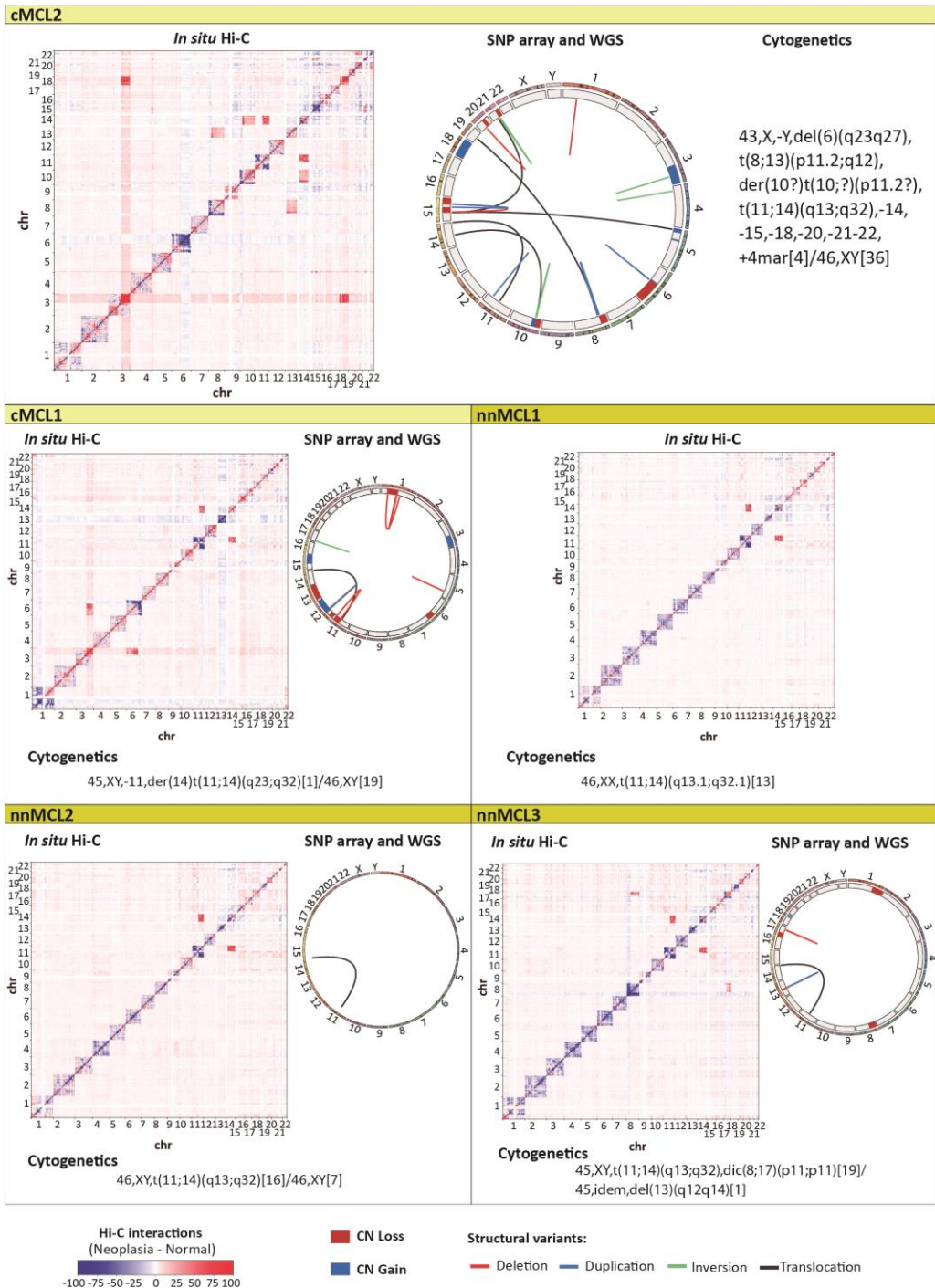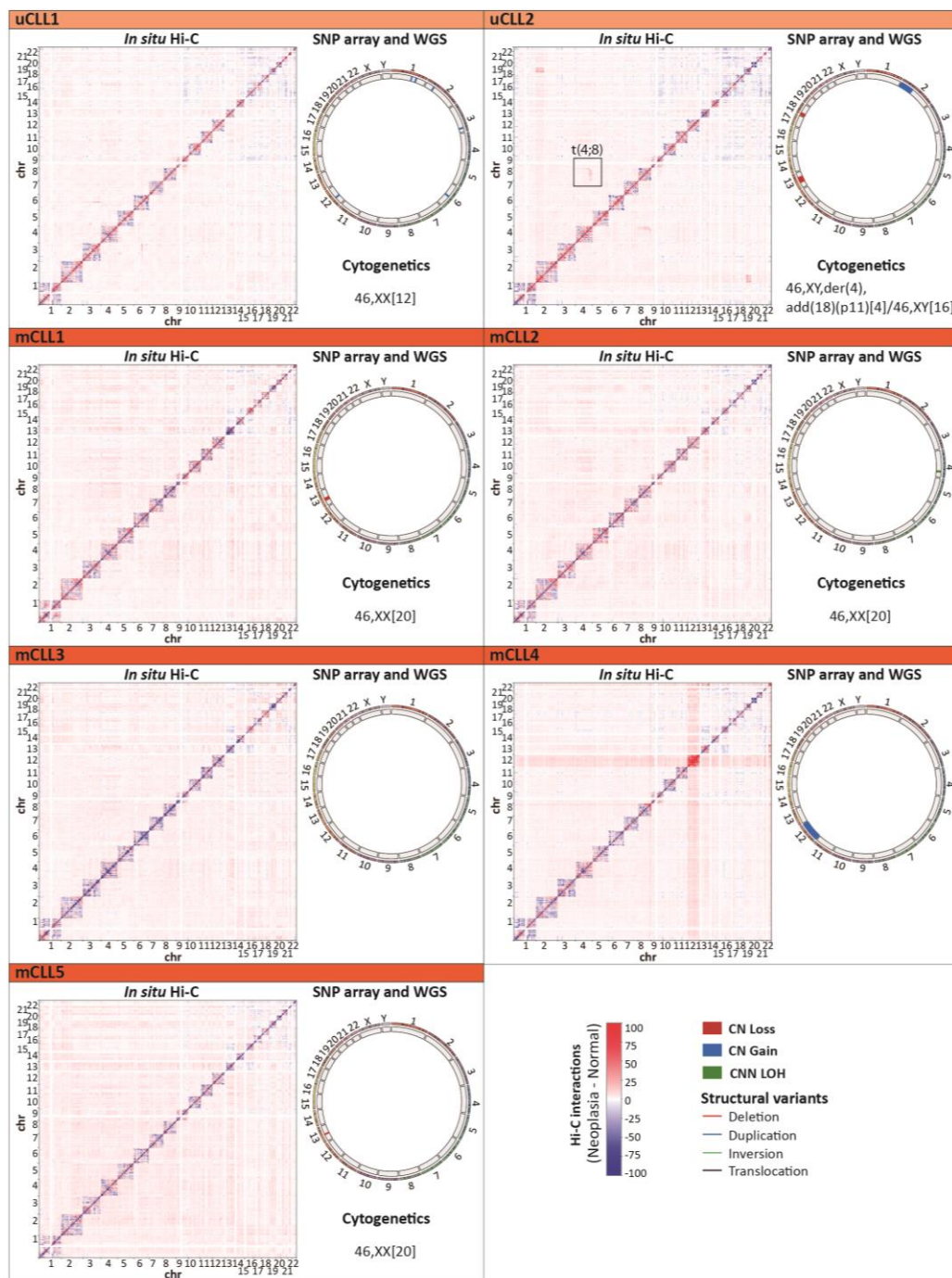

Figure 48. CLL structural variants. *In situ* Hi-C, raw chromosome conformation capture maps at

1Mb resolution obtained subtracting the respective normal contact matrix from the neoplastic sample. Color code indicates the difference of Hi-C interaction matrix. To each unmutated CLL (uCLL) case, the naive B cell matrix was subtracted and for each mutated CLL (mCLL) case the memory B cell matrix. chr, chromosome. SNP array and WGS data is presented on a circus plot with each chromosome represented. Copy number variation (CNV) data is represented as loss (red) or gain (blue). Inner lines represents the structural variants (deletion, red; duplication, blue; inversion, green and translocation, black) obtained per whole genome-sequencing (WGS) analysis. Cytogenetics data is the karyotype. Data available per each case is represented.

Besides, cytogenetic analysis in the uCLL2 cases identified the existence of a derivative chromosome involving chromosome 4, suggesting that this chromosome was structurally rearranged. In the interaction matrix, however, a "butterfly" appearance pattern could be observed between chromosome 4 and chromosome 8 suggesting a balanced translocation t(4;8), (squared in **Figure 48**).

To sum up, chromosomal translocations can be visually detected easily on the chromosomal conformation maps at 1Mb resolution. This stands in contrast to cytogenetic analyses, where reciprocal translocations implicating small regions could not be determined. Hi-C also allowed the detection of translocations implicating the extreme of the chromosomes, which could not be identified by WGS analysis.

In addition to chromosomal translocations, I could also visually detect chromosomal deletions in both neoplasias. For example, the chromosome territory of chromosomes 1, 6, 11 and 13 of cMCL1 displayed less Hi-C interactions (observed as more intense blue spots at the diagonal) suggesting losses of genetic material, which were validated by SNP array and WGS data as shown at the circos plot (**Figure 47**). Furthermore, small losses of chromosome 13 in uCLL2, mCLL1 and mCLL2 were detected with genomic analysis (circos plot) and could be also visually observed at the Hi-C contact matrices (**Figure 48**). Interestingly, the contact matrix of mCLL4 showed an increase in interactions at the entire chromosome 12. SNP array data confirmed the trisomy 12, a frequent aberration detected in up to 20% of CLL cases (Matutes et al., 1996; Swerdlow et al., 2017). Overall, chromosomal conformation maps showed less structural rearrangements in CLL cases as compared to MCL. These results are in line with multiple published studies showing that MCL cases present a high number of structural alterations per case, and that CLL cases frequently show a low number of SVs (Bea et al., 2013; Puente et al., 2018).

Overall, these analyses demonstrated that Hi-C data seems to be the most reliable method to identify most of the SVs. Due to its current technical complexity and elevated cost, it is unlikely that it will be introduced in the diagnostic setting, but shall be considered if a detailed characterization of SVs is needed, for instance in the context of research of genome instability analyses.

2.2 TAD borders are enriched in structural variants

Next, I wanted to analyze the relationship between SVs and modulation of 3D structure. As observed in the previous analyses, MCLs are clearly more unstable than CLLs, so I centered this analysis on the MCL cases. I wondered whether those cancer-specific SVs were somehow associated with the topologically associating domains (TAD) borders, and if so how they affect the chromatin structure. Several scenarios were considered (**Figure 49A**). For instance, I hypothesized that SVs could create new cancer-specific TAD borders. However, other possibilities may be contemplated, i.e. SVs might not affect the TAD borders but just take place in preexisting TAD borders (**Figure 49A**).

To allow a proper study of the impact of the structural variants on TADs in MCL, I first defined and characterized the TADs per each sample at a 50Kb resolution. I characterized these building blocks looking at the number of TADs, which seem to be similar between the normal B cells (NBC and MBC) and MCL cases (**Figure 49B**). I also analyzed each B-cell replicate separately (**Figure 49C**) and no differences in the number of TADs was detected with the merged replicates of NBC and MBC. Also, I analyzed the number of TAD on the other B-cell neoplasia, CLL, and I could determine that the number of TADs was similar to MCL suggesting that the number of building blocks is more or less homogeneously in normal and neoplastic B cells (**Figure 49C**). Moreover, I also analyzed the mean size per TAD between the normal B cells, NBC and MBC, and the MCL cases, being similar in all the samples (**Figure 49D**). The mean size of the TADs was maintained when looking at the B-cell subpopulation replicates separately and also when comparing with CLL (**Figure 49E**). Overall, MCL samples did not presented alterations in the general TADs properties as compared to B cells and also CLL.
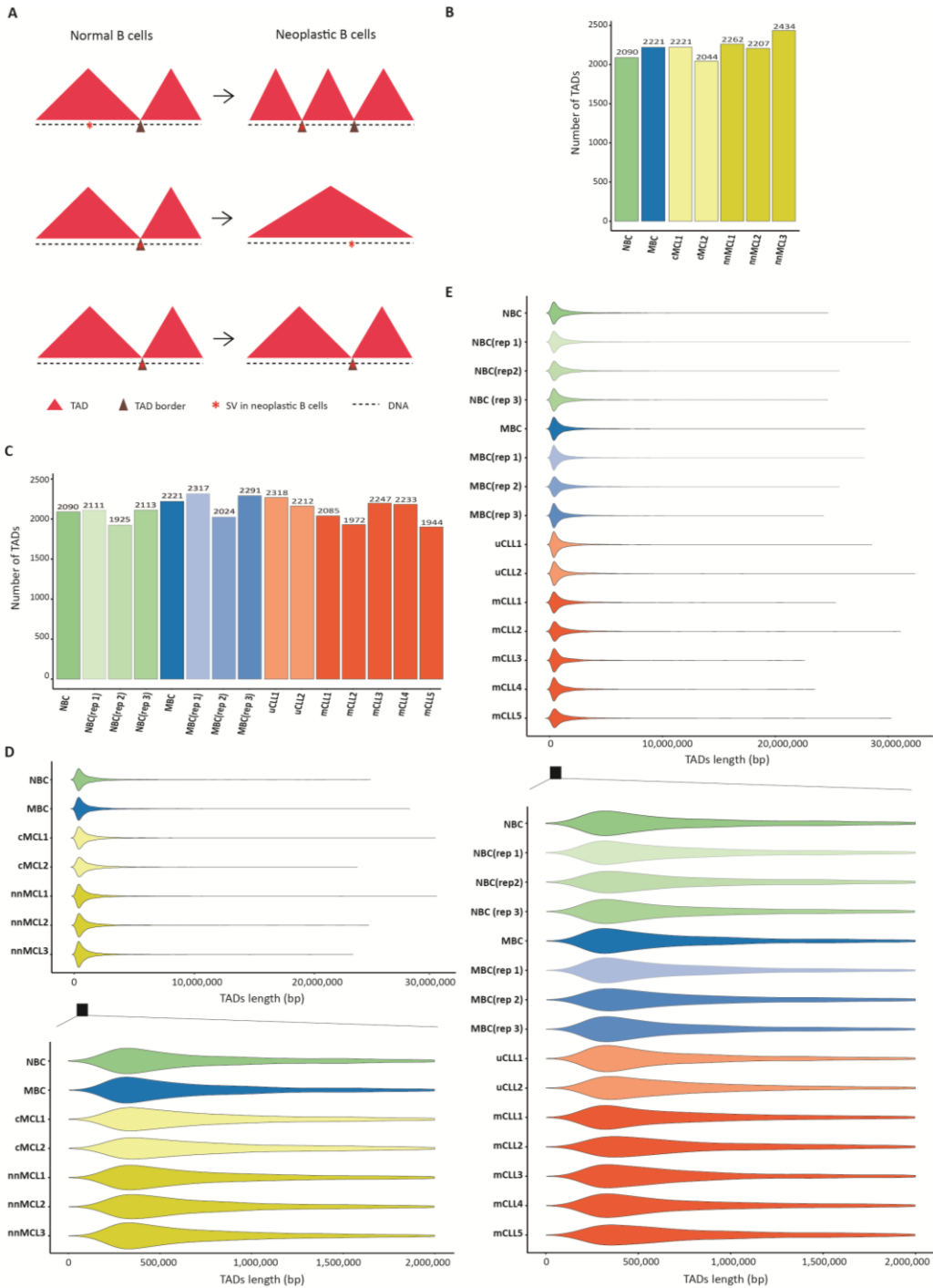
Figure 49. Topologically associating domains characterization. A. Different scenarios in which neoplastic B-cell structural variants (SV) could modulate the 3D structure. Three options may be considered: SV may be involved in creating a new topologically associating domain (TAD) border

(*upper* panel), fussing to TADs (*middle* panel) or not modifying the TAD border (*lower* panel), as compared to normal B cells. **B** and **C.** Number of topologically associating domains (TADs) in NBC, MBC and MCL cases (**B**)**,** NBC and MBC (merged and replicates) and CLL cases (**C**). **D** and **E.** Violin plots representing distribution of TAD lengths (*upper* panel) and a zoom into the 0 to 2Mb distribution (*lower* panel) in NBC, MBC and MCL cases (**D**), NBC and MBC (merged and replicates) and CLL cases (**E**).

Next, per each MCL case I considered the SVs determined by WGS analysis with its own defined TADs. Particularly, from five MCL cases, there were four in which WGS was performed. From those, this technique identified SVs in three cases. A total of 37 SVs were determined, being 10 from cMCL1, 19 from cMCL2, 1 from nnMCL2 and 7 from nnMCL3 (**Appendix Table 8**). These SVs created 74 breaks as each of the SVs was associated with two different breaks. Then, I analyzed if each specific break position was located on a TAD border. To do so, a window of 50Kb around the TAD border coordinate (25Kb up- and downstream from the TAD border coordinate) was considered in order to overcome the resolution drawback between both techniques, as TADs were defined at a 50Kb resolution while the SVs by WGS were defined at the single bp resolution (**Figure 50A**). A window of 100Kb was also analyzed (50Kb up- and downstream from the TAD border coordinate) but I observed that most of the SVs were located more proximal to the TAD border coordinate. Therefore, a window of 50Kb was used for further analyses (**Figure 50A**).
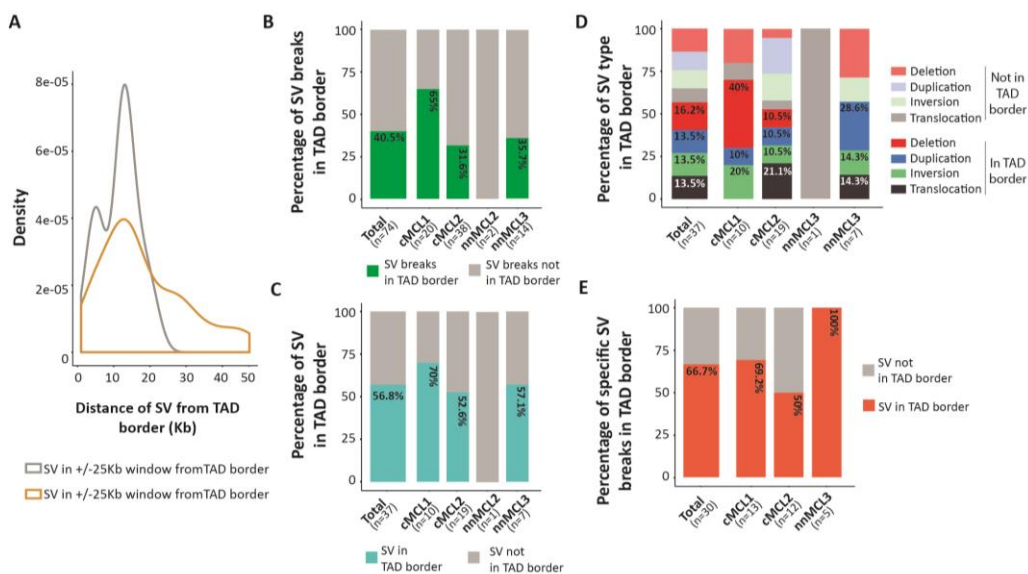
Figure 50. **Structural variants located in TAD borders. A.** Distribution of the structural variants (SV) located at TAD borders in a window of 50Kb or 100Kb from TAD border coordinate. **B to E.** Percentage of SV breaks (**B**), SV (**C**), types of SV (**D**) and case-specific SV breaks (**E**) located in TAD borders using a 50Kb window for total cases and each case individually. The *n i*ndicates total number of SV breaks or SV per sample.

The analysis overlapping TAD borders and SVs revealed that 40.5% (30/74) of the total SVs breaks were taking place in TAD borders (**Figure 50B**) which accounted for 56.8% (21/37) of the total structural variants (**Figure 50C**). However, there was no association of specific SV type (i.e. translocations, inversions, duplications or deletions) at those borders (**Figure 50D**). A step further was done analyzing whether those TAD borders were newly created in a specific MCL case as compared to normal B cells (NBC and MBC) and the rest of MCLs. To do so, each case-specific SV break was overlapped with the TAD structure of the rest of the samples. From this analysis, I detected that 66.7% (20/30) of all the SVs breaks located at TAD borders were creating a *de novo* TAD border on the specific MCL case bearing the SV (**Figure 50E**).

Next, I focused on the t(11;14) translocation, the primary genetic change occuring in more than 95% of MCL cases which leads to the upregulation of *CCND1* (Jares et al., 2012b). I observed that the chromosome 11 breaks mainly fall on a TAD border. Next, a more detailed analysis was performed to characterize the TAD borders in the *CCND1* region using a 20Kb resolution. Interestingly, the 11q13 breakpoint of the aforementioned translocation in each case was coinciding with a new TAD border (**Figure 51**). This finding suggests that the break might lead to a *de novo* 3D chromatin reconfiguration of the region, creating a tumor-specific TAD which would favor *CCND1* overexpression.
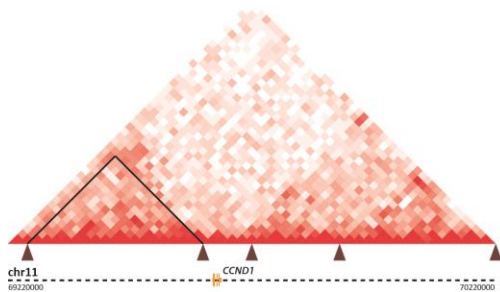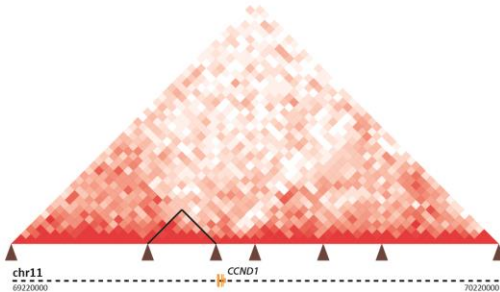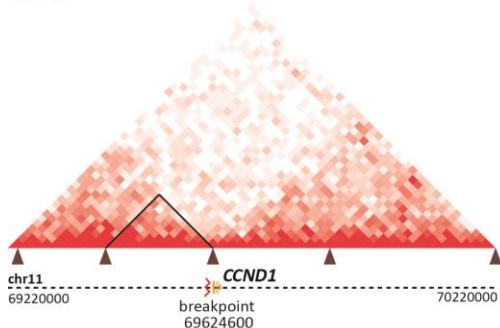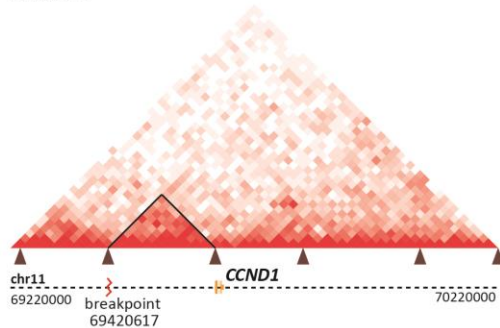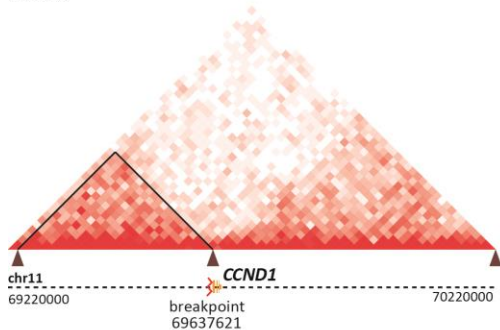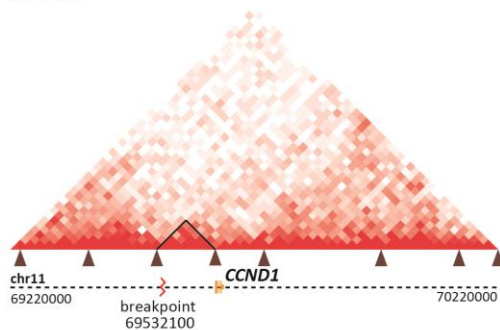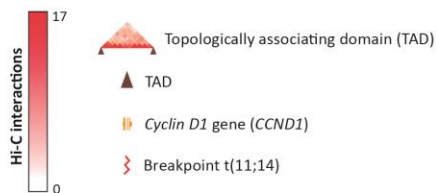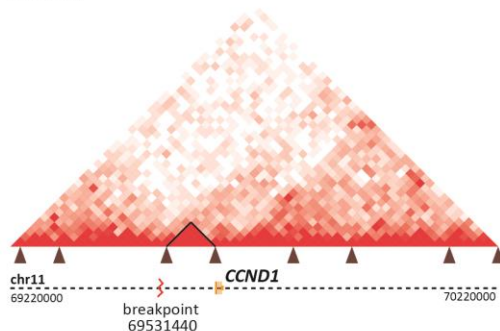
Figure 51. Characterization of TADs within the locus of translocation leading to *CCND1*

overexpression. Hi-C contact matrices for normal B cells and its related distinguished MCL subgroups. NBC and the two related cMCL cases and MBC and the three related nnMCL cases. Color code indicated the Hi-C interactions. The location of the *CCND1* gene, the TAD borders and t(11;14) breakpoint are marked accordingly. The coordinates of the represented region are chr11: 69,220,000-70,220,000; GRCh38.

The chromosome 11 breakpoint prompted a gain of regulatory marks likely caused by the fusion with the active enhancer/promoter region of the IGH in chromosome band 14q32 (**Figure 52**). Of note, *CCND1* was always found at the distal border of the TAD, confirming that these novel chr11 promoter/enhancer regions contribute to dysregulate *CCND1* by creating a new specific building blocks.
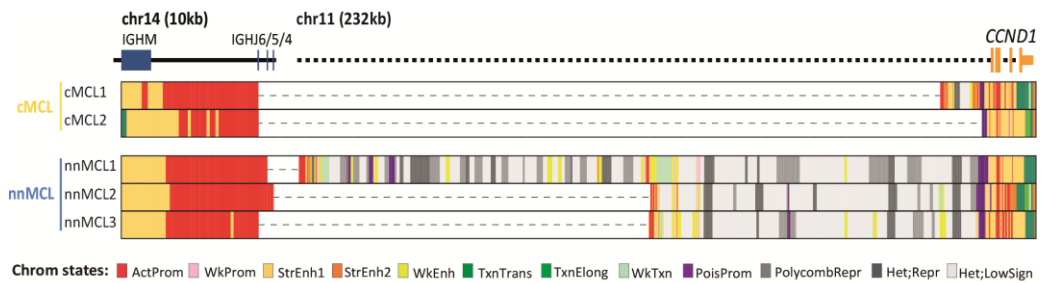
Figure 52. Functional reconstruction of the translocated *CCND1* region on MCL cases. Building of the chromatin states of the IG/*CCND1* translocated allele from chromosome 14 and chromosome 11, respectively. Dashed lines indicate the junction of both chromosomes to maintain the breakpoints relative to *CCND1*. ActProm-, Active Promoter; WkProm, Weak Promoter; StrEnh1, Strong Enhancer 1; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; TxnTrans, Transcription Transition; TxnElong, Transcription Elongation; WkTxn, Weak Transcription; PoisProm, Poised Promoter; PolycombRepr, Polycomb Repressed; Het;Repr, Heterochromatin-Repressed; Het;LowSign, Heterochromatin-Low Signal.

2.3 Association between MCL mutations and TAD borders

Apart from chromosomal alterations, I wondered whether TAD borders were enriched in mutations, which could also modify the chromatin structure, as explained for SVs (**Figure 49A**). To do so, the mutations determined per each case by WGS were interrogated following the very same method as for structural variants, using a 50Kb window around the TAD border coordinate. I observed that mutations were not specifically enriched within TAD borders. However, I analyzed the kataegis, or clusters of mutations, in each MCL case with WGS data, i.e. 6 mutations forming one cluster in

cMCL1, 56 mutations forming 5 different clusters in nnMCL2 and 63 mutations constituting also 5 different clusters in nnMCL3. Despite the low number of samples, approximately 20% of those mutations forming clusters were located in TAD borders. Specifically, one TAD border in chromosome 11 of the nnMCL2 case contained 9 mutations forming a kataegis and another in chromosome 2 of the nnMCL3 case encompassed 15 mutations. Those mutations were associated with the AID signature.

In spite of the rather limited sample size, all these analyses start to shed light into the importance of the chromatin structure as a new layer of information to consider when studying cancer-specific genetic alterations and data integration as demonstrated in some studies (Chaisson et al., 2019; Harewood et al., 2017). In the future, more samples shall be processed and integrated to perform significant analyses and reach more robust conclusions.

3. Conclusions

- Chromosomal alterations and specially translocations can be easily detected by visual inspection of chromosomal conformation maps.

- The use of Hi-C contact matrices allowed the identification of chromosomal rearrangements not detected by WGS or conventional cytogenetics.

- MCL structural variants were mainly located in *de novo* specific TAD borders.

- The MCL pathogenic translocation t(11;14) leads to a reconfiguration of the 3D chromatin structure as compared to their normal counterparts.

- MCL mutations were not significantly enriched in TAD borders at 50Kb resolution.

STUDY 3

Characterization of DNA methylome of mantle cell lymphoma and the aberrant SOX11 expression in clinically aggressive MCL

1. Introduction

Mantle cell lymphoma (MCL) is a B-cell lymphoma showing a broad spectrum of clinical behaviors and biological features (Jares et al., 2012c). In spite of the heterogeneity, the unifying factor in MCL is the t(11;14)(q13;q32) translocation leading to *cyclin D1* (*CCND1)* gene deregulation, which is considered to be a primary driver event in this disease (Jares et al., 2007). Two MCL groups have been recognized as clinically-aggressive conventional MCL (cMCL) and clinically-indolent leukemic non-nondal MCL (nnMCL) (Jares et al., 2012b; Navarro et al., 2012; Puente et al., 2018; Royo et al., 2012).

Although several studies attempt to describe MCL on molecular level, mostly from genetic prespective, the role of epigenetic changes such as DNA methylome in MCL remains largely unknown, as it has only been analyzed in promoter regions (Enjuanes et al., 2013; Halldórsdóttir et al., 2012; Leshchenko et al., 2010; Rahmatpanah et al., 2006). Therefore, the aim of this study was to obtain deeper insights into MCL epigenetics by applying an analytic strategy to deconstruct the DNA methylome of MCL in the context of the complete normal B-cell differentiation. Besides, an integrative whole-genome analysis of DNA methylation data combined with histone modifications, 3D looping and gene expression allowed us to detect novel distant regulatory elements associated with the pathogenesis.

2. Results

2.1 Sample recruitment

A total of 82 MCL samples were obtained from the Tumor Bank of the Hospital Clínic in Barcelona and the Institute of Human Genetics in Kiel. From those samples, IGHV-IGHD-IGHJ rearrangements and mutational status were analyzed using leader or consensus primer for the IGHV FR1 along with appropriate consensus primers, as previously described (Navarro et al., 2012), where sequences with ≥98% identity to the germ line were considered unmutated. In addition, SOX11 expression was evaluated either by quantitative RT-PCR, gene expression profiling or immunohistochemistry, which allowed to categorize the samples as positive or negative, as previously described (Navarro et al., 2012). Finally, mutations in *TP53, WHSC1*, *BIRC3, NOTCH2*, *MEF2B* and *TLR2* genes were detected by whole-exome sequencing (WES) or by Sanger

144 |

sequencing and analyzed with the Mutation Surveyor® (Softgenetics) (Bea et al., 2013). Clinical and biological characteristics of the patients are shown in **Appendix Table 9**. Besides, DNA methylation data from 10 sorted B-cell subpopulations spanning the entire B-cell differentiation, i.e., uncommitted hematopoietic progenitor cells, preB1 cells, preB2 cells, immature B cells, naive B cells from tonsils and peripheral blood, germinal center B cells, memory B cells from peripheral blood and plasma cells from tonsils and bone marrow (3-9 replicated of each ending up with 67 samples), were used as controls (Kulis et al., 2015). Lastly, 4 MCL cell lines were used as cellular models of MCLs. Specifically, the t(11;14)(q13;q32)-positive on MCL cell line Z-138, GRANTA-519 and JeKo-1 (SOX11-positive) and JVM-2 (SOX11-negative).

2.2 *In silico* purification of the samples

Genome-wide DNA methylation profiles of the 82 MCL samples were generated using the HumanMethylation450K BeadChIP (Illumina Inc), which covers 99% of RefSeq genes and 96% of CpG islands (Bibikova et al., 2011). Before deconstructing their DNA methylome, two potential confounding variables were studied which may affect the epigenomic MCL analyses, i.e. the biological origin of the samples (lymph node vs. peripheral blood) and the tumor cell content. In detail, from the first possible confounding factor, any consistent differential DNA methylation pattern between lymph node and peripheral blood sample was identified. In contrast, in spite of the high tumor cell content of the selected samples (median 89%, range=56% to 100%, **Appendix Table 9**), the purity was identified to be affecting the DNA methylation analyses (**Figure 53**).
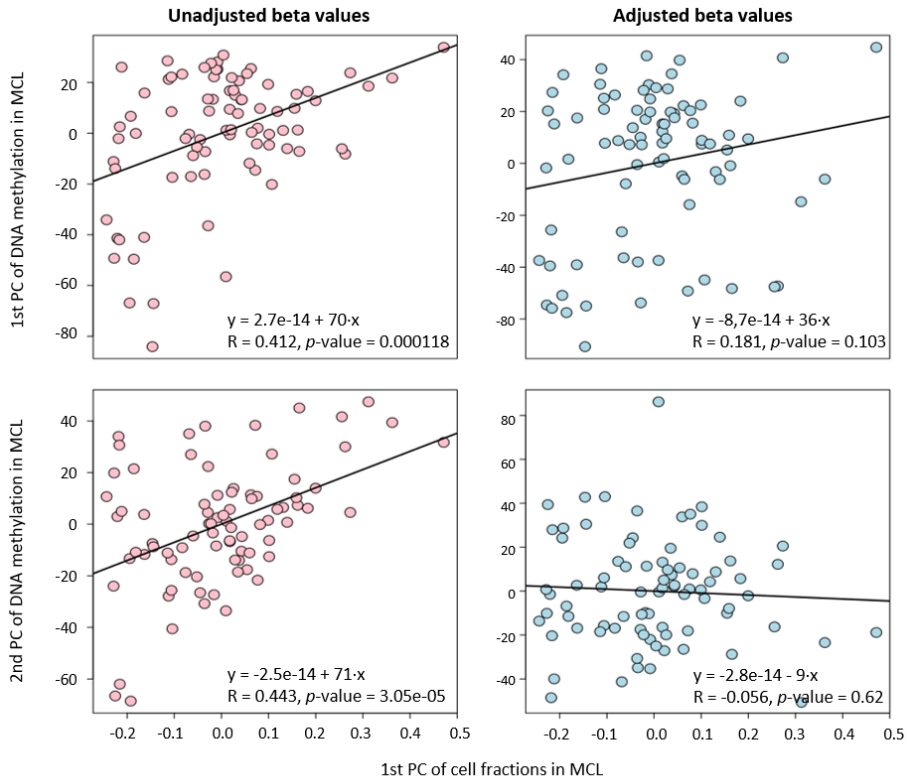
Figure 53. Effect of cell composition in DNA methylation variability. First and second Principal Component of DNA methylation in MCL samples against first Principal Component of the proportion of all 6 hematopoietic cell types in MCL before (*left*) and after (*right*) correcting DNA methylation estimates according to MCL purities. PC, Principal Component.

Consequently, a strategy to deconvolute the DNA methylation signal of mixed subpopulations and to isolate *in silico* the DNA methylation levels of the tumor cells was developed (**Figure 54A**). To do so, a statistical approach adapted for 450K Illumina arrays (Houseman et al., 2012; Jaffe and Irizarry, 2014) was tuned in order to estimate the fractions of 6 different hematopoietic cell types, i.e., B cells, CD8[+] T cells, CD4[+] T cells, NK cells, monocytes and granulocytes (Reinius et al., 2012) in MCL tumor samples, (**Figure 54B**). The normal B-cell fraction in MCL samples is estimated to be very low (0-0.3%) (Saba et al., 2016), therefore the total B-cell fraction was taken as a measure for the tumor fraction.
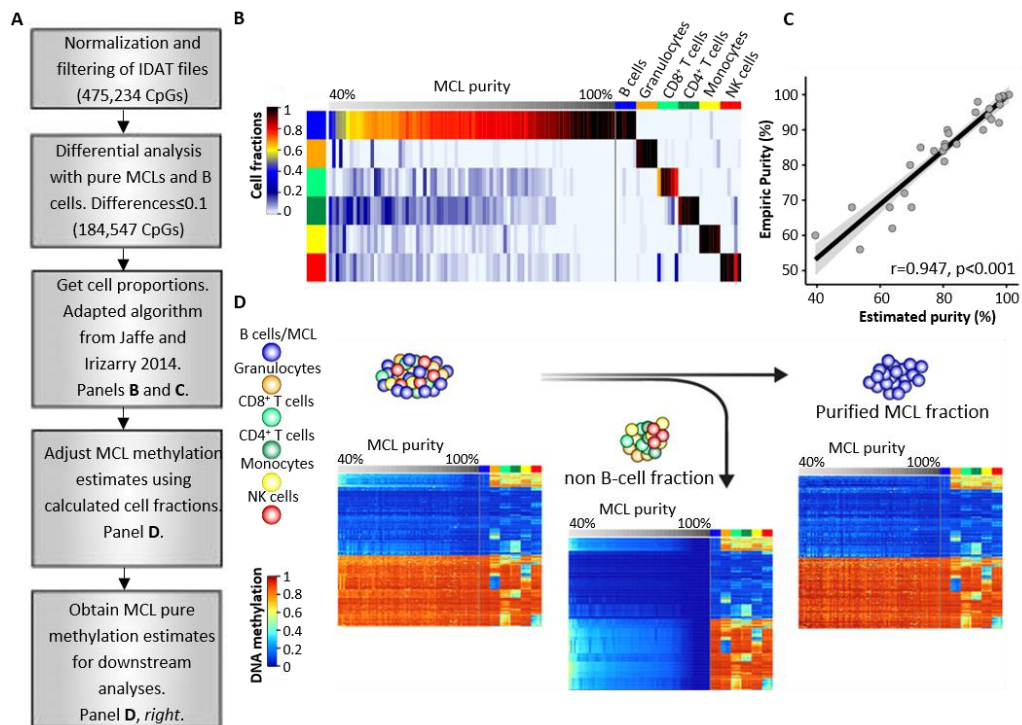
Figure 54. Deconvolution of DNA methylation data and *in silico* purification of MCL methylation estimates. A. Work flow of the deconvolution process in MCL samples. **B.** Estimation of the proportion of hematopoietic cell subpopulations in MCL samples and in sorted B cells, CD8+ T cells, CD4+ T cells, NK cells, monocytes and granulocytes. Sorted cell subpopulations (*right* part of the heatmap) are correctly predicted and MCLs show a gradient from lower to higher proportion of B cells (*left* part of the heatmap). **C.** The proportion of B cells in MCL samples as detected by flow cytometry and by the *in silico* prediction are highly correlated. **D.** Heatmaps of the CpGs representative of each cell type (n=580) showing the initial methylation estimates the MCL samples (*left*), the extraction of the DNA methylation signature from contaminating non-B cells (*middle*) and the final *in silico* purification of the DNA methylation estimates from MCL cells (*right*).

From the published algorithm (Jaffe and Irizarry, 2014) an adapted algorithm with three added steps was build, 1) selection of *a priori* set of CpGs for the deconvolution process, 2) capacity to remove unwanted CpGs and 3) ability to extract the set of CpGs used for the deconvolution. This new function was firstly tested on the selection of 32 MCLs samples with high purity (≥99, based on FACS data) which were then compared to flow-sorted purified B cells (Reinius et al., 2012). From this step, those CpGs showing

similar DNA methylation values were retained (i.e. a difference≤0.1 in methylation levels, n=184,547 CpGs). These CpGs represented those whose methylation levels globally remained stable in MCL lymphomagenesis and therefore were used for the deconvolution process. They were integrated into the modified version of the aforementioned algorithm. Initially, this approach was validated by comparing the *in silico* estimated tumor B-cell fraction with the sample purity measured by flow cytometry in 32 MCL samples (Pearson coefficient between *in silico* estimation and FACS data of r=0.947, *p*<0.001, **Figure 54C**), obtaining a more precise tumor B-cell content estimate than with the initial approach (Jaffe and Irizarry, 2014). Finally, the DNA methylation estimates of the normal non-B cell subtypes together with their respective proportions were used to extract the DNA methylation signature derived from the tumor B cells in each MCL sample (**Figure 54D**). These pure DNA methylation estimates of the tumor fraction were used for all downstream analyses.

2.3 Genome-wide DNA methylation analysis reveals two major MCL subgroups with distinct clinico-biological features

Firstly, an unsupervised principal component analysis (PCA) of DNA methylation data from normal B-cell subpopulations and MCL samples was performed (**Figure 55A**). The two first components ordered normal B cells according to their maturation stage, mainly separating germinal center-inexperienced B cells (uncommitted precursors, pre-B cells and NBC) from germinal center-experienced B cells (GCBC, MBC and PC). Particularly, principal component 1 showed that all MCLs are globally more similar to germinal center-experienced B cells (i.e. antigen experienced). In contrast, principal component 2 split MCLs into two subgroups named as cluster 1 (C1) (n=62) and cluster 2 (C2) (n=20), which respectively showed a DNA methylation pattern more similar to germinal center-inexperienced B cells and germinal center-experienced B cells. In fact, these two subgroups presented also significant clinico-biological differences (*p*<0.001) in e.g. IGHV mutation levels, SOX11 expression, number of copy number alterations, nodal presentation, and requirement of treatment at diagnosis (**Figure 55B**). Furthermore, C1 cases showed a significantly worse overall survival than C2 (*p*=0.026) (**Figure 55C**).
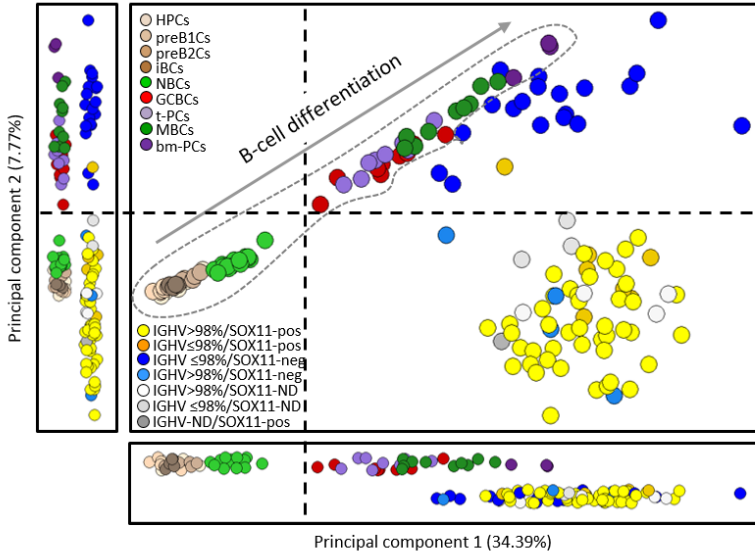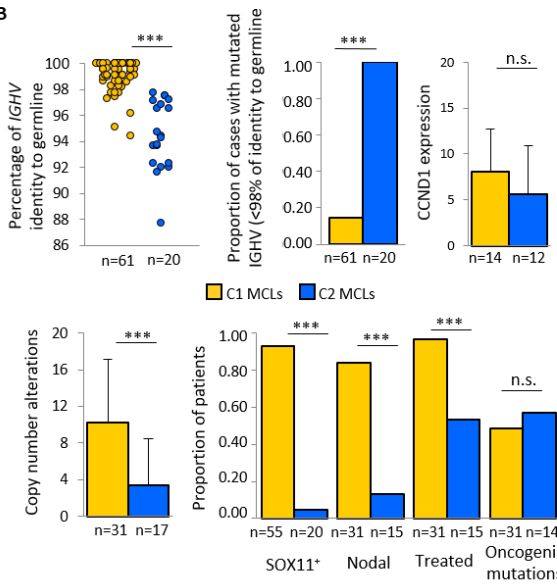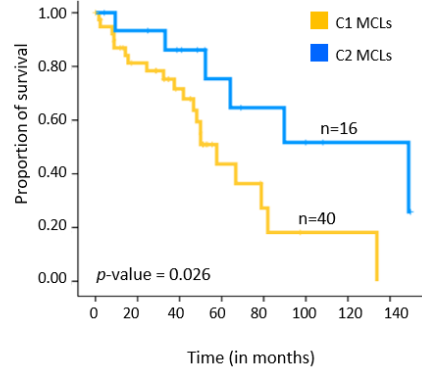
Figure 55. Identification of two MCL subgroups based on DNA methylation profiling. A. Unsupervised PCA of 82 MCLs and 67 normal B-cell subpopulations using the adjusted methylation values of all CpGs analyzed with the 450K array. The two main principal components are shown together in a 2D plot and separately. Vertical and horizontal dotted lines point to the cut-off value separating germinal center-inexperienced and -experienced B cells identified by means of linear discriminant analysis at each component. This cut-off value was used to identify which MCL samples were more similar to germinal center-inexperienced B cells and which were more similar to germinal center-experienced B cells. Normal B cells are surrounded by a dotted grey line. **B.** Comparison of biological and clinical features between the

two epigenetic subgroups (i.e. C1 and C2). The presence of oncogenic mutations is defined as having a mutation in at least one of the following genes: *BIRC3, MEF2B, NOTCH2, TLR2, TP53 and WHSC1*. Data show mean ± SD. ***p< 0.001; n.s., not significant, (Fisher's exact test for qualitative variables or t-test for independent samples in the case of quantitative variables with normal distribution, where a corrected *p*-value was used if the two groups had unequal variances. Statistical significance was defined as *p*<0.05). **C.** Kaplan-Meier plot showing that, as compared with C2, C1 MCLs globally show a worse overall survival. Only C1 and C2 MCL cases with available full clinical reports were used for this analysis (n=56).

Next, in order to define DNA methylation differences between the two clusters, C1 and C2 MCLs were compared, resulting in 13,691 differentially methylated CpGs (**Figure 56A**). Looking closely, most CpGs hypomethylated in C2 MCLs linked C1 cases to germinal center-inexperienced cells while C2 cases to germinal-center experienced B cells (**Figure 56A**) further supporting the concept shown in the second component of the PCA analysis (**Figure 55A**). In contrast, hypomethylation in C1 was predominantly a *de novo* event targeting regions that are highly methylated both in C2 MCLs and normal B cells (**Figure 56A**). Then, the precise location of those regions in the genome was inferred, finding out that were frequently targeting CpG island shores and gene bodies, (**Figure 56B**). Furthermore, the functionality of those regions was characterized by performing ChIP-seq on 6 histone marks and generating chromatin states from sorted NBC and MBC from healthy donors (**Appendix Figure 4**). From this analysis, it could be observed that hypomethylated regions in C1 MCLs were enriched for enhancers and transcribed regions (**Figure 56C**).
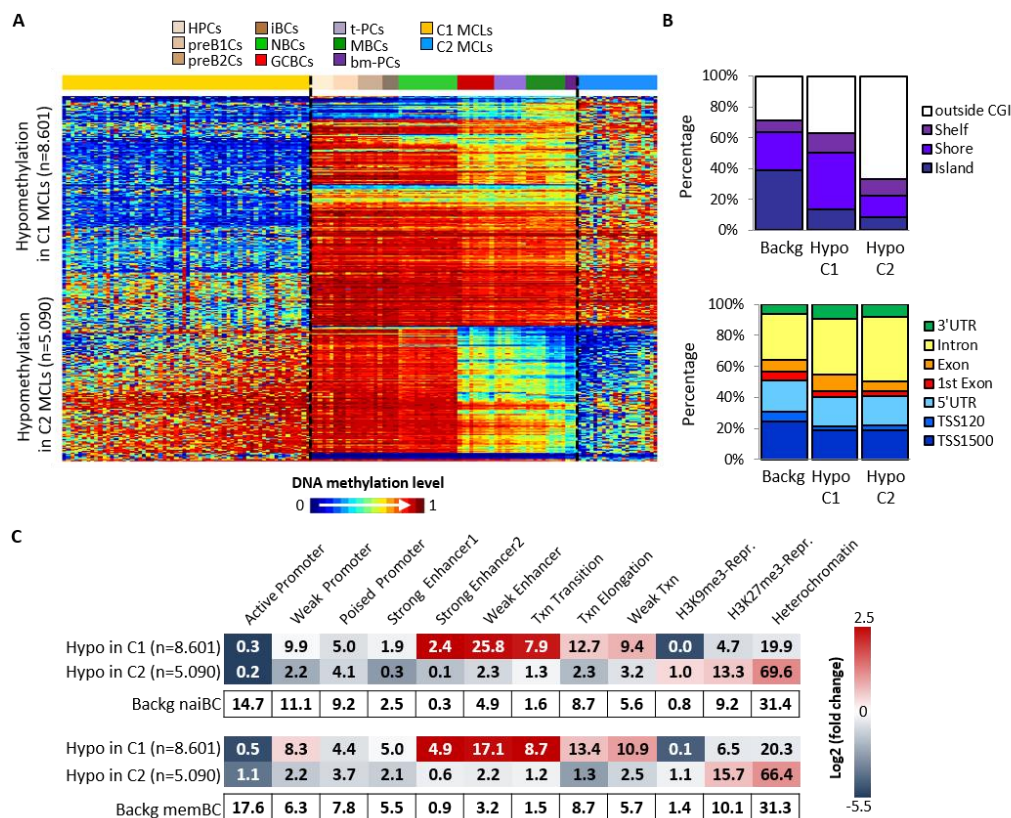
Figure 56. Differential CpGs between MCL subgroups. **A**. Heatmap of the CpGs differentially methylated in C1 compared with C2. **B.** Location of the hypo- and hypermethylated CpGs between C1 and C2 MCLs in the context of CpG islands (CGI) and gene-related regions. **C.** Chromatin states of naive (*upper* panel) and memory (*lower* panel) B cells of the differentially methylated CpGs between C1 and C2 MCLs. The numbers inside each cell point to the percentage of CpGs belonging to a particular chromatin state. The differentially methylated CpGs annotated in (**B**) and (**C**) are the same as those shown in (**A**). HPCs, hematopoietic progenitor cells; preB1Cs, pre-BI cells; preB2Cs, pre-BII cells; iBCs, immature B cells; NBCs, naive B cells from peripheral blood; GCBCs, germinal center B cells; t-PCs, plasma cells from tonsil; MBCs, memory B cells from peripheral blood; bm-PCs, plasma cells from bone marrow; C1 MCLs, germinal center-inexperienced MCLs; C2 MCLs, germinal center-experienced MCLs; Backg, background; Hypo, hypomethylation; Hyper, hypermethylation; TSS, transcriptional start site, UTR, untranslated region.

Lastly, in order to describe the target genes affected for those regions NBC and MBC were used as normal counterparts of C1 and C2 MCLs, respectively. Those two B-cell subpopulations had been previously suggested as potential cells of origin of IGHV

unmutated and mutated MCLs (Navarro et al., 2012). Interestingly, the genes affected by hypomethylation in C1 MCLs were significantly enriched (adjusted *p*<0.05) in several pathways, such as NOTCH signaling (**Figure 57**), which has been previously linked to MCL pathogenesis of the IGHV unmutated/SOX11-positive subgroup (i.e. C1) (Bea et al., 2013; Kridel et al., 2012).
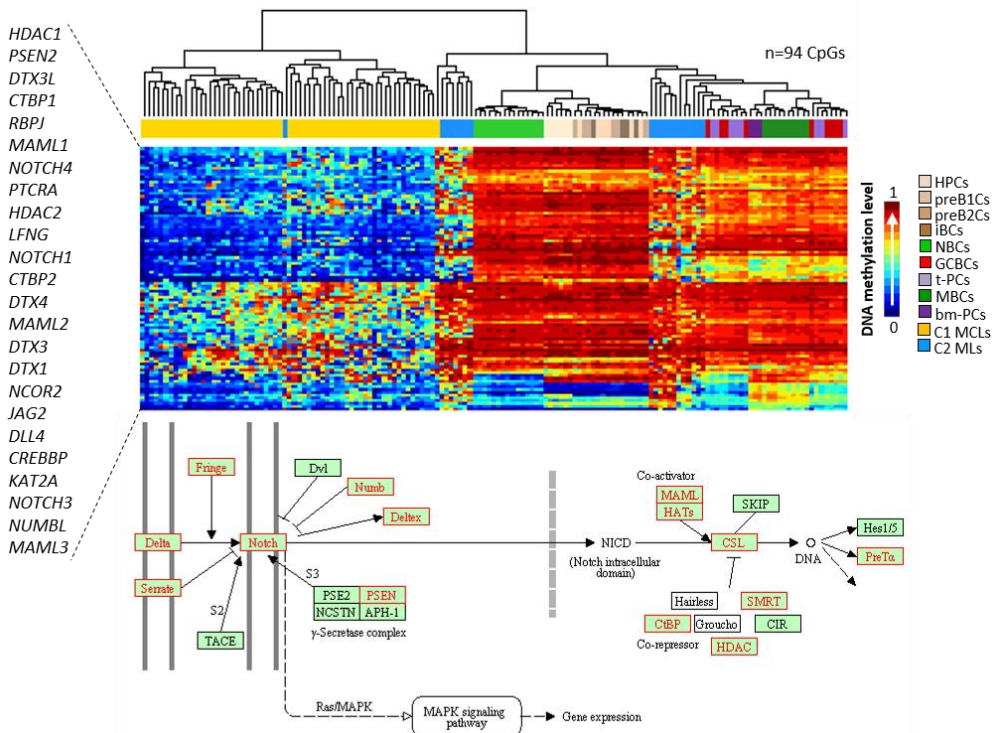
Figure 57. Biological data between the two MCL subgroups. Heatmap representation of the differentially methylated CpGs between MCL C1 and C2 belonging to genes in the NOTCH signaling pathway (*upper* panel) and the exact location of these genes within the NOTCH pathway (*lower* panel). Genes containing differentially methylated CpGs are presented in red in the graphical representation whereas genes with stable DNA methylation are represented in black.

2.4 Comparing MCL groups with their normal cell counterparts reveals a major epigenetic link with normal B-cell differentiation

Next, differences between the MCL subgroups as compared with their respective putative normal counterparts were assessed. In C1 MCLs 60,622 differentially

methylated CpGs (78% hypomethylated) were observed in comparison with NBC, while in C2 MCLs 5,469 CpGs (84% hypomethylated) in comparison with MBC (**Figure 58A**).
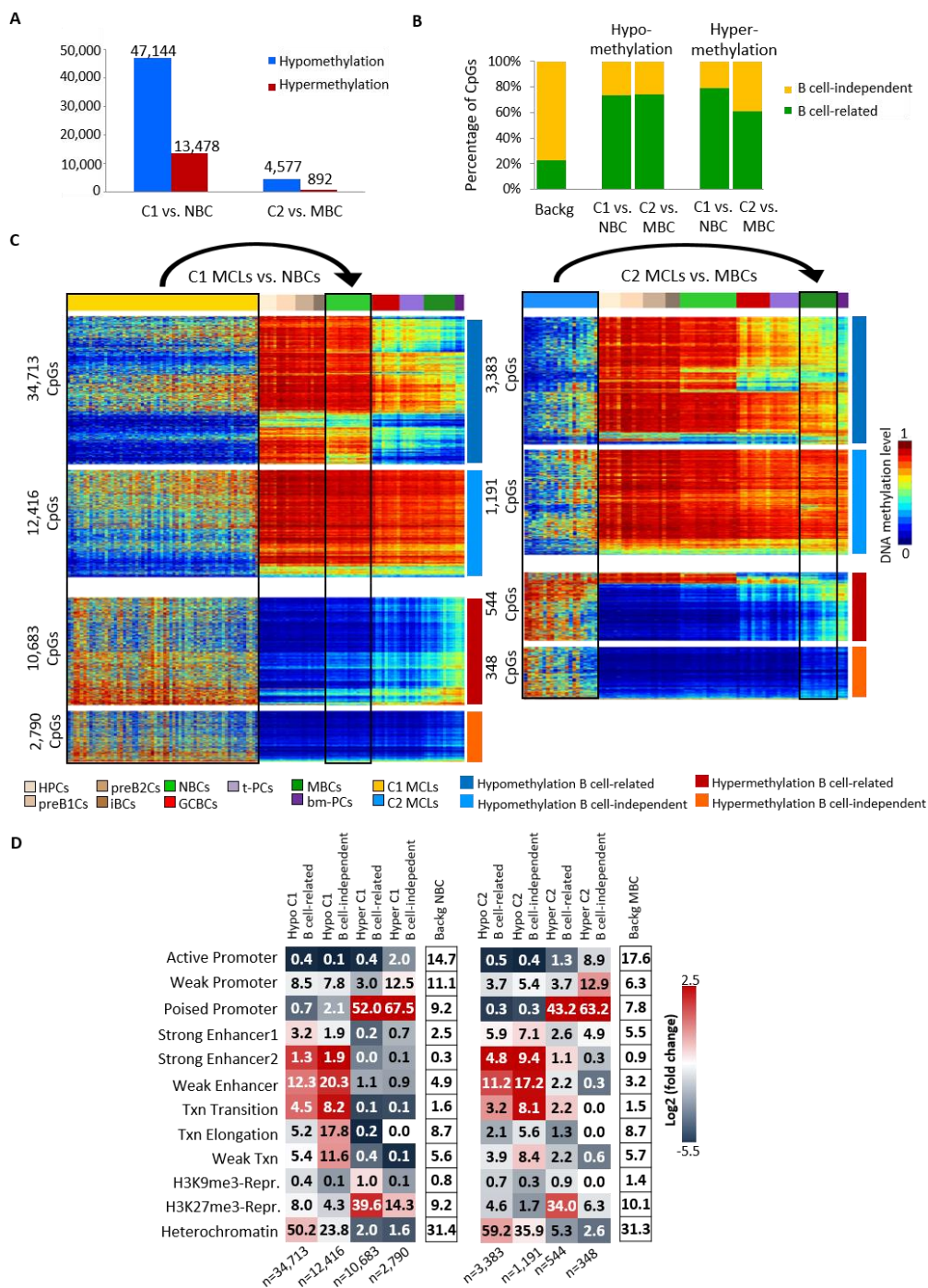
Figure 58. DNA methylation of MCL subgroups versus their respective normal B-cell counterpart. A. Number of differentially methylated CpGs between C1 and NBCs, and between C2 and MBCs. **B.** Percentage of B cell-related and B cell-independent CpGs differentially methylated in each comparison. **C.** Heatmaps of differentially methylated CpGs in C1 MCLs as compared to NBCs (*left*) and in C2 MCLs as compared to MBCs (*right*) in the context of normal B-cell differentiation. **D.** Chromatin states in NBCs and MBCs of the differentially methylated CpGs between C1 and NBCs (*left*), and between C2 and MBCs (*right*), respectively. The numbers inside each cell point to the percentage of CpGs belonging to a particular chromatin state.

Interestingly, 61% to 79% of these CpGs overlapped with those previously described to show variable DNA methylation levels during normal B-cell differentiation (Kulis et al., 2015) (**Figures 58B** and **58C**). This finding suggests that only a fraction of the DNA methylation changes in MCLs as compared with their normal counterparts is unrelated to normal B-cell differentiation and thus, strictly tumor-specific. So, those CpGs dynamically methylated both in MCL and B-cell differentiation were referred as B cell-related CpGs, and those exclusively changing in MCL were named as B cell-independent CpGs. Indeed, these regions were enriched in different chromatin states defined in NBC and MBC (**Figure 58D**). Overall, hypomethylation in MCL in both the B cell-related and independent fractions was enriched for enhancer elements. On the contrary, B cell-related hypermethylated CpGs in MCL were located both in H3K27me3-repressed and poised promoters, whereas those in the B cell-independent fraction were mostly associated with poised promoters (**Figure 58D**).

To identify chromatin state transitions in relationship with DNA methylation changes, ChIP-seq profiles and chromatin states from two MCL cases representative for C1 and C2 were generated, and compared to those from naive and memory B cells, respectively (**Figures 59A** to **59C**). Overall, from these analyses 56% of the regions did not seem to change their chromatin state in MCL upon a DNA methylation alteration. However, repressed regions losing DNA methylation were observed to most likely change towards chromatin states related to activating histone modifications (especially H3K4me3 in poised promoters and H3K4me1 in weak enhancers) (**Figure 59A**) being this phenomenon more prominent in the B cell-independent than in the B cell-related fraction in C1 (52% vs. 18%, *p*<0.001). In the case of hypermethylated regions in C1 MCLs, active and weak promoters in NBCs were detected to turn into poised promoters in MCL and poised promoters turn into H3K27me3-repressed regions (**Figure 59B**).
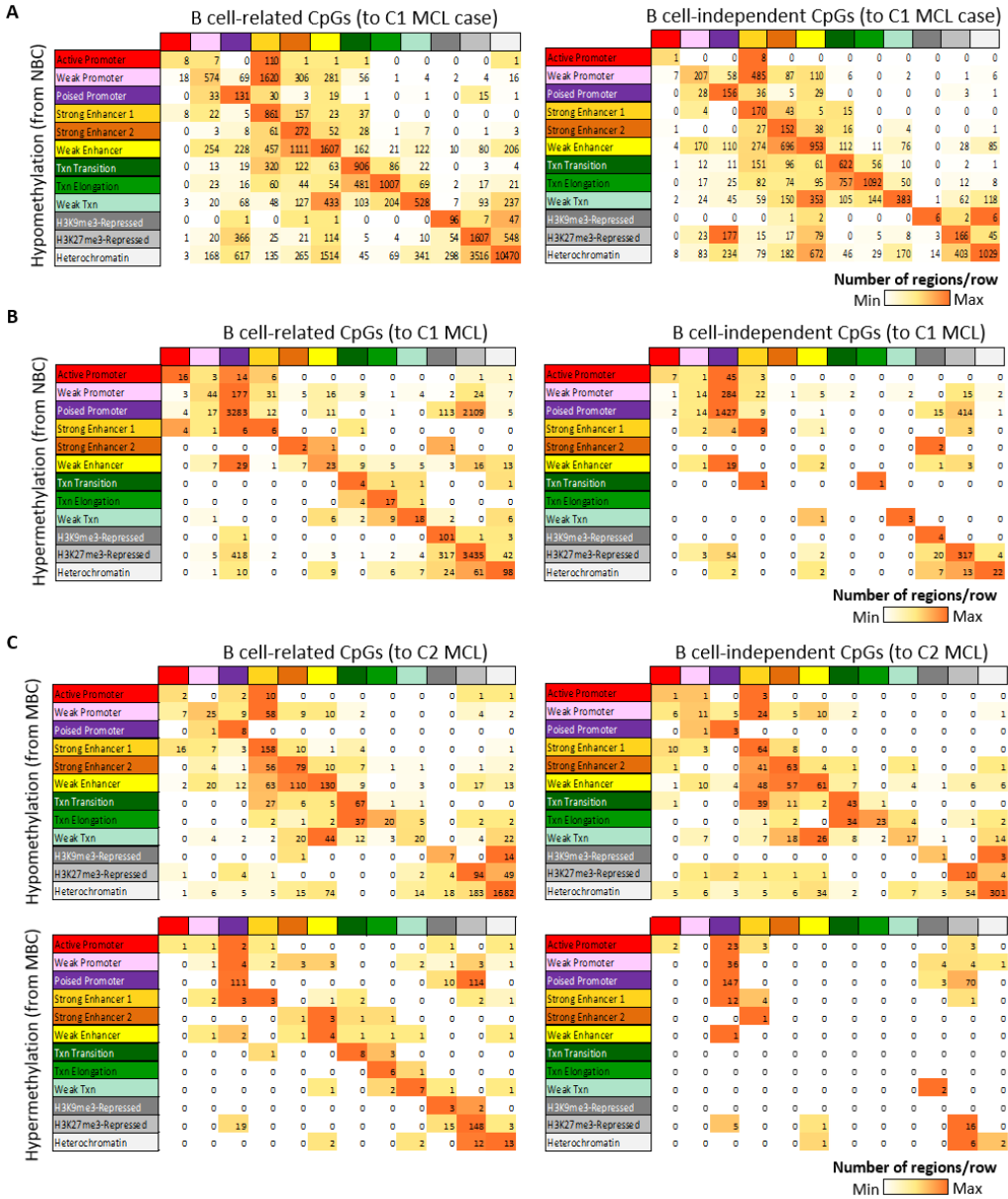
Figure 59. **Transitions among chromatin states in relationship with DNA methylation changes.** **A.** Transition of the chromatin states from NBCs to a C1 MCL case in the B cell-related and B cell-independent hypomethylated CpGs. The numbers inside each cell point to the total number of CpGs in each transition. **B.** Transitions in chromatin states from NBCs to C1 MCL at the B cell-related and B cell-independent hypermethylated CpGs. **C.** Transitions in chromatin states from MBCs to C2 MCL at the B cell-related and B cell-independent hypomethylated and hypermethylated CpGs.

2.5 Individual epigenetic heterogeneity in MCL

The data presented so far suggests that both MCL groups are epigenetically heterogeneous. Based on these observations, a second analytic strategy was applied to tackle individual epigenetic variation of MCL cases in the context of the entire B-cell maturation program. In order to do so, the DNA methylome of each individual MCL case was compared with the hematopoietic progenitor cells (HPCs) (using as cut-off an absolute difference of methylation values of at least 0.25). This seemingly unorthodox approach has the advantage that it uses a fixed reference point for B-cell neoplasms with different normal counterparts, allowing not only to precisely dissect but also to compare the DNA methylation modulation of each individual MCL sample from the moment of B-cell commitment up to and beyond its cell of origin.

Firstly, the total number of changes per case was evaluated and it turned to be highly variable both in C1 and C2 MCLs (ranging from 62,888 to 143,925 CpGs) (**Figure 60A**). Additionally, the DNA methylation levels of the MCLs was observed to correlate less among each other than within normal B cells, showing that the inter-sample heterogeneity is much higher in MCLs than in normal B cells (**Figure 60B**). Furthermore, a total of 318,659 unique CpGs (98% of the 106,552 B cell-related and 53% of the 368,442 B cell-independent CpGs measured by the 450K array) showed a DNA methylation changes as compared to HPCs in at least one MCL case, suggesting that a large fraction of the human methylome can be modulated in normal and neoplastic B cells.
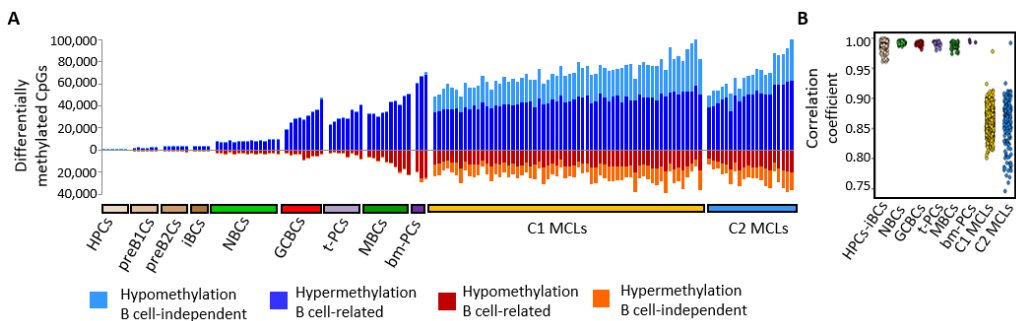
Figure 60. Association between B cell-related and B cell-independent DNA methylation changes in MCL. A. Number of differentially methylated CpGs for each individual normal B-cell subpopulation and MCL compared with HPCs. **B.** Correlation coefficient among samples of the different groups.

Next, to describe regions that may play a role in MCL development, B cell-independent CpGs with recurrent differential methylation in C1 and C2 MCLs were identified. Doing so, it could be observed that the majority of the differentially methylated sites between MCL and HPC were present in one or few MCLs, and that highly recurrent changes were rare events (**Figure 61A**).
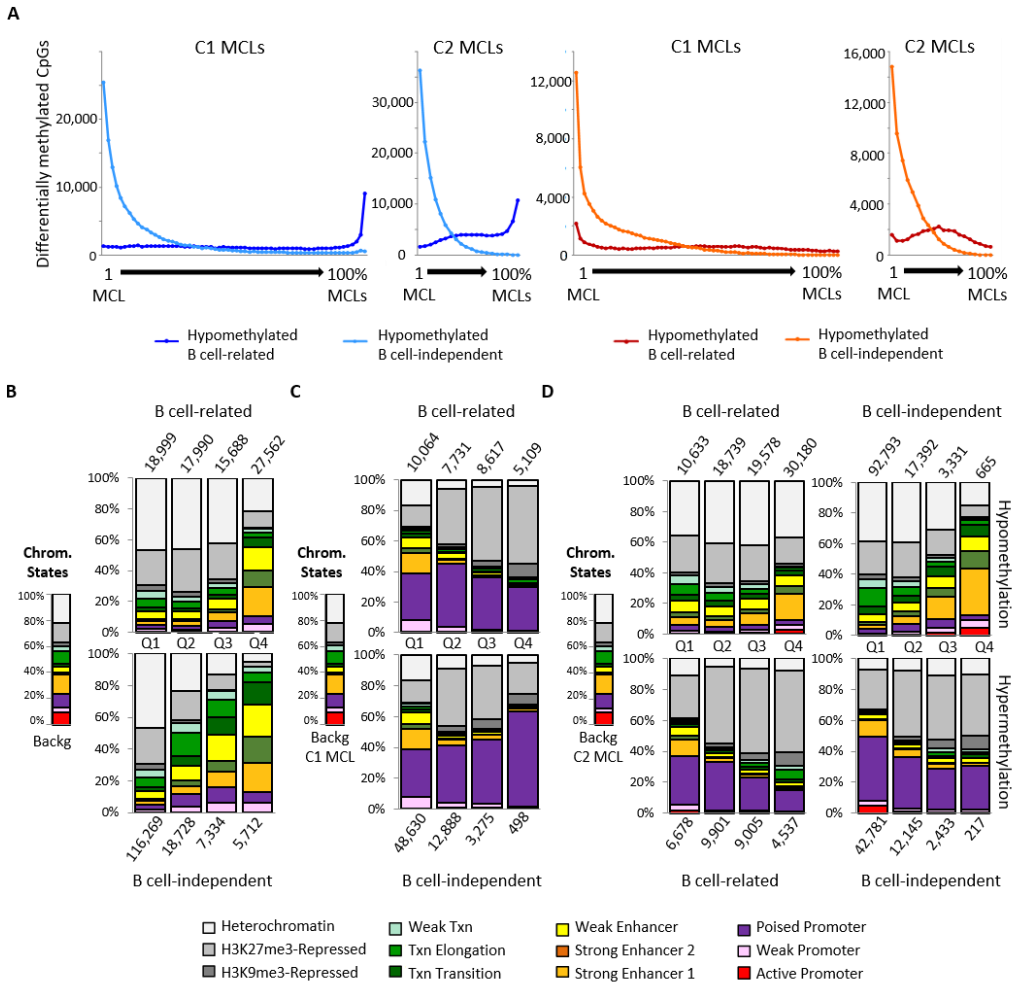
Figure 61. Chromatin states in recurrently altered CpGs. A. Number of B cell-related and B cell-independent differentially methylated CpGs based on their level of recurrence in C1 (1st and 3rd panel) and C2 (2nd and 4th panel). **B.** Chromatin states, defined in a MCL primary case representative of C1 cases, of the hypomethylated CpGs between C1 and naiBCs divided into quartiles based on their level of recurrence. Q1, recurrent in 0-25% of patients; Q2, recurrent in 25-50% of patients; Q3, recurrent in 50-75% of patients; Q4, recurrent in 75-100% of patients.**C.**

Chromatin states, defined in a MCL primary case representative of C1 cases, of the hypermethylated CpGs between C1 MCLs and HPCs divided in quartiles based on their level of recurrence. Q1, recurrent in <25% of the cases; Q2, recurrent in 25-50% of the cases; Q3, recurrent in 50-75% of the cases; Q4, recurrent in >75% of the cases. **D.** Chromatin states, defined in a MCL primary case representative of C2 cases, of the differentially methylated CpGs between C2 and HPCs divided in quartiles based on their level of recurrence.

Furthermore, the relative proportion of differentially methylated regions marked by particular chromatin states (as defined in primary MCL cases), such as heterochromatin and enhancers, was related to the level of recurrence of the DNA methylation changes (**Figures 61B** to **61D**). These findings suggest that most B cell-independent changes in individual MCLs seem to target non-functional regions (i.e. heterochromatin) while commonly altered CpGs, although rare, target regulatory elements (i.e. enhancers).

An additional interesting aspect of this analysis of individual variation was that the number of B cell-related and B cell-independent differentially methylated CpGs per MCL case were linearly related (Pearson r=0.82 and 0.91 for C1 and C2 MCLs, respectively; *p*<0.001) (**Figure 62A**). This association suggests that the mechanisms underlying differentially methylation in B cell-related and B cell-independent CpGs are shared, even though different cases show different degrees of epigenetic changes. Furthermore, in C1 MCLs, 6,245 CpGs were detected with an inverse correlation between their DNA methylation levels and the percentage of IGHV somatic hypermutation (SHM) (**Figures 62B** to **62E**) a phenomenon not observed in C2 cases. The fact that some C1 MCLs concurrently show some degree of SHM and DNA demethylation suggests that C1 MCLs may be derived from germinal center-inexperienced B cells at different maturation stages, ranging from those lacking SHM to those showing low but variable degrees of SHM, which correlate with epigenetic changes (Kolar et al., 2007; Sims et al., 2005).
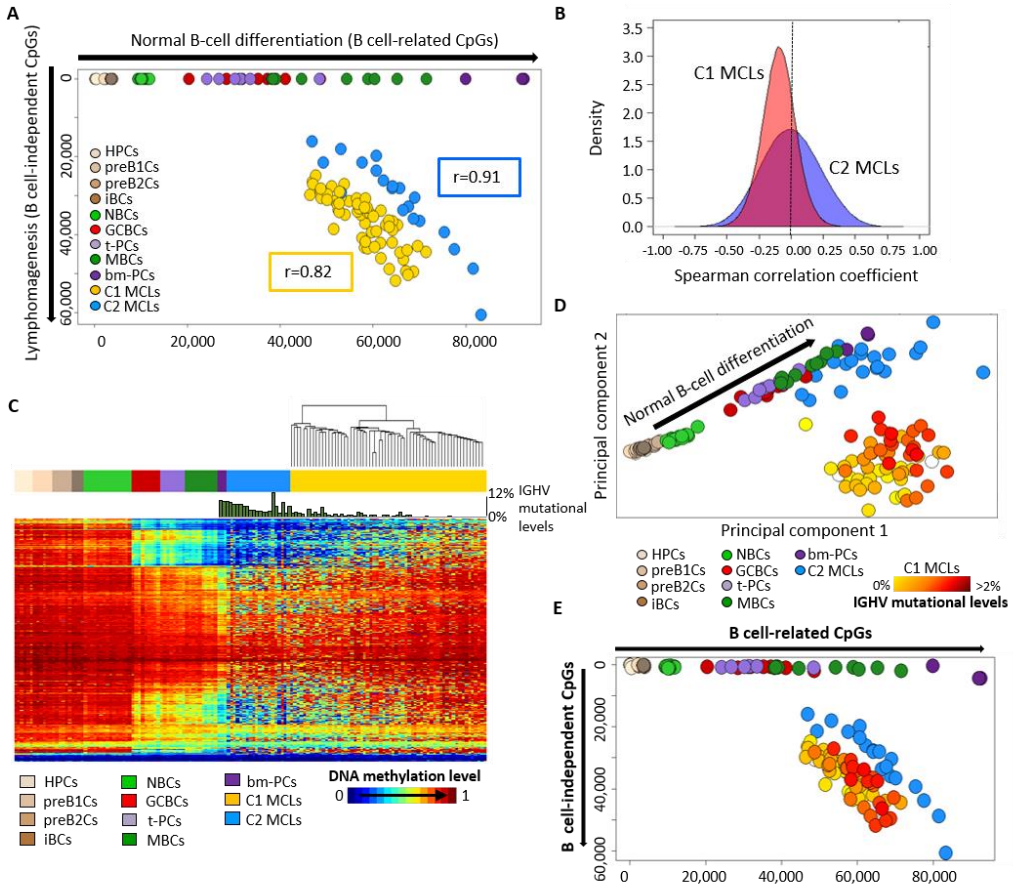
Figure 62. Association between DNA methylation and IGHV mutation levels in recurrently altered CpGs. A. Scatter plot showing the number of B cell-related (x-axis) and B cell-independent (y-axis) CpGs differentially methylated in individual MCLs and normal B cells as compared to HPCs. **B.** Density plot showing the results of the correlation analysis between DNA methylation levels of all 450K CpGs and IGHV mutational levels, for both MCL C1 (red) and C2 (blue). C1 MCLs show a shift towards an inverse correlation whereas C2 MCLs show a distribution centered around 0 (i.e. no correlation). **C.** Heatmap representation of the 6,245 CpGs showing a significant negative correlation (correlation coefficient<-0.4 and FDR<0.05) with IGHV mutational levels in the C1 group. **D-E.** New versions of the unsupervised PCA defined in *Figure 55A* (**D**) and scatter plot defined in *Figure 62A* (**E**) including a color code for C1 MCLs based on their somatic hypermutation level. These graphics indicate that already at the global and unsupervised level, C1 MCLs show different DNA methylation patterns depending of the level of IGHV somatic mutations and that these patterns follow the direction of the normal B-cell differentiation.

2.6 Deep characterization of the MCL methylome by WGBS

To obtain a deeper insight into the MCL methylome, the entire DNA methylome of two highly pure (95% and 99% tumor cells) representative MCLs were sequenced at a single base-pair resolution (48x mean coverage).
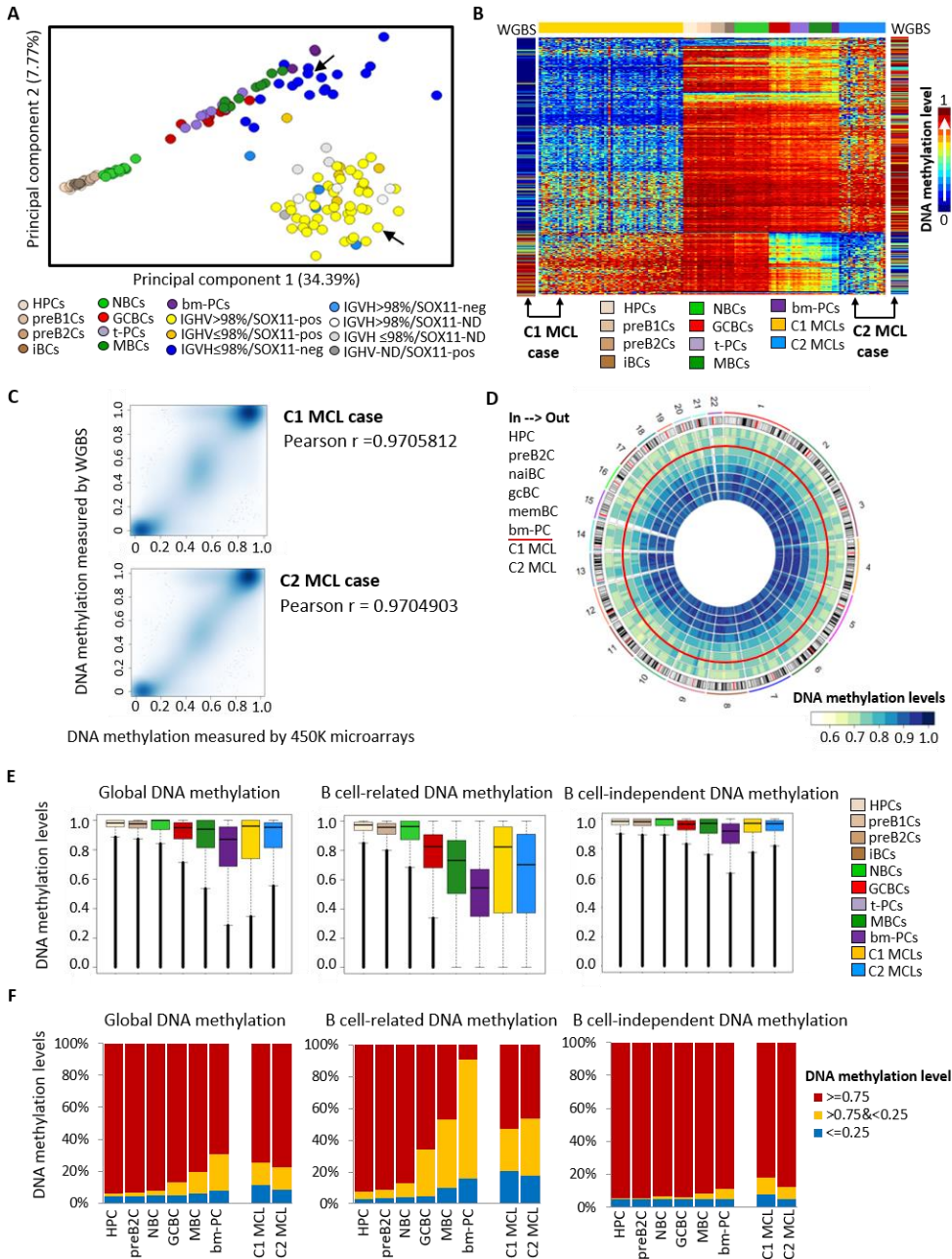
Figure 63. Analysis of the MCL methylome by WGBS. A. Identification of the two representative MCL cases used for WGBS in an unsupervised PCA of microarray methylation data. The two black arrows points to the two MCLs that were further analyzed by WGBS. **B.** Heatmap of the differentially methylated CpGs between C1 and C2 as defined in *Figure 56* that overlap with the CpGs mapped by WGBS in the two representative MCL primary cases. **C.** Scatter plots showing the correlation between the DNA methylation values measure by 450K microarrays and WGBS for the two MCL primary cases. **D.** Circular representation of the DNA methylation levels for HPC, preB2C, NBC, GCBC, MBC and bm-PC, as well as two MCLs representative for C1 and C2, respectively. CpG methylation levels are averaged over 10Mb genomic windows (WGBS, n=11,384,077 CpGs). **E.** Boxplots summarizing the distribution of DNA methylation levels per sample for the 10.4 million CpGs with methylation estimates in all 8 samples. **F.** Global distribution of DNA methylation in each sample showing the percentage of methylated (red), partially methylated (yellow) and unmethylated (blue) CpGs. For both panels (**E**) and (**F**), all CpGs are shown on the *left*, B cell-related CpGs in the *middle* and B cell-independent CpGs on the *right*.

These two WGBS sequenced cases were previously analyzed by 450K microarrays (one from MCL C1 and one from MCL C2; **Figures 63A** to **63C**), and were also analyzed in the context of the DNA methylome of the B-cell lineage (**Figures 63D** to **63F**). The methylation estimates obtained by the two methods were highly comparable (Pearson r=0.97 for both cases, **Figure 63C**).

Then, each MCL was compared to HPCs as fixed reference in order to define differentially methylated CpGs (DMCs) and differentially methylated regions (DMRs). Determining DMRs increased the detection of regulatory regions as compared to detecting DMCs, and therefore the subsequent analyses were performed using the DMR strategy (**Figures 64A** to **64D**). Hence, CpGs were split within DMRs into B cell-related and B cell-independent CpGs, where 55-92% were observed to overlap with those modulated during B-cell differentiation (**Figure 64E**). Intriguingly, most DMRs in MCL either contained only B cell-related or a mixture of both B cell-related and B cell-independent CpGs, and few were exclusively B cell-independent (**Figures 64F** to **64H**). More specifically, in the C1 MCL case, only 9.3% of the hypomethylated and 5.6% of the hypermethylated DMRs were B cell-independent, and these numbers respectively dropped to 1% and 1.2% in the C2 MCL case (**Figure 64G**). This analysis suggests that those regions prone to acquire differential methylation during normal B-cell differentiation seem to be predisposed to be further altered in the context of malignant

transformation, and that regions with pure tumor-specific DMRs seem to be a rare phenomenon.
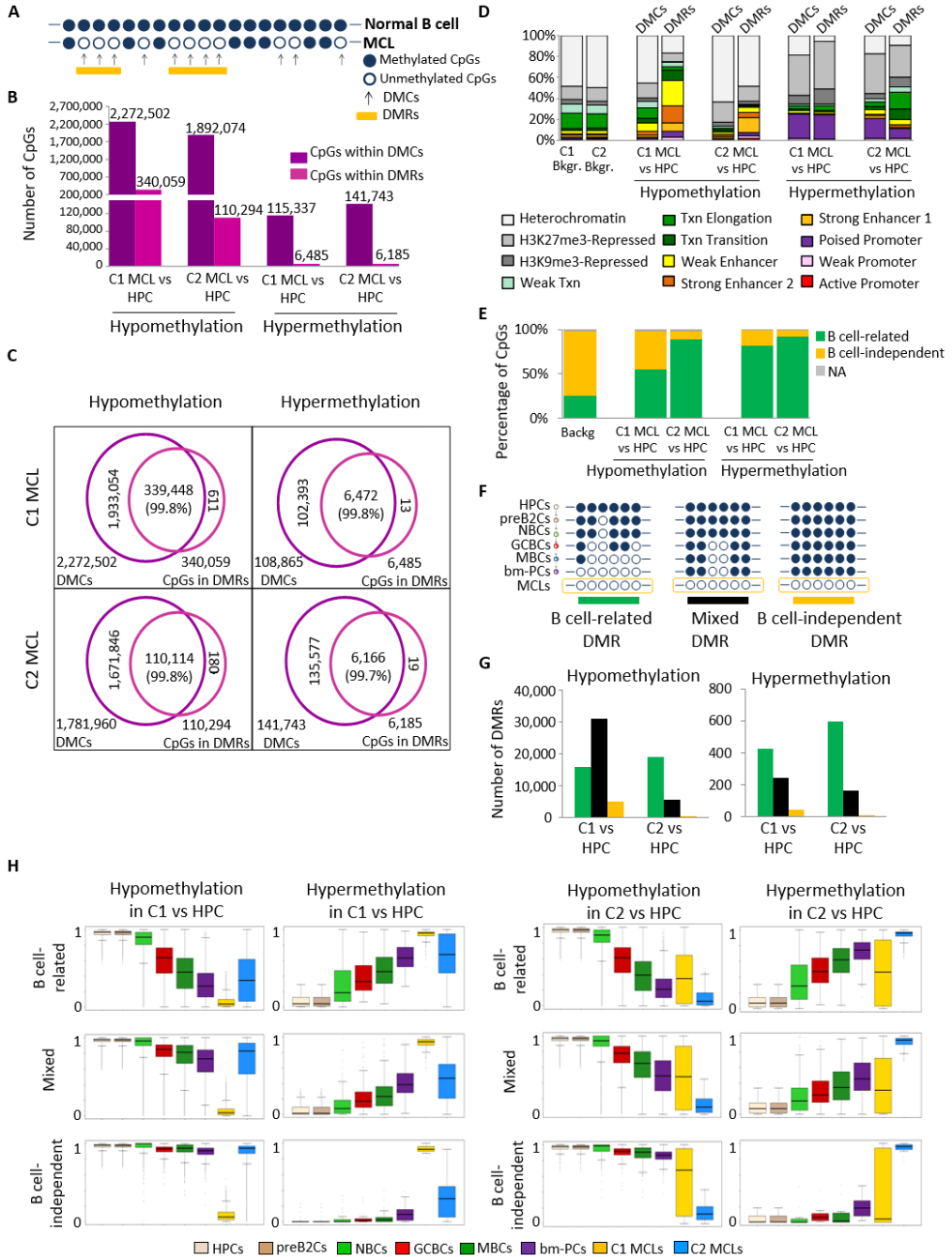
Figure 64. WGBS data. A. Graphical representation of the strategies used to define differentially

methylated CpGs (DMCs) and differentially methylated regions (DMRs). DMCs were defined at a single base resolution whereas DMRs where characterized by a minimum of 100bp and 3 CpGs in a row that showed the same tendency in DNA methylation. **B.** Number of differentially methylated CpGs detected by DMCs and DMRs methods. **C.** Comparison of the chromatin states of the differentially methylated CpGs detected by DMCs and DMRs methods. **D.** Venn diagrams showing the overlap of CpGs detected by each of the strategies. **E.** Percentage of B cell-related and B cell-independent CpGs differentially methylated in C1 MCL and C2 MCL versus HPC. **F.** Graphical representation of the different DMR types: DMRs with only B cell-related CpGs are defined as B cell-related DMRs (*left*), DMRs containing both B cell-related and B cell-independent CpGs are defined as mixed DMRs (*middle*), and DMRs with only B cell-independent CpGs are defined as B cell-independent DMRs (*right*). **G.** Number of B cell-related, mixed and B cell-independent DMRs between C1 versus HPC and between C2 versus HPC. **H.** Distribution of DNA methylation levels for the different DMRs types defined between C1 MCL and HPC and between C2 MCL and HPC. Boxplots show upper and lower quartiles and the median, and whiskers represent minimum and maximum, with outer points indicating outliers.

2.7 Identification of potential epigenetic drivers in MCL and detection of distant *SOX11* enhancers

Next aim was to study whether DMRs between C1 and C2 MCLs can lead to the detection of potential functional regulatory regions that are differentially active in these two groups. By comparing them, 26,603 DMRs hypomethylated in C1 and 4,457 DMRs hypomethylated in C2 were detected. Approximately 60% of these DMRs contained a mixed pattern of B cell-related and B cell-independent CpGs (**Figure 65A**).
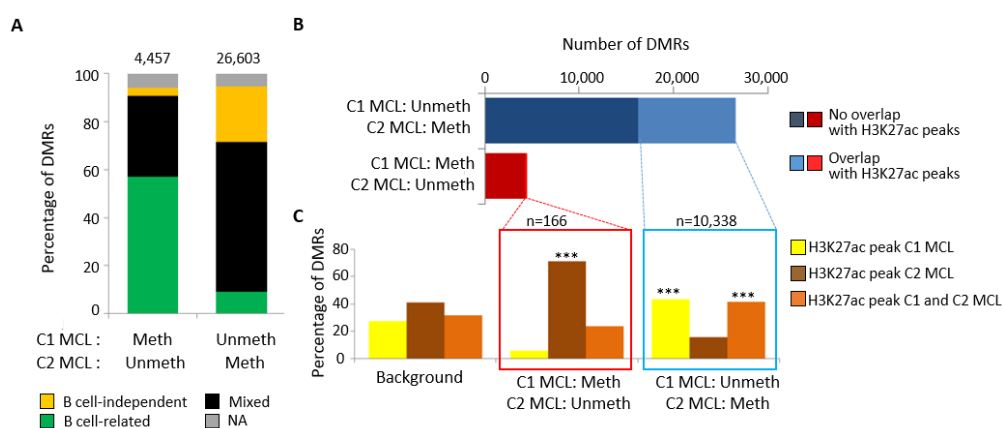
Figure 65. Integrative analysis of differentially methylated regions and histone modifications. A. Distribution of differentially methylated regions (DMRs) defined by WGBS between the MCL cases representative of C1 (SOX11-positive) and C2 (SOX11-negative) into three different DMR

types (B cell-dependent, B cell-independent or mixed DMRs; NA = non-assigned). **B.** Number of DMRs between the C1 and C2 MCL cases and their overlap with H3K27ac peaks in these MCL cases. **C.** Distribution of the DMRs showing an overlap with H3K27ac peaks in the C1 MCL case only, the C2 MCL case only or in both cases. The background represents all H3K27ac peaks in the C1 and C2 MCL case, and shows which percentage is unique for these cases (yellow and darkbrown) and which percentage overlaps (lightbrown). ***p<0.001 (Fisher's test).

Subsequently, ChIP-seq profiles were generated of the same MCL cases studied by WGBS and the detected DMRs were overlapped with the genomic regions simultaneously containing H3K27ac, which marks active regulatory elements (Heintzman et al., 2007). This analyses showed that hypomethylated DMRs in the C1 MCL case had a substantial overlap (39%) with H3K27ac peaks, which were predominantly present either in the MCL C1 case only or in both MCL cases (**Figures 65B** and **65C**).

Then, the chromatin architecture of the DMRs within H3K27ac peaks was characterized in further detail by taking into account H3K4me1, mostly marking enhancers, and H3K4me3, marking promoters (**Figures 66A** to **66C**). The hypomethylated DMRs in the C2 MCL case that are located within H3K27ac peaks (n=118, 2.6%) in the corresponding MCL case but not in the C1 MCL case, normal naive or memory B cells, showed simultaneous presence of H3K4me1 and H3K4me3 (**Figure 66A**) suggesting that these regions represent *de novo* active promoters. The hypomethylated DMRs (n=4,452, 16.7%) in the C1 MCL case within H3K27ac peaks only in the corresponding MCL case and not in MCL C2, normal naive or memory B cells, showed enrichment for H3K4me1 (**Figure 66B**) pointing towards *de novo* activation of enhancers at these regions. Furthermore, the DMRs within H3K27ac peaks appeared to be significantly enriched (p<0.001) in mixed and B cell-independent DMRs (**Figures 66A** to **66C**).

Figure 66. Characterization of DMRs through histone modifications. A-C. Heatmaps showing the read density of H3K27ac, H3K4me1 and H3K4me3 ChIP-seq in the C1 MCL case, C2 MCL case, naive B cells (NBC) and memory B cells (MBC) at selected DMRs (±10Kb). Only the DMRs showing significant differences *versus* the background in *Figure 64C* were used for these heatmaps, i.e. unmethylated regions in the C2 case that overlap with H3K27ac peaks in the C2 case only (**A**), unmethylated regions in the C1 case that overlap with H3K27ac peaks in the C1 case only (**B**), or with H3K27ac peaks in both the C1 and C2 case (**C**). In the lower part of these panels, the percentage of these respective DMRs within the B cell-related, mixed and B cell-independent DMRs is represented. ***$p<0.001$.

Similar results were obtained analyzing the overlap between DMRs and superenhancers (**Figure 67**). Overall, these results show that DMRs between C1 and C2 MCLs may point towards differential active enhancers and promoters in these samples, especially when they contain CpGs that change only in MCL (i.e. mixed/B cell-independent DMRs).

Figure 67. Analysis of superenhancers. A. Number of DMRs between the C1 (SOX11-positive) and C2 (SOX11-negative) MCL cases and their overlap with superenhancers in these MCL cases. **B.** Distribution of the DMRs showing an overlap with superenhancers in the C1 MCL case only, the C2 MCL case only or in both cases. The background represents all superenhancers in the C1 and C2 MCL cases, and shows which percentage is unique for these cases (yellow and

darkbrown) and which percentage overlaps (lightbrown). ***Significantly higher than the background, *p*-value<0.001, as determined by a Fisher's test. **C and D.** Heatmaps showing the read density of H3K27ac, H3K4me1 and H3K4me3 ChIP-seq in the C1 case, the C2 case, naive B cells (NBC) and memory B cells (MBC) at selected DMRs (±25.000bp). Only the DMRs showing significant differences versus the background in (**B**) were used for these heatmaps. These are the unmethylated regions in the C2 case that overlap with superenhancers in the C2 case only in (**C**) and the unmethylated regions in the C1 case that overlap with superenhancers in the C1 case in (**D**). In the lower part of these panels, the percentage of these respective DMRs within the B cell-related, mixed and B cell-independent DMRs is represented, showing a significantly higher (***p<0.001) percentage in the mixed and B cell independent DMRs. n.s, non-significant.

One striking example is a cluster of mixed DMRs hypomethylated in the C1 MCL case overlapping with an enhancer region, marked with H3K4m1 and H3K27ac, located 624-653Kb downstream of *SOX11* only in the SOX11-expressing MCL C1 case (**Figure 68**).
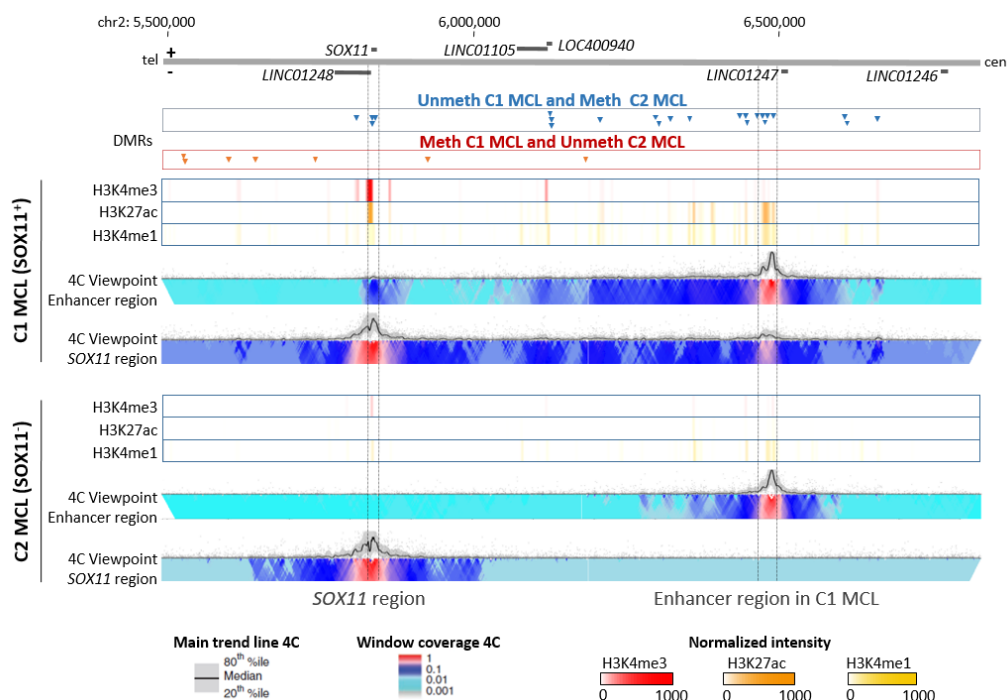
Figure 68. Analysis of the epigenetic and 3D structure of the *SOX11* locus. Differentially methylated regions (DMRs), ChIP-seq levels and 4C-seq signals around the *SOX11* locus. The represented region covers chr2:5,492,778-6,834,378; GRCh37. Unmethylated DMRs in respectively the C1 (SOX11-positive) and C2 (SOX11-negative) MCL cases are represented in the

upper part of the panel by the blue and red arrows. In the lower 2 panels, normalized ChIP-seq intensities for H3K4me3, H3K4me1 and H3K27ac are depicted for the C1 and C2 MCL case. Furthermore, normalized 4C-seq intensities are indicated using the enhancer in MCL C1 (chr2:6,465,559-6,496,708; GRCh37) or the *SOX11* region as viewpoint. tel, telomere; cen, centromere.

To shed light into the interaction and regulation between the described enhancer and *SOX11* gene, I applied a three-dimensional (3D) conformation technique called 4C-seq (Simonis et al., 2007b; van de Werken et al., 2012a) on a set of samples, i.e. two MCL patient samples, two normal B-cell subpopulations, NBC and MBC, and several MCL cell lines. From each representative C1 and C2 primary MCL cases, taking the enhancer region as viewpoint I observed that in the C1 MCL case this region presented high contact frequencies with the *SOX11* gene in three-dimensional (3D) space whereas those were not obtained in the C2 MCL case (**Figure 68**). The reciprocal experiments, i.e. considering the *SOX11* gene as the viewpoint showed a high number of contact frequencies with the enhancer region at the C1 MCL case, validating the previous 4C-seq results (**Figure 68**). Furthermore, I corroborated these results using *SOX11* gene as a viewpoint in three SOX11-positive MCL cell lines, Z-138, GRANTA-519 and JeKo-1, and one SOX11-negative MCL cell line, JVM-2 (**Figures 69A**). For these samples, I validated the result performing the reciprocal experiments, taking the enhancer region as the viewpoint (**Appendix Figure 4A**). Then, I wanted to analyze, as control, the normal B-cell subpopulations considered cells of origin for both MCL subgroups, NBC and MBC. However, high cell numbers of a single B-cell subpopulations were difficult to obtain. Thus, I optimized the 4C-seq technique in order to lower the amount of starting material (**Appendix Figure 4B**). On this way, the experiment could be perfomed in NBC and MBC where no interactions were obtained between the enhancer region and *SOX11* gene, similar to what was observed in the C2 MCL case (**Figures 69A**).

Interestingly, the enhancer region contained around 29Kb of active enhancer activity being much larger than general enhancer regions. Hence, a superenhancer of MCL C1 could be defined taking into account the H3K27ac peaks of the region. Next, to investigate whether the association between DNA hypomethylation of this distant enhancer and the expression of SOX11 is a recurrent phenomenon in MCL primary cases, I analyzed the DNA methylation status of this region by bisulfite pyrosequencing in additional primary SOX11-positive (n=12) and SOX11-negative MCL cases (n=10). In

this way, I confirmed that the identified regulatory region is *de novo* demethylated in SOX11-positive (average methylation level 14-21%) as compared to SOX11-negative cases (average methylation level 63-85%, *p*<0.01) or naive B cells (average methylation level 79-91%) (**Figure 69B** and **Appendix Table 10**). However, whether this demethylation is a cause or a consequence of the enhancer activation and SOX11 expression, remains to be elucidated (further examined in **Study 4**). Altogether, these data suggest a model in which aberrant SOX11 expression in MCL is associated with a *de novo* activation of a distant enhancer element that interacts with the *SOX11* locus in 3D space (**Figure 69C**).
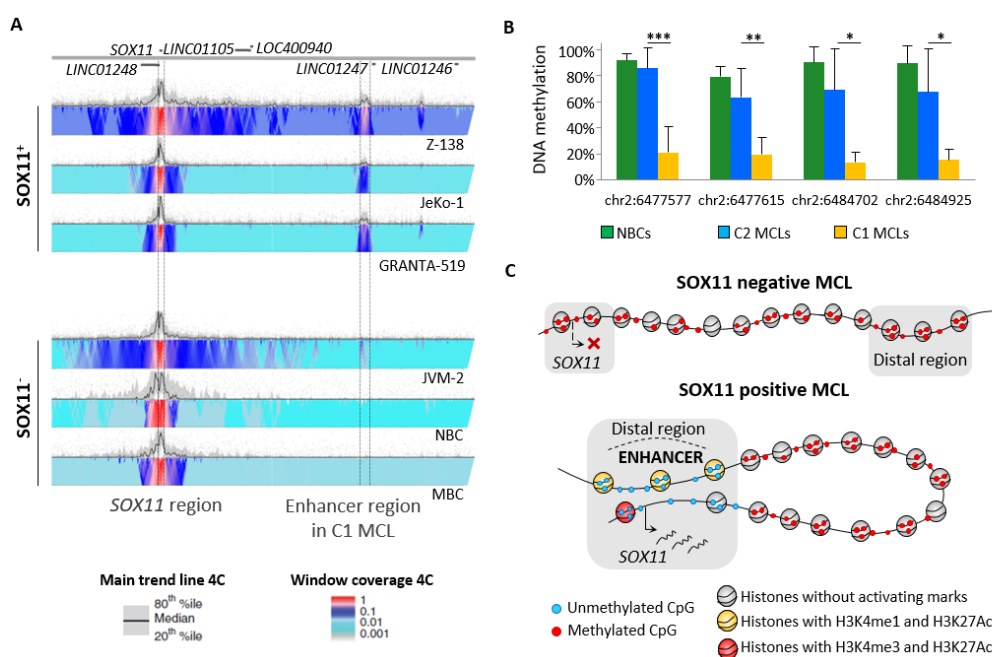
Figure 69. Characterization of the specific SOX11-positive MCL enhancer region. A. Normalized 4C-seq intensities taking the *SOX11* region as viewpoint in 3 SOX11-positive MCL cell lines (Z-138, JeKo-1, GRANTA-519), one SOX11-negative MCL cell line (JVM-2) and in normal naive and memory B cells (NBC and MBC). **B.** Mean methylation levels of 4 CpGs within the SOX11-positive MCL enhancer region in NBC (green, n=4), SOX11-negative (blue, n=10) and SOX11-positive (orange, n=12) MCLs as analyzed by bisulfite pyrosequencing. Data show mean ± SD. **p*<0.01, ***p*<0.001, ****p*<0.0001 (Wilcoxon test for independent samples). **C.** Model of the *SOX11* locus in SOX11-negative MCL (*upper*) and SOX11-positive MCL (*lower*).

2.8 Link among epigenetic burden, genetic changes and clinical outcome of MCL patients

The analyses and results obtained so far suggested that in addition to the significant survival difference between C1 and C2 MCLs, the epigenetic burden (i.e. number of differentially methylated sites regardless of their relationship to normal B cells) might be as well associated with clinical behavior (**Figure 55C**). Indeed, in both MCL subgroups, the numbers of DNA methylation changes as compared to HPCs were found to show a significant linear association with the clinical outcome, approximately doubling the risk of death with each 10,000 methylation changes (**Figures 70A** and **70B**).
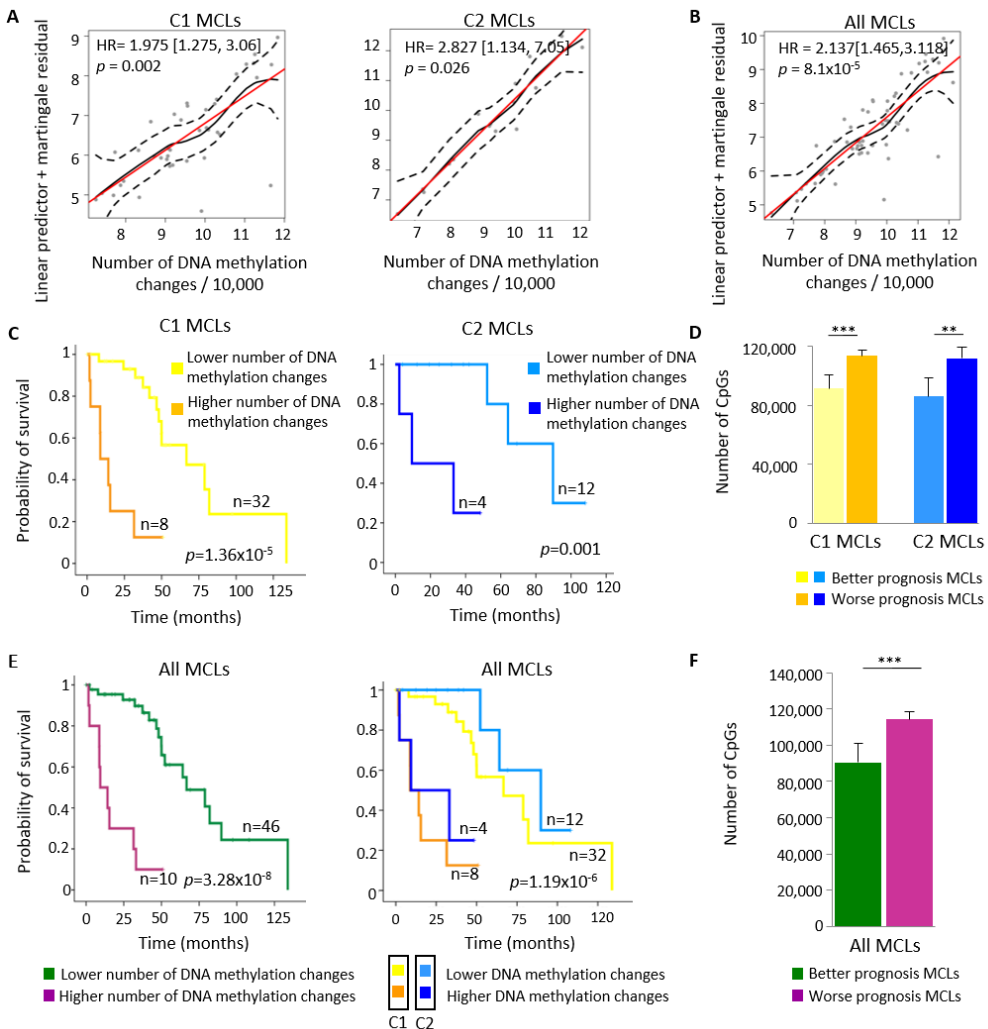
Figure 70. Link between the number of DNA methylation changes and prognosis. A and **B**. Relationship between the number of epigenetic changes and overall survival through a linear predictor in C1 and C2 MCLs (**A**) and all MCLs (**B**). Red line, perfect linear relationship; black line, observed regression line; dash line, 95% confidence interval of observed regression. The local regression line suggests a linear predictor between the number of changes and the log hazard. **C** and **D.** Kaplan-Meier plots of MCLs with lower vs. higher number of differentially methylated CpGs compared to HPCs in C1 and C2 MCLs (**C**) and all MCLs (**D**). **E**. Merged Kaplan-Meier plots with lower vs. higher number of differentially methylated CpGs compared to HPCs in C1 MCL and C2 MCL from panel (**C**). Yellow and orange lines refer to C1 MCL cases whereas light and dark blue lines refer to C2 MCL cases. **F** and **G.** Number of differentially methylated CpGs between the subgroups with different prognosis defined in panel *C* (for **F**) and *D* (for **G**). Data show mean ± SD. ***p<0.001 (t-test for independent samples).

Beyond this quantitative association, the threshold of DNA methylation changes that maximizes the difference in clinical outcome between two subsets of patients was calculated (**Figures 70C** to **70G**). Furthermore, DNA methylation changes were compared with the presence of mutations using a set of six recurrent driver genes in MCL (Bea et al., 2013). This analysis showed that cases with gene mutations in C2 MCLs, but not C1 MCLs, displayed a significantly higher number of CpG methylation changes (**Figures 71A** and **71B**). To determine whether these observations can be linked to cell proliferation, the proliferation signature in 25 of our MCL cases was calculated (Navarro et al., 2012). As expected, MCL C1 cases are in general more proliferative than C2 cases (**Figure 71C**) but the proliferation signature was positively correlated with the number of epigenetic changes only in C2 MCLs (**Figure 71D**).
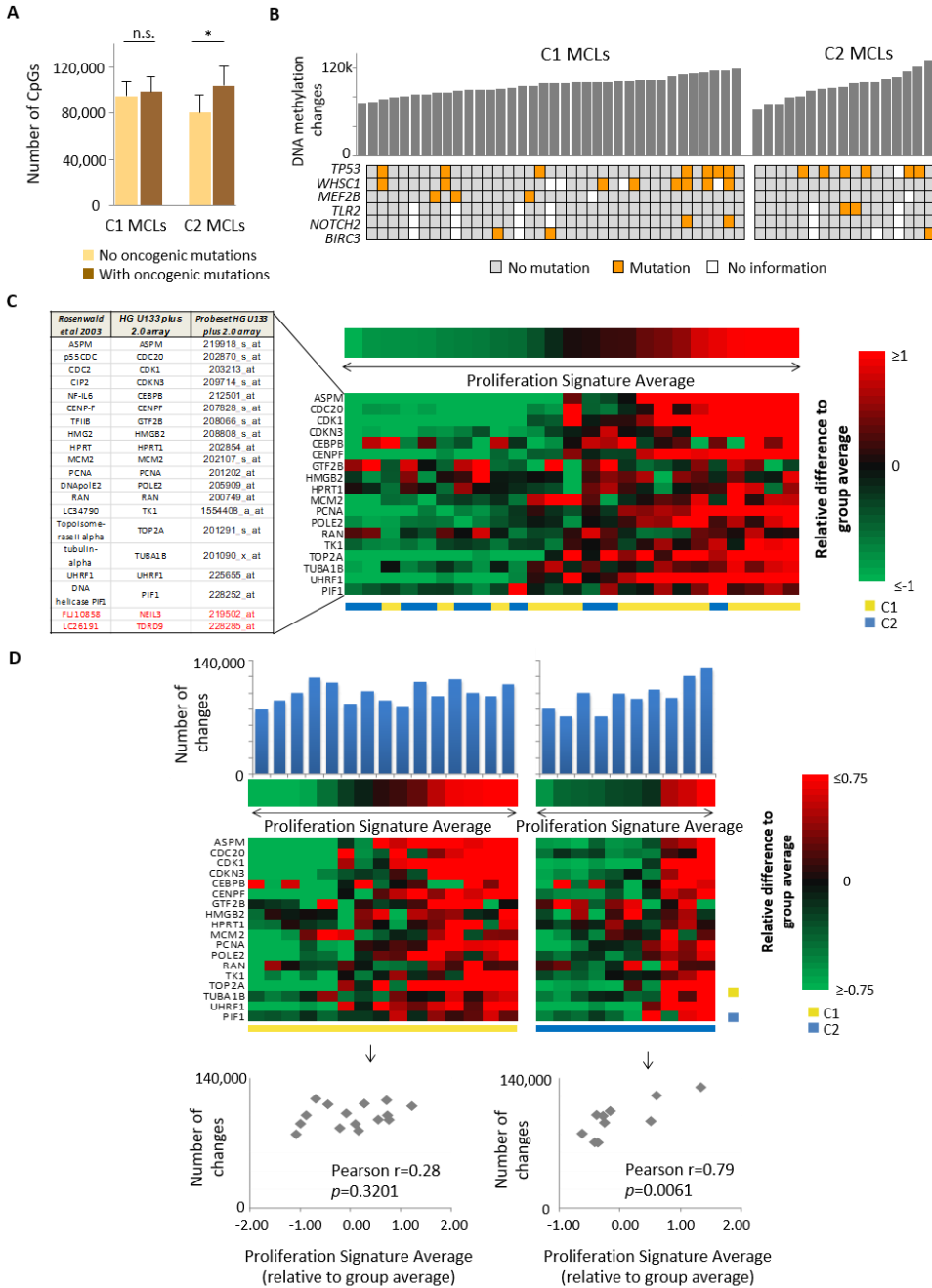
Figure 71. Prognosis and proliferation signature. A. Association between the number of differentially methylated CpGs and the presence of oncogenic mutations (in *BIRC3*, *MEF2B*, *NOTCH2*, *TLR2*, *TP53* and *WHSC1* genes) for both C1 and C2 MCLs. Data show mean ± SD. *p<0.05; n.s. not significant (t-test for independent samples). For cases without or with

mutations, the sample sizes are, respectively: C1 (n=16 and n=15) and C2 (n=6 and n=8). **B.** Representation of epigenetic changes and the presence of oncogenic mutations in both C1 and C2 subgroups. **C.** Proliferation signature average of 25 MCL cases (columns) using the genes and probesets indicated in the last 2 columns of the left panel. *NEIL3* and *TDRD9* were excluded as they did not show expression in the MCL cases. **D.** Proliferation signature of 15 (C1) and 10 (C2) MCL cases (heatmaps, using the same probesets as in panel *C*), the number of DNA methylation changes (*upper* panels, bar graphs) and the correlation between the proliferation signatures and the number of changes (*lower* 2 panels) in C1 (*left*) and C2 (*right*). *P*-values were calculated using the student's t-distribution.

Finally, a multivariate Cox regression model with 6 variables was performed and related to MCL prognosis. Particularly, the variables used as input were: age (quantitative), morphology (classical or blastoid), IGHV somatic hypermutation (quantitative), presence of mutations (positive or negative, based on the analysis of the 6 most frequent drivers: *TP53, WHSC1, MEF2B, TLR2, NOTCH2* and *BIRC3*), SOX11 expression (positive or negative) and number of DNA methylation changes (quantitative). This analysis revealed that the number of DNA methylation changes was the strongest independent prognostic factor in the MCL series of this study (*p*-value=1.4×10^{-5}) followed by IGHV identity levels (*p*-value=0.0015) and age (*p*-value=0.0019) (**Table 6**). Altogether, these data suggest that patients with more epigenetic changes have a worse clinical outcome and that in C2 MCLs, this correlates with the acquisition of genetic changes and increased cell proliferation.

Variables	HR	95% CI	*p*-value
DNA methylation changes/10,000	4.14	2.18-7.86	1.4×10^{-5}
IGHV identity	1.4	1.14-1.73	0.00147
Age	1.11	1.04-1.19	0.00192

Table 6. Results of the multivariate Cox regression model.

An (epi)genetic model of MCL pathogenesis could be built based on the results obtained throughout this study (**Figure 72**), in which C1 MCL cases derive from a range of germinal-center inexperienced B cells that carry the t(11;14) translocation and show absence or low levels of IGHV somatic hypermutation (Navarro et al., 2012). Early

during transformation, these cells acquire genetic and epigenetic changes and show expression of SOX11, which prevents these cells from entering the germinal center (Palomero et al., 2016). C2 MCLs also carry the t(11;14) translocation but, in contrast to cases from C1, they lack SOX11 expression and show IGHV somatic hypermutation. This fits with the hypothesis that they are derived from germinal-center experienced B cells, most likely MBC (Navarro et al., 2012). C2 MCLs with an indolent clinical course lack oncogenic mutations and acquire few epigenetic changes, whereas C2 MCLs with a more aggressive clinical behavior acquire mutations and present extensive DNA methylation changes.
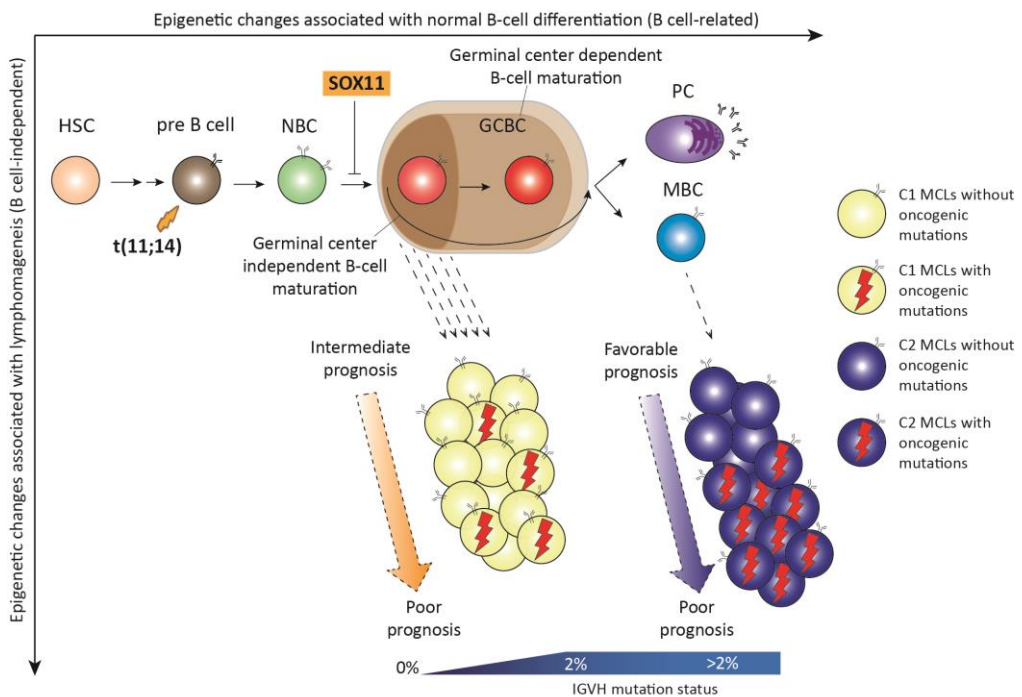
Figure 72. Proposed epi(genetic) model of MCL pathogenesis. Overview of the results obtained decoding the epigenome of MCL in the light of the entire B-cell differentiation.

3. Conclusions

- A method to deconvolute the cellular composition of the sample and in silico purify the DNA methylation estimates of the MCL samples was developed.

- Overall, all MCLs have a DNA methylation profile more similar to antigen-experienced cells. However, MCLs with unmutated and mutated IGHV retain DNA methylation signatures of germinal center-independent (naive) and germinal center-experienced (memory) B cells, respectively, supporting the concept of their differential cellular origin.

- The number of DNA methylation changes in individual MCL cases was highly heterogeneous, and mostly affecting heterochromatic regions without an apparent functional impact.

- An integrative approach combining the DNA methylome and histone modification patterns in primary MCL cases allowed us to identify DMRs with potential functional impact.

- The biomarker SOX11 is connected by a three-dimensional loop structure with a cluster of hypomehtylated DMRs located 650Kb downstream of *SOX11* carrying enhancer marks (H3K4m1 and H3K27ac), which regulates the gene in MCL.

- From the clinical perspective, the magnitude of DNA methylation changes is the most relevant independent prognostic factor in our MCL series.

STUDY 4

Insights into the mechanisms underlying the aberrant SOX11 expression in MCL

1. Introduction

SOX11 was identified as one of the most representative genes differentially expressed in distinct clinico-biological forms of MCL, being this oncogene highly expressed in cases with aggressive behavior named as conventional MCL (cMCL), but not or very lowly expressed in non-nodal MCL (nnMCL) (Beekman et al., 2018b; Swerdlow et al., 2017).In the **Study 3** of this doctoral thesis, I present an integrative analysis of the epigenome in primary MCLs, which hinted the regulatory region involved in SOX11aberrant expression in SOX11-positive MCL. This distant superenhancer region was identified to be specific for SOX11-positive MCL cases, as it was not active in other cells expressing *SOX11* (human embryonic stem cells, skeletal muscle myoblasts and human lung fibroblasts) studied by the ENCODE consortium (Beekman et al., 2018b; Dunham et al., 2012; Ernst et al., 2011).

The specific goal of this study was to shed light into the potential mechanisms promoting the establishment of a *de novo* three-dimensional conformation at the *SOX11* genomic region in SOX11-positive MCLs.

2. Results

2.1 Analysis of the 3D loop in SOX11-positive MCL under the microscope

Fluorescence *in situ* hybridization (FISH) was applied on cells from SOX11-negative and positive MCL cases to corroborate the previously identified 3D loop between the *SOX11* locus and the distant superenhancer in SOX11-positive MCLs (examined in **Study 3**). Two differentially labeled DNA probes were established, one spanning the SOX11 oncogene and one spanning the superenhancer. If the 3D loop is present, I expect the two signals are expected to be fused whereas if there is noin the absence of the loop, I expect a significant separation between the differentially-labeled probes would be observed. Indeed, the results showed a significant increase in the distance between the *SOX11* locus and the superenhancer region in SOX11-negative MCLs as compared to the SOX11-positive MCLs (SOX11-positive, mean distance=0.013, SD=0.019; SOX11-negative, mean distance=0.088, SD=0.054; *p*-value=0.002,) (**Figure 73A** and **73B**). In addition to validate the 3D interaction detected by 4C-seq, FISH allows to assess single cells and therefore I could examine each allele individually observing that on most of the SOX11-positive MCLs cases the loop formation between *SOX11* locus and

superenhancer was occurring in a biallelic fashion. This finding makes unlikely that a particular non-coding genetic alteration accounts for the activation and looping of this region, as this should be in most cases monoallelic.
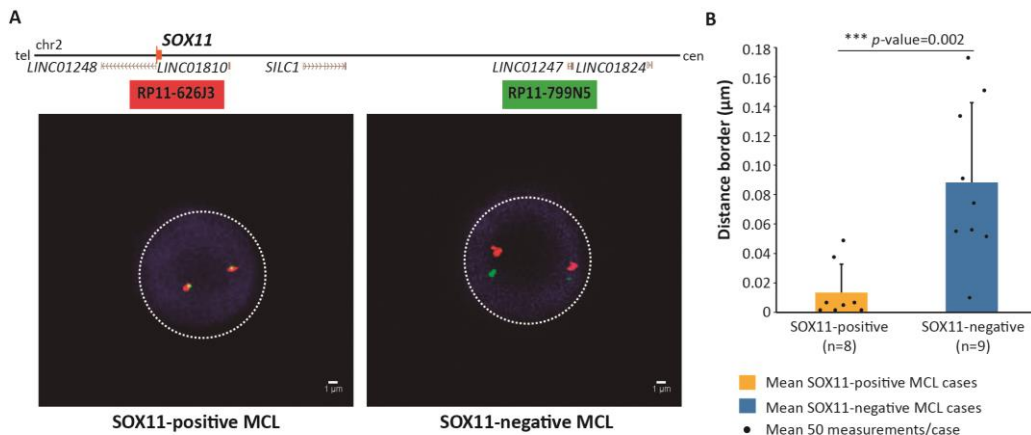
Figure 73. FISH validation results of SOX11-enhancer loop. A. Scheme of the 3D FISH experiment *(top)*. Two BACs were used one for the *SOX11* locus (red) and another for the SOX11-positive MCL superenhancer (green). Two experimental 3D FISH images *(bottom)* of a SOX11-positive *(left)* and a SOX11-negative *(right)* MCL case. Nucleai were stained with DAPI. **B.** Mean distances between *SOX11* locus and the superenhancer region were analyzed for 50 different signals per each case. SOX11-positive group was composed by: n=5 MCL primary cases and n=3 MCL cell lines (JeKo-1, GRANTA-519, Z-138) while the SOX11-negative group was formed by: n=4 MCL cases, n=1 MCL cell lines (JVM-2), n=1 B-lymphocyte cell line (GM12878) and n=2 peripheral blood form normal donors.

2.2 Inferring *SOX11* topologically associating domaing

I next wonder if the looping structure in SOX11-positive MCLs was occurring due to changes of higher order three-dimensional structures such as the 3D building blocks or so-called topologically associated domains (TADs). These structures are considered as regulatory units, in which alterations have been described in many disease and cancers (Flavahan et al., 2015; Franke et al., 2016; Giorgio et al., 2015; Lupiáñez et al., 2015; Spielmann et al., 2012), such as prostate cancer (Taberlay et al., 2016) as well as T-cell acute lymphoblastic leukemia (T-ALL) (Hnisz et al., 2016). To shed light into this phenomenon in MCL, TAD structures at 20Kb resolution were defined in five MCL cases, 2 SOX11-positive and 3 SOX11-negative, with available *in situ* Hi-C data (presented in **Study 1**). The features of chromosome 2 TADs were further assessed to control that all

samples had similar number of TADs, similar density of interactions and TAD length discarding technical issues (**Appendix Figure 5**). Once TADs were defined, I hypothesized that a loss of TAD border in SOX11-positive MCL cases might result in the aberrant 3D loop between the defined enhancer and the *SOX11* locus. However, the TAD structures encompassing the *SOX11* locus and the studied superenhancer were similar in expressing and non-expressing SOX11 MCL cases, **Figure 74.** This result indicated that the *de novo* 3D reorganization was not caused by disruptions of the structural blocks. Then, to analyze whether there were intra-TAD changes in MCL expressing SOX11 cases, I accurately inspected the TAD interactions containing the *SOX11* locus and its SOX11-positive MCL superenhancer. An increase of interactions could be observed in the specific TAD containing *SOX11* at the SOX11-positive MCL cases as compared to the rest of the chromosome (mean density of interactions of the TAD=1.25; SD=0.004 at 20Kb resolution) and as compared to the SOX11-negative cells (mean density of interactions of the TAD=0.96; SD=0.091 at 20Kb resolution). As a result of these interactions, a compartment shift could be defined on the domain containing *SOX11*. Precisely, more active compartments were distinguished in SOX11-positive cMCL cases than in SOX11-negative nnMCL cases (**Figure 74**). Overall, these results ruled out the presence of a large-scale structural disruption as the cause of *SOX11* deregulation and points out that the chromatin loop reconfiguration induces an increase of interactions at a TAD level on the *SOX11* expressing MCLs.
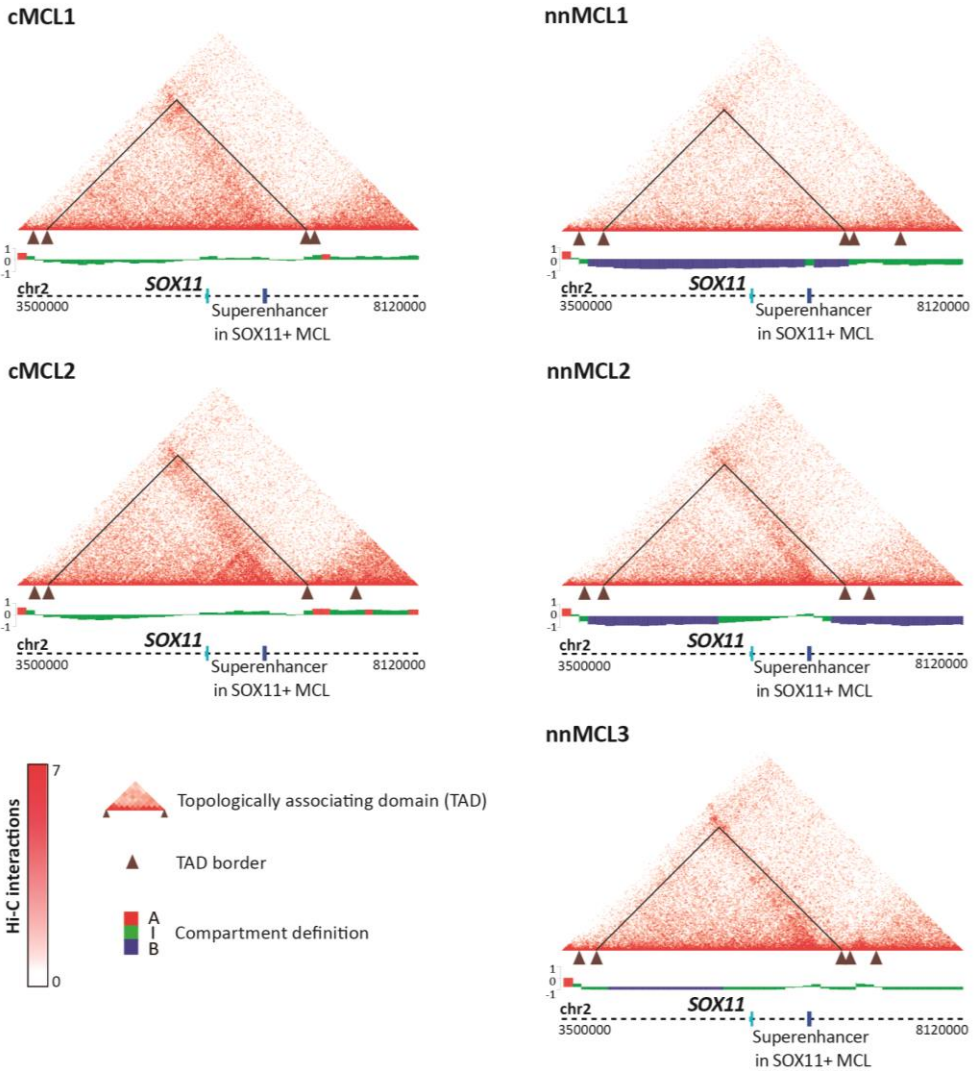
Figure 74. Chromosome conformation in MCLs within *SOX11* locus. Hi-C contact matrices for two conventional MCL cases, SOX11-positive (cMCL1 and cMCL2, *left*) and three leukemic non-nodal MCL cases, SOX11-negative (nnMCL1, nnMCL2 and nnMCL3, *right*). Color code indicated the Hi-C interactions. TAD contacting *SOX11* locus and the SOX11-positive MCL superenhancer is marked with a black line. Below, TAD borders and compartment definition were indicated per sample, as well as, the location of the *SOX11* locus and the superenhancer in SOX11-positive MCL. The coordinates of the represented region are chr2: 3,500,000-8,120,000; GRCh38.

Then, the chromosomal conformation was assessed for MCL cell lines. *In situ* Hi-C was performed for the SOX11-positive MCL cell line, Z-138, and the SOX11-negative MCL cell line, JVM-2 (**Figure 75A** and **Appendix Table 3**). Similar to the primary cases, the domain structure containing the *SOX11* locus and the superenhancer was also characterized and maintained in both cell lines. Moreover, an increase on the density of interactions were identified at the SOX11 expressing MCL cell line, Z-138 (density of interactions of the TAD=1.13 at 50Kb resolution) as compared to SOX11 negative MCL cell line, JVM-2 (density of interactions of the TAD=0.87 at 50Kb resolution), being the compartment definition more inactive in JVM-2. These results replicated those observed in primary cases, demonstrating that 3D conformation in these *in vitro* MCL models is representative of both MCL subgroups. Therefore, further experiments evaluating the chromosome structure using cell lines could be then performed.
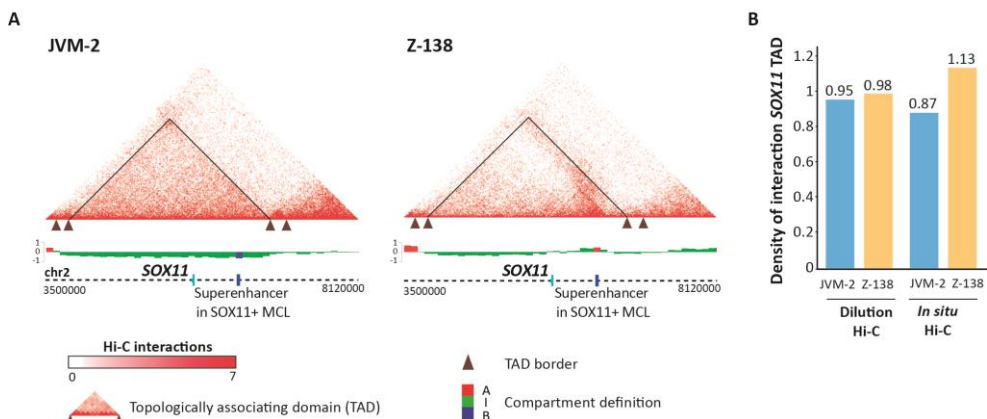
Figure 75. Cell line model chromosome conformation structure within *SOX11* locus. A. Hi-C contact matrices for the SOX11-positive MCL cell line, Z-138, and the SOX11-negative MCL cell line, JVM-2. TAD contacting *SOX11* locus and the SOX11-postive MCL superenhancer is Marked with black line. Color code indicated the Hi-C interactions. Below, TAD borders and compartment definition are indicated per sample, as well asthe location of the *SOX11* locus and the superenhancer in SOX11-positive MCL. The coordinates of the represented region are chr2: 3,500,000-8,120,000; GRCh38. **B.** Density of interactions on the TAD comprising *SOX11* locus with the SOX11-positive MCL superenhancer, determined at 50Kb resolution for JVM-2 and Z-138, using the two genome-wide chromosome conformation capture techniques, Hi-C and *in situ* Hi-C.

Moreover, this TAD example allowed me to assess the differences between the *in situ* Hi-C protocol performed with 2 million cells of starting material and the original Hi-C protocol, performed combining two different libraries using 25 million cells per library. I determined the TAD structures in both cell lines for both protocols at a 50Kb resolution (instead of 20 Kb) to avoid technical biases between the two methods. I observed that, eventhough the TAD structure boundaries were pretty similar in both protocols, the density of interactions at the SOX11 domain was lower with the Hi-C protocol. Using this method, the differences were not so clear between the SOX11 positive and negative MCL cell lines as with *in situ* Hi-C (**Figure 75B**). This example demonstrated that, apart from requiring lower amount of cells, *in situ* Hi-C also reduces the frequency of spurious contacts observed in standard or dilution Hi-C, as has been already described (Rao et al., 2014), which in turn leads to increased number of specific interactions, as I observed in the case of the SOX11 locus.

2.3 Dynamic interactions on the SOX11 locus upon normal B-cell differentiation

Taking advantage of the genome-wide chromosome conformation capture data I performed in cell subpopulations spanning normal B-cell differentiation (presented on **Study 1**), I also examined the TAD structures in these normal B cells at 20Kb resolution. I initially performed some technical and quality controls in these samples and discarded any associated technical issues in the data from chromosome 2 (**Appendix Figure 5**). I also compared those traits between each B-cell replicate and the merge from three replicates. In this way, I could discard the total number of reads as a confounding factor on TADs characterization, so for this study I also merged the data from the three biological replicates (**Appendix Figure 6**). Once a global TAD characterization of chromosome 2 was performed, I analyzed the TAD structure containing the *SOX11* locus. As observed for the MCL cases, a TAD structure in normal B cells also contained both the *SOX11* locus and the superenhancer (**Figure 76A**). Subsequently, I analyzed the interactions of the aforementioned TAD and I could observed that in GCBC and to a lesser extent in PC, the density of interactions were slightly increased as compared to NBC and MBC (NBC=0.94, GCBC=1.16, MBC=0.95, PC=1.02 density of interactions on the SOX11 TAD determined at 20Kb resolution by TADbit (Serra et al., 2017)), (**Figure 76B**).

These results suggest that the TAD structure encompassing the *SOX11* locus gain interactions in some B-cell subpopulations. However, these interactions did not affect the *SOX11* gene expression which remained silent in all B-cell subpopulations. Although those interactions resulted in subtle changes in the compartment classification among B-cell subpopulations (**Figure 76A**), these changes were far more clear in SOX11-positive MCL cases (**Figure 74** and **Figure 76B**). Interestingly, the B-cell subpopulations with lower density of interactions in the *SOX11* locus were the putative cells of origin of MCL subtypes, NBC and MBC (Navarro et al., 2012; Queirós et al., 2016), (**Figure 76B**).
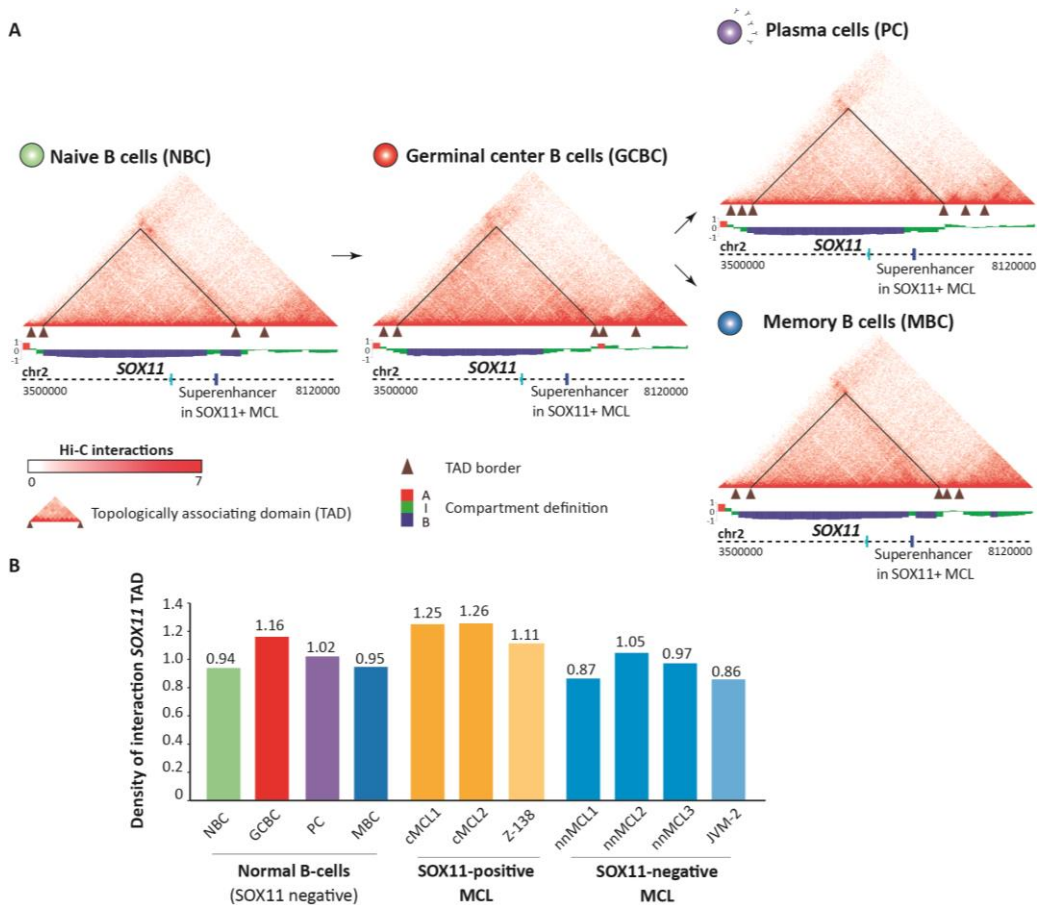
Figure 76. Chromosome conformation through mature normal B-cell differentiation within *SOX11* locus. A. Hi-C contact matrices for the four human normal B-cell subpopulations studied, naive B cells (NBC), germinal center B cells (GCBC), plasma cells (PC) and memory B cells (MBC). Color code indicated the Hi-C interactions. Below, TAD borders and compartment definition are indicated per sample, as well as, the location of the *SOX11* locus and the superenhancer in

SOX11-positive MCL. The coordinates of the represented region are chr2: 3,500,000-8,120,000; GRCh38. **B.** Density of interactions calculated at the TAD containing *SOX11* locus and SOX11-positive MCL superenhancer among all the samples studied.

2.4 Specific active sites on the SOX11-positive MCL superenhancer

In order to shed light into potential mechanisms underlying the *SOX11* superenhancer activation in cMCL cases, I focused my attention into the regulatory region. Analyzing the chromatin states in this region (**Figure 77**), I corroborated that the superenhancer defined in SOX11-positive MCLs was highly activated and correlated with active chromatin at the *SOX11* locus (2 MCL primary cases and the cell line Z-138). In contrast, in both SOX11-negative MCL cases (3 MCL primary cases and the cell line JVM-2) and normal B-cells lacking SOX11 expression, the gene was maintained silenced by poised-promoter marks, while the superenhancer region was mostly heterochromatic and only partially showed weak enhancer marks.
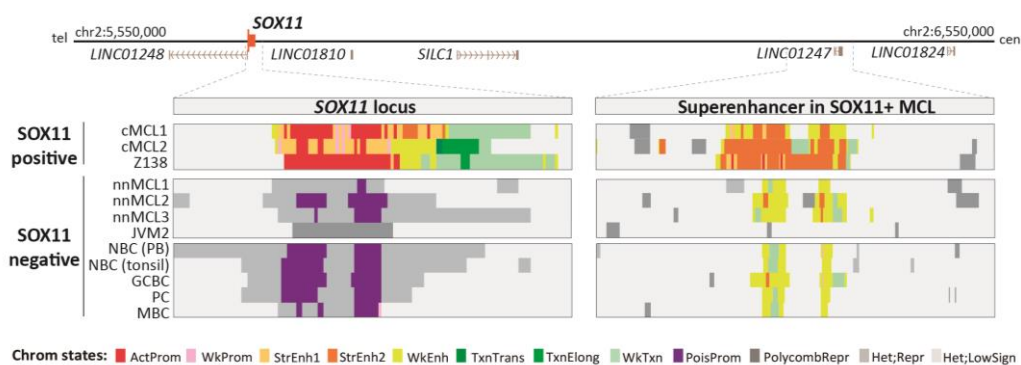
Figure 77. Active chromatin function on *SOX11* locus and superenhancer at SOX11-positive MCL. Chromatin states of the *SOX11* locus and the SOX11 positive MCL superenhancer in SOX11-positive (n=2 MCL cases and n=1 cell line, Z-138) and SOX11-negative samples (n=3 MCL cases, n=1 cell lines JVM-2 and n=5 normal B-cells, a representative B-cell subpopulation was chosen from the three replicates). cMCL, conventional MCL; nnMCL, leukemic non-nodal MCL; NBC, naive B cell; GCBC, germinal center B cell; PC, plasma cell; MBC, memory B cell, PB, peripheral blood. The coordinates of the all the represented region are chr2:5,550,000-6,550,000; GRCh38. The *SOX11* locus coordinates are chr2:5,683,948-5,710,104; GRCh38, and the SOX11-positive MCL superenhancer coordinates are chr2:6,314,751-6,381,194; GRCh38.

The high superenhancer activation in SOX11-positive MCL samples was defined by a great increase of histone acetylation in that region (**Figure 68** from **Study 3**). Due to

this observation, I hypothesized that proteins with acetyl-lysine binding modules or bromodomains would be recruited to the regulatory region to facilitate transcriptional activation (Dhalluin et al., 1999), as has been reported for Myc. In fact, inhibitors of the bromodomain and extra-terminal motif (BET) proteins have been postulated as a therapeutic strategy to treat cancer by targeting e.g. pathologic activation of *MYC* (Delmore et al., 2011). In MCL, the BET bromodomain inhibitor, CPI203, has been used to downregulate MYC expression, which combined with lenalidomide was suggested to have an antitumor activity in specific MCL cases resistant to the bortezomib therapy (Moros et al., 2014). Consequently, to study the effect of BET-inhibitors in SOX11 expressing MCLs and the superenhancer activity, I treated the SOX11-positive Z-138 cell line with CPI203, an inhibitor which targets the BRD4 bromodomain and I also tested a broader BET-inhibitor, which has similar potencies for BRD2, BRD3, BRD4 and BRDT bromodomains, named as CPI0610 (CM1818). Firstly, to assess the right dose, different time points (24, 48 and 72 hours) as well as different concentrations per each inhibitor (200, 500 and 700nM) were considered. Cell proliferation was evaluated by means of MTT assay while cell death was assesed by annexin and propidium iode. A decrease in cell proliferation was observed while no effects could be reported on cell death, at a dose of 500nM after 72 hours of treatment (same CPI203 dose reported on the MCL study (Moros et al., 2014)). Gene expression was assessed for *SOX11* and *MYC,* which was used a positive control. Once the cells were treated, a decrease of *MYC* expression was observed but no effects on *SOX11* could be detected. Overall, these results suggest that bromodomains play a role in MYC activation but not in the case of SOX11 regulation in cMCLs.

Then, I speculated that perhaps a transcription factor might be involved in the aberrant *SOX11* gene expression, as it has been shown for other oncogenes. For example, the studied oncogenic transcription factor TAL1, which activates the *MYB* oncogene in human T cell acute lymphoblastic leukemia (T-ALL) (Lee and Young, 2013; Sanda et al., 2012). To assess this hypothesis, I studied the chromatin accessible regions (as these regions contain an increase probability of transcription factor binding) in the described regulatory region using ATAC-seq in the very same MCL cases analyzed as well as in MCL cell lines and normal B cells. I detected five different accessible regions, named peak 1 to 5, in the SOX11-positive MCL cases (**Figure 78**).
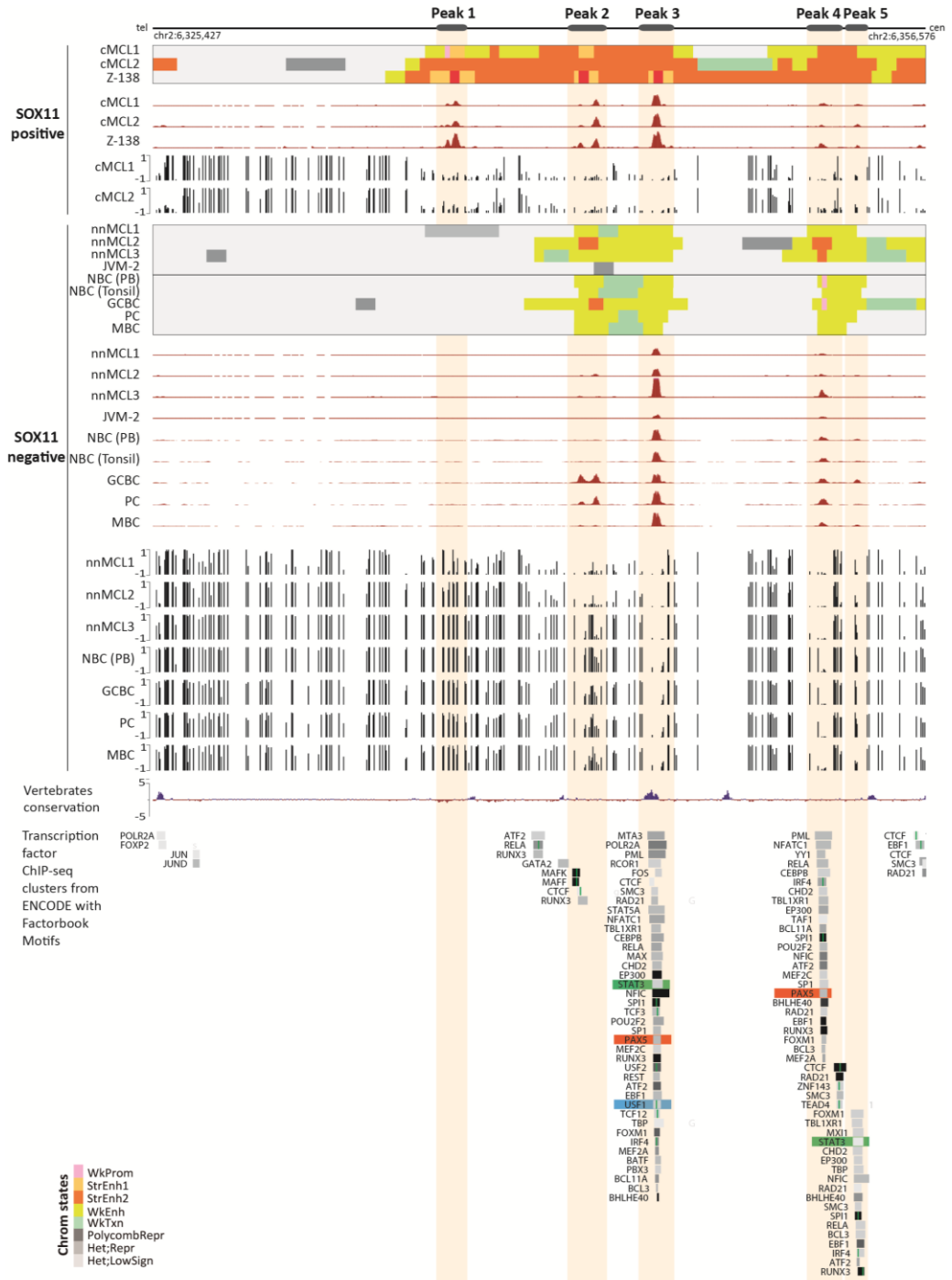
Figure 78. Functional characterization of the chromatin within SOX11-positive MCL superenhancer region. For both SOX11-positive and negative samples (including normal and neoplastic samples) the following tracks are shown, from up to bottom: 1)chromatin states 2).

chromatin accessibility (based on ATAC-seq signals, y-axis signal from 0 to 100). The chromatin accessibility signal shown from each B-cell subpopulation is the median from three different replicates. 3) DNA methylation profiles per each CpG obtained by the quantification of 5-mC signal by WGBS. Below, vertebrate conservation track is shown, considering 100 vertebrate species data (track available on the GRCh37 genome (chr2:6,465,559-6,496,708), followed bythe transcription factor ChIP-seq clusters from ENCODE with factorbook motifs track d. ENCODE TFBS ChIP-seq data represents peaks from 161 transcription factors in 91 different cell types, the track shows occupancy regions for each transcription factor and the motif sites identified within the regions (track available on GRCh37 genome). The transcription factors highlighted are USF2 in blue, PAX5 in orange and STAT3 in green. The coordinates of the represented region are chr2:6,325,427-6,356,576; GRCh38.

Particularly, peak 1 was remarkably interesting as it was present only in the SOX11-positive MCL samples, primary cases and cell line, but not in SOX11-negative MCL samples or normal B cells. Peaks 2 and 5 were present in SOX11-positive MCL cells but also in two more B-cell subpopulations, GCBC and PC. In fact, these two specific peaks may be participating in the slight increase in interactions observed in GCBC and PC, as it was identified by the Hi-C contact matrix (**Figure 76A** and **76B**). Besides, peaks 3 and 4 were detected in all SOX11 positive and negative MCL samples as well as normal B-cell subpopulations. Interestingly, the peak number 3 was the only one conserved among vertebrates (**Figure 78**). All those samples showing accessible peaks displayed less DNA methylation as observed by WGBS analyses (**Figure 78**). Peaks 2 to 5 were identified in some or all B cells lacking SOX11 expression, suggesting that B-cell transcription factors might somehow be also involved in *SOX11* deregulation in MCL (**Figure 78**). Therefore, I analyzed the target region using the transcription factor ChIP-seq data from the ENCODE (Dunham et al., 2012), (**Figure 78**). Transcription factors were binding mainly to the peaks present in all samples.

Furthermore, the active form of STAT3 has been recently described to repress SOX11 transcription (Mohanty et al., 2019). I analyzed these findings in the context of our chromatin data observing that peaks 3 and 5 contained STAT3 binding sites (**Figure 78**). These results suggest that the described function of STAT3 might be exerted on the accessible regions present both in normal B cells as well as SOX11 positive and negative MCL cases, and therefore, does not seem to be involved in the *de novo* overexpression of SOX11 in cMCL. To obtain further insights into potential TFs involved in SOX11 overexpression, I focused specifically on two peaks. The peak number 1, present

exclusively in SOX11-positive MCL samples lacked TF binding sites using the ENCODE ChIP-seq data, suggesting that this accessible peak occurs *de novo* in SOX11-positive MCLs and could be the trigger for the oncogene expression on the aggressive MCLs. This peak number 1 was named as SOX11-positive MCL peak. Besides, peak number 3 present both in MCL and normal B cells was further studied. This peak might be involved, as explained above, in the negative regulation on *SOX11* (Mohanty et al., 2019). However, as within this accessible peak 3 region, the enhancer changed from weak to active state only in SOX11-positive MCLs (**Figure 78**), I hypothesize that upon SOX11-positive MCL transformation, this particular region might contribute to enhance *SOX11* gene expression. In order to identify potential TFs and proteins binding to peak 1 and peak 3, I performed a targeted mass spectrometry (MS) in the SOX11-positive MCL cell line, Z-138.
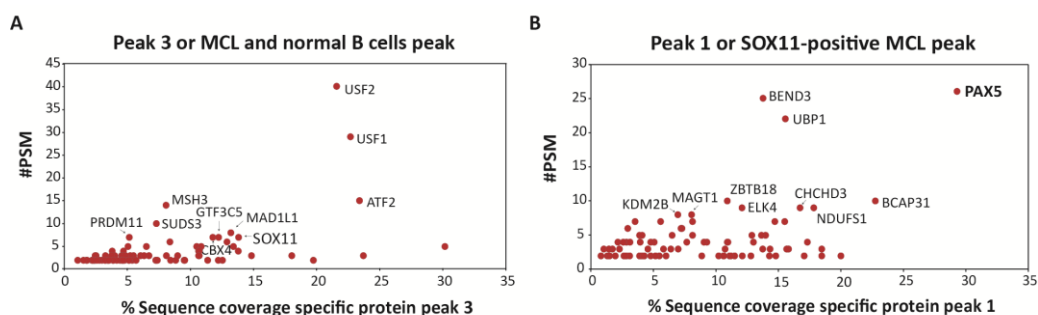
Figure 79. Proteins present at specific chromatin accessible peaks. Graphs showing at the x-axis the PSM, peptide spectrum matches, which is the total number of identified peptide sequences for protein and at the y-axis the percentage of the protein sequence identified in the analysis or coverage. **A.** Specific proteins on peak 3 or MCL and normal B cells peak. **B.** Specific proteins on peak 1 or SOX11-positive MCL peak. The name of the top ten protein genes ordered from higher to lower PSM and coverage are listed. A protein can be estimated to be more abundant if the sequence coverage is high and the peptide is abundant at the sample (high PSM).

From the list of filtered proteins obtained per each peak, I identified those that were specific for each of these two regions (**Figure 79**, **Appendix Table 11** and **12**). Among the proteins associated with peak 3, USF (USF1 and USF2) was detected, which is a regulator of transcription for many genes during cellular differentiation (Anantharaman et al., 2011; Sirito et al., 1994). In a liver cell line, it was observed that

USF1 and USF2 interact together with thousands of genomic loci, emphasizing our finding that the heterodimer is predominant active form of this TF (Rada-Iglesias et al., 2008). Also, SOX11 protein was located in this peak suggesting that SOX11 might by regulating itself by binding on this chromatin accessible peak (**Figure 79A** and **Appendix Table 11**). Indeed, a SOX11 positive feedback loop was already suggested by analysis of transcription-factor binding motifs enrichment at the described SOX11-positive MCL superenhancer, which revealed the presence of potential binding sites for SOX protein family members (Beekman et al., 2018b). In the case of the SOX11-positive MCL peak (or peak 1), the B-cell transcription factor PAX5 was detected to be highly enriched (**Figure 79B** and **Appendix Table 12**). To further study this finding, I analyzed the PAX5 binding motifs in peak 1 and 3, and I could detect potential binding motifs for PAX5 on both peaks (**Figure 80** and **Appendix Table 13**), although they contained different sequences. Furthermore, the peak 4 contained the very same binding motif as the SOX11-positive MCL peak. Indeed, the transcription factor ChIP-seq data from ENCODE revealed the presence of PAX5 both in peak 3 and 4. The *SOX11* locus itself contained also both PAX5 binding sites (**Appendix Table 13**).
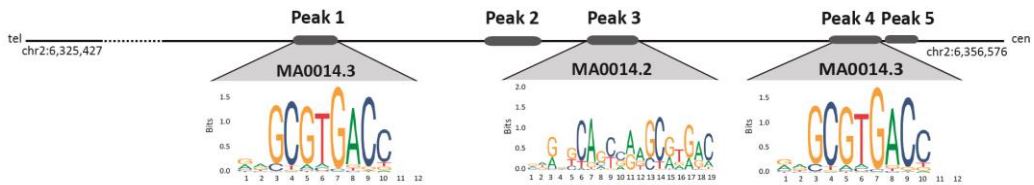
Figure 80. PAX5 binding motifs on SOX11-positive MCL superenhancer. Two distinguished PAX5 binding motifs were detected on the superenhancer regulating SOX11 in SOX11-positive MCL cases. The coordinates of the represented region are chr2:6,325,427-6,356,576; GRCh38.

In contrast to the ENCODE ChIP-seq data, my MS results did not identify PAX5 binding to peak 3 (i.e. present in MCL and normal B cells), which I speculate could be a false negative result due to the high amount of transcription factors binding on this conserved peak. Most remarkably, PAX5 is the most clearly enriched TF observed in the MS from the specific peak 1 (i.e. present only in SOX11-positive MCL) suggesting that it might be associated with SOX11 upregulation by aberrantly binding to this specific site. To further demonstrate binding of PAX5 to peak 1, ChIP-qPCR using an antibody against PAX5 was performed in Z-138, the MCL cell line expressing *SOX11*, and JVM-2, a MCL

cell line lacking *SOX11* expression. I detected an increased enrichment of PAX5 in peak 1 as compared to the peak 3 in Z-138. However, peak 3 was also slightly enriched in PAX5 as compared to a negative control (**Figure 81**). In JVM-2, PAX5 binding was not observed in neither of the studied peaks (**Figure 81**).
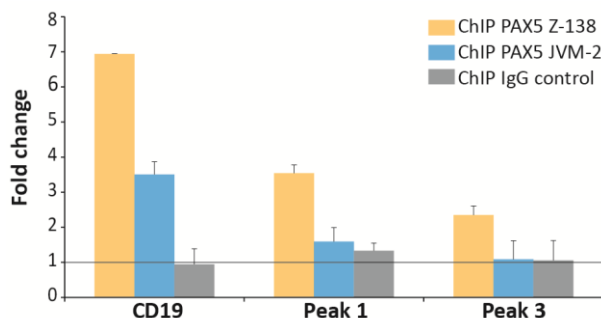
Figure 81. Binding of PAX5 onto MCL SOX11 superenhancer in Z-138. ChIP-qPCR validation of a selected positive region (CD19) and the two chromatin accessible peaks, 1 and 3, on a SOX11 positive cell line, Z-138 (yellow) and negative MCL cell line, JVM-2 (blue). The fold change enrichments relative to a negative region are presented.

These results shed light on the mechanisms leading to the *SOX11* oncogene expression in MCL. The B-cell TF, PAX5, expressed from the pro-B to the mature B cell stage (Fuxa and Busslinger, 2007a), which is in charge of controlling the differentiation, function and identity of B-lymphocytes (Cobaleda et al., 2007), might be leading *SOX11* aberrant expression on MCL cases. It is indeed intriguing that a B cell-specific TF can be related to the expression of a gene such as *SOX11*, which is not expressed in the B-cell lineage. It seems that in the context of MCL lymphomagenesis, PAX5, which may be already present in the B cell-specific peaks 3 and 4 of the superenhancer, also binds *de novo* to peak 1. This *de novo* binding of PAX5 may recruit chromatin remodeling complexes as well as histone-modifiers and transcription factors to induce chromatin changes (McManus et al., 2011), which in turn may lead to *SOX11* overexpression in MCL. Although these findings contribute to our understanding of the potential causes of SOX11 expression in MCL, further studies are required to completely elucidate the factors and chain of events associated with this oncogenic activation.

3. Conclusions

- Fluorescence *in situ* hybridization analyses in SOX11-positive and SOX11-negative MCL samples and normal B cells indicated that the 3D chromatin loop between *SOX11* locus and the SOX11-positive MCL superenhancer was biallelic.

- The topologically associating domain (TAD) containing *SOX11* locus was maintained throughout normal and neoplastic B cells. Although the density of 3D interactions in the SOX11 domain in SOX11-positive was higher than in SOX11-negative MCL cases.

- Five different accessible chromatin peaks can be detected in the SOX11-positive MCL superenhancer that loops to the *SOX11* locus.

- Out of the five accessible peaks, only one was exclusively present in MCLs expressing SOX11, and the remaining were present in normal B cells and/or SOX11-negative MCLs.

- The PAX5 transcription factor was binding to the accessible peak specifically present in the SOX11-positive MCLs, suggesting that it may be involved in the aberrant expression of the *SOX11* oncogene in MCL.

DISCUSSION

1. Compartmentalization of the three-dimensional structure

Spatial folding of chromosomes and their three-dimensional organization in the interphase nucleus is being recognized to play a crucial role in the complex process of regulating gene expression. To obtain further insights into this relatively understudied layer of epigenomic information, this doctoral thesis has focused on analyzing the 3D chromatin structure of normal and neoplastic B cells as cellular models. A comprehensive analysis of the 3D genome structure, going from specific chromatin loops to genome-wide assemblies was performed to examine in detail how the nuclear architecture is modulated during normal B-cell differentiation and upon neoplastic transformation into chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL). An initial exploration of the distribution of Hi-C eigenvector coefficients led to a novel categorization of the genome into three components. Integrating these three components with nine additional omic layers including DNA methylation, chromatin accessibility, six histone modifications and gene expression, revealed the presence of an intermediate (I) component between the active (A) and repressed (B) compartments. This I-type compartment contained more inter-compartment interactions than fully active or inactive chromatin, and was enriched in the H3K27me3 histone mark associated with poised promoters and polycomb-repressive chromatin states. Thus, this categorization into three components seemed to be functionally more appropriate than the well-established dichotomous separation of the genome into A and B compartments (Lieberman-Aiden et al., 2009). Moreover, the three component distribution resembles the traditional chromatin structure organization into euchromatin and heterochromatin further distinguishing the two subcategories of heterochromatin, constitutive and facultative. Hence, the I-type compartment may represent the facultative heterochromatin, a labile state of the high-order chromatin organization that can evolve either into active or inactive chromatin compartments (Janssen et al., 2018). Alternative models to the two classical A and B compartments have been already reported in the literature based on long-range interactions. Three compartments were identified by k-means clustering describing a third gene-poor cluster (Yaffe and Tanay, 2011) or a six subcompartment categorization was reported according to their chromatin signature (Rao et al., 2014). Among the six compartments, the B1 subcompartment, associated with compartment B, was defined and functionally characterized to correlate positively with H3K27me3. Strikingly, the H3K27me3 mark

was initially associated with A compartment (Lieberman-Aiden et al., 2009) and later with the subcompartment B1 (Rao et al., 2014), maintaining genes inactivate on an active environment. An analysis correlating both types of compartmentalizations was performed, and I observed that the I-type compartment was composed of different percentages from the six subcompartments identified by Rao and co-workers. The I-type compartment showed the highest proportion of B1 but also contained significant fractions of other compartments (**Figure 82A**). Similar results were obtained when these two types of segmentations were performed using Hi-C data from the GM12878 cell line (**Figure 82B**). These results show that, although there is some overlap between I-type and B1 compartments, they appear to reflect distinct structures. However, these differences may also be influenced by distinct approach used by the two compartmentalization methods. The six subcompartments clustering was based on a subset of the inter-chromosomal contact data while a more straight-forward approach including all interactions was use to determine the I-type compartment.
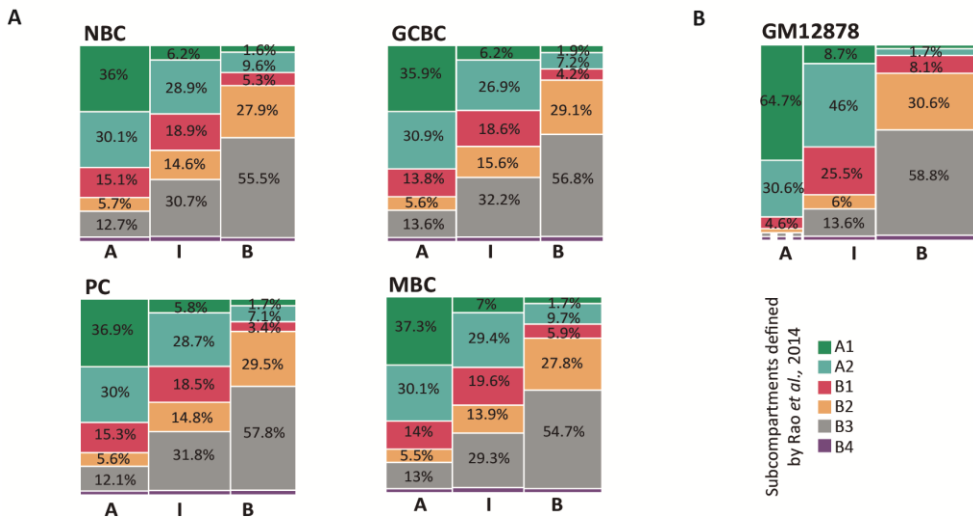
Figure 82. Correlation among different compartment segmentations. **A.** Correlation between the three compartment segmentation in normal B-cell subpopulations, as defined in this doctoral thesis, and the six subcompartments defined by Rao and collaborators. **B.** Correlation of the two methods in GM12878. The percentage (>1%) of each subcompartment is indicated. The discontinuous line represents the subcompartment not enriched in the correlation.

The existence of the I-type compartment is further supported by several additional lines of published evidences. For example, during T-cell commitment, a correlation between intermediate compartment scores (i.e. between A and B) with intermediate levels of gene expression was observed (Hu et al., 2018). Also, the poised chromatin mark, H3K27me3, has been observed to enrich both A and B compartments in plants, such as *Arabidopsis thaliana* (Dong et al., 2017), and also in *Drosophila melanogaster* (Boettiger et al., 2016). Recently, using super-resolution imaging, it was found that some compartments could belong to active or inactive states depending on the observed cell (Nir et al., 2018), which could resemble an intermediate compartment in a population-based analysis such as Hi-C. Finally, these evidences are also in line with the observation that members of the key elements of the polycomb (PcG) complex, polycomb repressive complex 1 (PRC1), seems to form discrete subnuclear chromatin domains (Boettiger et al., 2016; Kundu et al., 2017; Wani et al., 2016). Thus, the facultative heterochromatin compartment in the nucleus might be further characterized by small clusters of PcG to larger sized PcG bodies, which have been suggested to bring PcG-regulated genes together (Wani et al., 2016). These clusters are frequent and spread along the nucleus while PcG bodies are rare (Wani et al., 2016), generally localized close to pericentromeric heterochromatin and linked with gene repression (Saurin et al., 1998). Thus, PcG proteins seem to form phase-separated condensates assembled through liquid-liquid phase separation (LLPS) (Tatavosian et al., 2019). The phase-separation phenomenon described as membrane-less organelles that concentrate specific molecules (Hyman et al., 2014; Shin et al., 2018) was firstly suggested on heterochromatin-mediated gene silencing interceded by the HP1α protein (Larson et al., 2017). However, PcG-bound chromatin might create a local chromatin environment as shown on mouse and human cell models (Kundu et al., 2017; Nir et al., 2018; Rada-Iglesias et al., 2018; Schoenfelder et al., 2015; Vieux-Rochas et al., 2015), (**Figure 83**), and has been further described to be dynamically modulated during cell differentiation (Mas et al., 2018). Interestingly, *EZH2* encoding the catalytic subunit of PCR2, a highly conserved H3K27 methyltransferase, is in charge of the epigenetic repression of genes which enables the proliferative phenotype in GC, and is absent in quiescent NBC (Béguelin et al., 2017; Guo et al., 2018). In fact, chemical inhibitios of EZH2 has been observed to enhance plasma cell formation (Scharer et al., 2018). Apart for B-cell differentiation, this protein is relevant in B-cell tumors, as gain

of function mutations in *EZH2* have been described in B-cell lymphomas (Ernst et al., 2010; Morin et al., 2010) which lead to changes in chromatin structure and transcription activity (Donaldson-Collier et al., 2019). Overall, with this studies in mind one may suggest that structural compartments containing polycomb might be of crucial relevance for gene regulation.
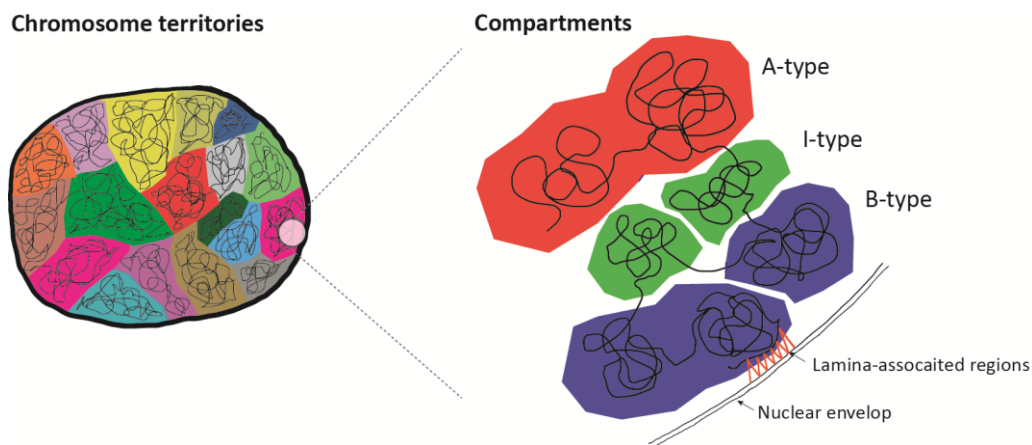
Figure 83. **Representation of the I-type compartment.** Each chromosome occupies a distinct territory in the cell nucleus and each chromosome might be partitioned into three distinct compartments.

2. Three-dimensional chromatin structure during B-cell differentiation

The genome architecture of the B-cell differentiation program has been previously studied in cell lines and primary mouse cells (Johanson et al., 2018; Kieffer-Kwon et al., 2013; Lin et al., 2012; Martin et al., 2015; Mumbach et al., 2017; Stadhouders et al., 2018) or during the human germinal center reaction (Bunting et al., 2016). In this doctoral thesis, I made one step ahead and I studied four mature human B-cell subpopulations sorted from healthy donors and I analysed their chromosomal organization by *in situ* Hi-C. I observed that 71.9% of the compartment structure was conserved whereas the remaining 28.1% of the genome was dynamically altered in particular B-cell maturation transitions. This proportion is in line with the compartment transitions observed during *in vitro* differentiation of human embryonic stem (ES) cells into four cell lineages, where the switch from A to B compartments and vice versa accounted for 36% of the genome in at least one of the lineages (Dixon et al., 2015). Furthremore, a similar proportion was observed in the reprogramming of mouse

somatic cells into induced pluripotent stem cells (iPS) (Krijger et al., 2016; Stadhouders et al., 2018). In that study, 28% of the genome was changing compartments in at least one of the founder cells studied (pre-B cells, neural stem cell, macrophage or fibroblast) (Krijger et al., 2016). Finally, in 21 human cells and tissues, 59.6% of the genome showed compartment transitions, indicating that as we increase the number of cell states studied a larger proportion of the genome shows changes in the 3D genome structure (Schmitt et al., 2016b). Collectively, these findings indicate that the genome architecture is a highly dynamic trait.

The compartment modulation linked to B-cell maturation was mainly related to two biological phenomena. First, I observed a large-scale activation from NBC to GCBC and second, a reversion of the 3D genome organization in MBC back to the one observed in the pre-germinal center NBC subpopulation. The GCBC-specific structural activation supports a previous study in which the chromatin structure of GCBC undergoes global decompaction (Bunting et al., 2016), which in turn is in line with a other report indicating that compartment activation implies a decreased number of mid-range 3D interactions and therefore, a less compact chromatin (Le Dily et al., 2014). The outcome of the germinal center reaction is PC or MBC, which are phenotypically and functionally distinct subpopulations. GCBC and PC show an overall high level of conservation of their 3D genome organization, but the differentiation into MBC is related to extensive changes. Remarkably, roughly three quarters of the compartment changes from GCBC to MBC were reverting back to the compartment profile observed in NBC, as mentioned above. This reversibility of the higher-order chromatin structure is very much in line with the previously observed similarity of histone modifications, chromatin accessibility and gene expression profiles in NBC and MBC. In sharp contrast to this congruent behavior of chromatin-based traits and the transcriptome, DNA methylation is rather different between NBC and MBC, as this mark follows an accumulative pattern during cell differentiation (Kulis et al., 2015; Shearstone et al., 2011) and can be used to faithfully track the lineage trajectory of the cells (Gaiti et al., 2019). This apparent lack of association between DNA methylation and 3D genome organization may not be always a rule, as it may be exemplified in the study of different brain cell types, where significant overlaps have been described between cell-type specific chromatin interactions and differentially methylated regions (DMRs), suggesting that for some specific loci, DNA methylation indeed may correlate

with 3D genome interactions (Lee et al., 2019). In MCL, the integrative approach applied in this doctoral thesis combining DNA methylation and histone modification patterns led to the identification of DMRs with potential functional impact. In this context, epigenetic drivers were characterized as recurrent DMRs containing at least some B cell-independent CpGs and a concurrent change in the chromatin activation state, such as the distant regulatory region in SOX11-positive MCL cases. This integrative methodology identified locus-specific associations between DNA methylation changes and histone modifications along with the 3D genome structure.

The study of the 3D conformation during B-cell differentiation represents a valuable resource as it permits to identify and place specific chromosome conformation changes found in different haematological malignancies in a proper context of accurate cell of origin. For instance, a previous study in multiple myeloma combining Hi-C data, whole-genome sequencing (WGS) and gene expression data used the lymphoblastoid GM12878 cell line as control. However, the study pointed out that some of the recognized changes in myeloma might be due to the difference between lymphoblastoid B cell and plasma B cells (Wu et al., 2017), the cell of origin of multiple myeloma, suggesting the importance of using the right normal B-cell subpopulations to identify tumor-specific changes in B cell neoplasms.

3. Changes in the three-dimensional chromatin structure in B-cell neoplasms

The two B-cell neoplasms studied in this doctoral thesis, CLL and MCL, showed tumor-specific changes in their 3D genome organization, which represent a novel type of molecular alteration in these diseases. Compartment shifts as observed in CLL and MCL compared to normal B cells have also been reported comparing mouse pro-B and lymphoma cells with embryonic stems cells (Augusto et al., 2019), normal T-cells and leukemia (T-ALL) patient samples (Kloetgen et al., 2020), as well as in normal breast cancer (MCF-10A) and its cancerous counterpart (MCF-7) (Barutcu et al., 2015). Interestingly, although MCL is chromosomally more instable than CLL and shows more genetic rearrangements, at the level of 3D genome compartments, MCL as a whole is more similar to normal B cells than CLL. Even so, the compartments changes identified in CLL and MCL span over large region and contained genes linked to their pathogenesis. An example of compartment shift between normal B cells and CLL is a large 2Mb region involving the *EBF1* gene, which encodes a key developmental

regulator that orchestrates B-cell fate. *EBF1* is expressed in B cells from their early development, and upon CLL transformation, the gene is silenced along with a long-range remodelation of its chromatin structure as compared to its normal counterpart. In fact, it has been reported that upon B-cell committment, the *EBF1* locus switch compartments and that it was linked to changes in transcription signature (Lin et al., 2012). *EBF1* locus is silenced and tightly associated with nuclear lamina in multipotent progenitors but relocated away from the lamina in committed pro-B cells, leading to its activation (Lin et al., 2012). In CLL, *EBF1* inactivation linked to a change in its 3D structure may result in a chromatin organization at that locus similar to less differentiated cellular stages.

As in the case of *EBF1*, another region of particular interest was the observed structural activation of 6.1Mb affecting the entire chromosome band 2p25.2 in aggressive cMCL, which contains the *SOX11* oncogene, a biomarker whose expression defines this MCL subtype (Fernandez et al., 2010) and plays key functional roles in its pathogenesis (Beekman et al., 2018b). However, mechanisms underlying the aberrant expression of this oncogene in MCL are not yet well established. From the 3D genome perspective, it was not related to an increase in insulation around the *SOX11* gene as was shown at neuronal cells where the gene is expressed (Bonev et al., 2017). In MCL, the oncogene is embedded into a long-range alteration in the 3D genome structure. This change is not only linked to *SOX11* overexpression but seems to be related to the simultaneous overexpression of multiple genes within the target region. This phenomenon of long-range 3D genome changes also involve the opposite phenomenon i.e. a switch from compartment A to compartment B, as it was described between normal B cells (GM12878) and multiple myeloma (U266 and RMP1-8226) cell lines leading to the downregulation of a cytokine receptor gene cluster at 2q11.1-q12.1 (Wu et al., 2017). Thus, the phenomenon of long-range epigenetic changes occurs in two flavors, as epigenetic silencing (LRES) and epigenetic activating (LREA). These large epigenetic remodeled regions are in fact characteristic from several cancer types, such as colorectal, bladder, non-small cell lung cancer, breast, prostate, and Wilms tumor (Achinger-Kawecka et al., 2016; Bert et al., 2013; Dallosso et al., 2009; Frigola et al., 2006; Hitchins et al., 2007; Novak et al., 2008; Rafique et al., 2015; Seng et al., 2008; Stransky et al., 2006; Taberlay et al., 2016). Particularly, long-range epigenetic activation (LREA) regions have been identified in prostate cancer harboring oncogenes,

microRNAs and cancer biomarker genes (Bert et al., 2013). Those regions have been associated with gains in active histone marks and loss of repressive histone marks (Bert et al., 2013). Furthermore, at the level of DNA methylation, DNA hypermethylation over one chromosomal band of 4Mb has been linked to silencing of several genes in colorectal cancer (Frigola et al., 2006). DNA methylation is related to genomic architectural proteins, such as the methylation sensitive DNA-binding protein CTCF, which is a key regulator of higher-order chromatin structures (Wang et al., 2012). One of the best characterized CTCF chromatin architectures was identified at the imprinted Igf2/H19 locus and regulates imprinting. It was found that Igf2 repression on the maternal allele is achieved by preventing the interaction between the gene and a distal enhancer through the formation of chromatin loops mediated by CTCF. In contrast, its binding to the imprinted control region (ICR) and insulator looping are prevented by DNA methylation on the paternal allele, allowing the Igf2 gene to contact the distal enhancer by transcription factor-mediated looping (Murrell et al., 2004; Nativio et al., 2009, 2011). However, at a small subset of sites, CTCF is a methyl-sensitive protein (Feldmann et al., 2013; Maurano et al., 2015; Merkenschlager and Nora, 2016; Wang et al., 2012). Indeed, a recent study developed single-nucleus methyl-3C sequencing to capture chromatin organization and DNA methylation and showed that, in reality, just a minority of CTCF binding sites may be subjected to DNA methylation regulation (Lee et al., 2019). Nonetheless, CTCF binding at the LREA/LRES boundary appeared to be unaltered in the cancer cells, so other deregulated CTCF cofactors may be involved in promoting switching between the two opposing epigenetic domains (Bert et al 2013). Another report suggested that DNA methylation might not regulate the transcription factor landscape, such as the case of CTCF, but is rather targeted at labile occupancy sites (Maurano et al., 2015). Although is interesting to point out that DNA binding factors have been suggested to mediate turnover of DNA methylation on maintenance and reprogramming of regulatory regions (Feldmann et al., 2013). These studies revealed an intricate relationship between epigenetic players and their implication in cancer biology. In fact, the most studied epigenetic layer, DNA methylation, represents an important biomarker to differentiate different subtypes of cancers. In the case of MCL, DNA methylation signatures are able to separate samples into two distinct subgroups based on similarities to normal B-cell subpopulations. Hence, C1 MCLs were described presenting an aggressive clinical behavior, acquired mutations and a DNA

methylation pattern resembling germinal-center independent B cells (and mostly correspond to cMCL). In contrast, C2 MCLs showed an indolent clinical course, lack of oncogenic mutations and resemble experienced germinal center reaction B cells (and mostly correspond to nnMCL). Moreover, DNA methylation distinguished three different CLL clinico-biological entities (Kulis et al., 2012; Oakes et al., 2016). Additionally, five different CpGs defined at enhancer chromatin regions were able to distinguish the CLL subgroups (Bhoi et al., 2016; Queirós et al., 2015). These findings indicate that DNA methylation can be used as a biomarker for CLL, as described also for other cancers or diseases where specific DNA methylation is able to diagnose and predict prognosis (Locke et al., 2019). Moreover, specific chromosomal conformations have been already identified as new possibilities for less invasive blood-based biomarkers (Babu and Fullwood, 2015; Bastonini et al., 2014; Mukhopadhyay et al., 2014). Those studies open the door to new detection systems and also different therapeutic strategies. Indeed, at the epigenetic level, several classes of anticancer drugs have been described including DNA methyltransferase inhibitors, histone deacetylase (HDAC) inhibitors, lysine-specific demethylase 1 (LSD1) inhibitors or enhancer zeste homolog 2 (EZH2) inhibitors and bromodomain and extra-terminal motif (BET) inhibitors (Doroshow et al., 2017). Moreover, recently curaxins have been characterized as an anti-cancer agent that intercalates DNA inducing dissociation of CTCF from DNA and changing thus the 3D genome organization (Kantidze et al., 2019; Schütte et al., 2019).

4. Three-dimensional chromatin conformation and structural variants

As chromosomes are distributed in chromosome territories within the interphase nucleus (*see section 1.4.1.1 Chromosome territories*), it is logical to postulate that the presence of chromosomal rearrangements, in particular those affecting more than one chromosome, will affect the results detected by Hi-C. In order to address this, I wondered if I could detect structural variants (SV) using the Hi-C data. I studied the changes in the 3D genome architecture associated from SV taking place in the CLL and MCL samples under study. I visually distinguished the chromosomal translocations in each sample on the contact matrix of 1Mb resolution. However, to neatly identify SVs I extracted from the interaction matrix of each case the interactions present on the cell of origin described per each case, allowing a more straight-forward visual inspection.

Comparing this technique with standard karyotyping or next-generation sequencing technologies spotlights Hi-C as a method not only to validate previously reported SVs but event to identify new SVs. In the literature, Hi-C has been applied to map translocations associated with DNA double-strand breaks (DSBs) in mouse pro-B cell genomes (Zhang et al., 2012). Besides, a recent publication uses Hi-C to infer chromosome rearrangements and copy number information from primary tumor material (Harewood et al., 2017). Also, in a proof-of-principle study, the low-C method, a derivative of Hi-C, was used to identify the t(3;14) translocation in a patient with diffuse large B-cell lymphoma, a rearrangement known to juxtapose the *BCL6* and IGH loci (Díaz et al., 2018). These studies pave the way to the usage of chromosomal conformation methods to detect and characterize structural variants, including both balanced and unbalanced chromosome rearrangements and copy number changes. The use of Hi-C overcomes some of the drawbacks of conventional cytogenetic analysis as well as next-generation sequencing (NGS) technologies and associated pipelines to detect SVs. Traditional cytogenetic G-band preparations need dividing cells in culture to generate metaphases while Hi-C is applicable for any tumor type regardless of the availability of dividing cells *in vitro*. Balanced rearrangements such as inversions or reciprocal translocations are also cryptic in some cases to be detected by cytogenetic analysis or assessed using array-based copy number variation (CNV)-based methods. In multiple myeloma, CNVs were observed to bias Hi-C data, even though contact matrices were used to detect them (Wu et al., 2017). Furthermore, whole-genome sequencing analyses are hampered by poorly mappable or repetitive regions. Therefore, it cannot detect rearrangements involving centromeric, heterochromatic or high homologous regions. Additionally, WGS cannot detect translocations positioned at the extreme of the chromosome. Importantly, Hi-C data provides information about the entire piece of rearranged chromosome, not just about the breakpoints. The integration of multiple lines of genomic information with 3D interactions may be an interesting way to assemble cancer genomes. For instance, CESAM, *cis* expression structural alteration mapping, is an approach to detect somatic copy number alterations (SCNAs) leading to gene dysregulation in *cis*, which integrates genome sequence, epigenetics and 3D genome information. However, Hi-C has some drawbacks in this context. The laborious experimental protocol and the cost of the technique still makes Hi-C an unaffordable method to be implemented in the clinics, at

least nowadays. Another inconvenience is the analytical method to detect rearrangements. In the last years, specific algorithms have been created to identify rearrangements (Burton et al., 2013; Harewood et al., 2017), but they need to be improved to reach a more objective method than the raw visual inspection. In fact, as the file is constantly moving forwards, a recent publication described an standardized bioinformatic algorithm to detect aberrations from Hi-C data (Wang et al., 2020). Altogether, SVs are essential to improve the detection and classification of hematological malignancies and finally lead to the development of targeted therapies (Schütte et al., 2019). Thus, their detection is of utmost importance in translational oncology, and the use of chromatin structure may represent a valuable asset in the precise characterization of SVs in cancer genomes.

In addition to detect SVs, Hi-C data can provide more details about chromatin structure dynamics related with chromosome rearragments. I found that 77.2% of case-specific SVs in MCL were taking place in TAD borders, suggesting either that breaks take place in TAD borders in normal B cells or that the SVs itself induces a TAD border. Comparing the TAD borders between MCL cases carrying specific SVs and normal B cells, I could identify that a large proportion of the TADs were *de novo* created in the tumor cells. This finding is in line with published evidence indicating that SVs mostly induce novel TAD borders. For instance, in prostate cancer, around 70% of the CNVs shared between two prostate cancer cell lines (LNCaP and PC3) seem to be located at newly formed cancer-specific domain boundaries (Taberlay et al., 2016). CNV breakpoints occurring near TAD boundaries have been also described in multiple myeloma, although at a lower degree, i.e. 30.7% of the CNV breakpoints were located within 120Kb of TAD boundaries (Wu et al., 2017). Hi-C may also help to characterize chromatin landscape of somatic copy number alterations (SCNAs), which are frequently considered to affect coding genes, e.g. a duplication containing an oncogene or a deletion containing a tumor suppressor gene (Li et al., 2018). However, less than one-quarter of recurrent SNCAs across multiple cancer types overlapped with known cancer-driving genes (Zack et al., 2013). Another way through which SNCAs extert their function in disease is by bringing enhancer elements close to genes that are not regulated by them, leading to their deregulation. This mechanism is known as enhancer hijacking (Northcott et al., 2017) or enhancer adoption (Lettice et al., 2011). In MCL, the *CCND1* gene is deregulated due to a translocation that juxtaposes the gene next to

enhancer elements of the IGH locus, which is highly active in B cells (Bosch et al., 1994). The translocation, as observed on the contact matrices from five different MCL cases in this doctoral thesis, seem to cause new TAD boundaries at the translocation breakpoints. Such new TAD may be related to aberrant *CCND1* expression. Overall, I addressed the importance of the chromatin structure as a new layer of information to consider when studying cancer-specific genetic alterations.

5. Transcription factors as potential drivers of chromatin reconfiguration

The genome is organized into epigenetically and transcriptionally appropriate domains. The role of the architect instructing structural changes in the genome has been assigned to transcription factors (TFs) (Natoli et al., 2010), a model that is supported by several lines of experimental evidence. A recent report shows that TFs drive topological genome reorganizations even before detectable changes in gene expression (Stadhouders et al., 2018). A detailed analysis of regions that become exclusively active in GCBC as compared to any other B-cell subpopulation revealed an enrichment in TF binding motifs of the MEF2 and POU families, which have been described to play a key role in the germinal center formation (Song and Matthias, 2018). Those transcription factors in GCBC could be linked to the dynamic behavior of the chromatin landscape by inducing compartment shifts. Indeed, the switching between compartments has also been described during the reprogramming of B cells into induced pluripotent stem cells, a process that is mediated by binding of the four Yamanaka TFs, i.e. OCT4, SOX2, KLF4 and MYC (OSKM), and the sequential expression of CCAAT/enhancer-binding protein alpha (C/EBPα) (Stadhouders et al., 2018). Apart from the important role of TFs in activating chromatin in GCBC, NFAT and TCF binding motifs were enriched in compartments specifically activated in CLL. These TFs have also been previously linked to *de novo* active regulatory elements in CLL and its pathobiology (Beekman et al., 2018a; Mallm et al., 2019).

Cell identity emerges from a dynamic interplay between chromatin-associated proteins, epigenetic modifications and a spatially organized genome. Cell fate-instructive TFs characterized by cell-type-specific expression, responsiveness to signals, DNA sequence specificity and ability to open non-permissive chromatin are matched to confer spatio-temporal alterations of the 3D chromatin landscape (Stadhouders et al., 2019). All these results are concordant with studies in which lineage-restricted TFs

have been proposed to establish and maintain a specific genome architecture (Heinz et al., 2010; Johanson et al., 2018; Montefiori et al., 2016; Natoli, 2010). For instance, PAX5 is an important regulator of B-cell commitment and development (Fuxa and Busslinger, 2007b). It regulates gene expression by inducing active chromatin at activated target genes and eliminating active chromatin at repressed genes in pro-B cells. In the presence of histone acetylation, PAX5 has been described to work as activator by interacting with histone acetyltransferases (HAT). However, corepressors from the larger histone deacetylase (HDAC) complex have been shown to convert PAX5 to a transcriptional repressor, which results in the shutdown of signaling systems (Cobaleda et al., 2007). Hence, PAX5 seems to act as an epigenetic remodeler by inducing chromatin and transcription changes by means of chromatin-remodelers, histone-modifiers and basal transcription factor complexes to its target genes (McManus et al., 2011). Surprisingly, PAX5 seemed to be present at a specific chromatin accessible peak in the SOX11-positive MCL superenhancer region. This finding suggests that PAX5 might be a relevant factor on the pathogenesis of MCL as it may be involved in upregulating the *SOX11* oncogene. PAX5 not only may underlie *SOX11* gene expression, but SOX11 itself has been shown to contribute to PAX5 up-regulation, which represses genes involved in plasma cell differentiation (Ferrando, 2013; Vegliante et al., 2013). PAX5 enables or limits activation of gene promoters by recruiting the chromatin remodeling complexes (CRCs), SWI/SNF or Mi-2/NuRD, respectively (Gao et al., 2009), a role shared with EBF1, which also regulates PAX5 (Medina et al., 2004). Thus, I hypothesize that PAX5 may bind to the SOX11 superenhancer due to the presence of EBF1, as binding sites of this TF were located at accessible peak 1 site and also at peak 2 and 4 within SOX11 superenhancer region. Indeed, ENCODE transcription factor ChIP-seq data showed EBF1 at the peak 4. Hence, the transcription factor EBF1 (Lin et al., 2012; Treiber et al., 2010) and PAX5 may bind to the B cell-related accessible peaks of the *SOX11* enhancer region in immature B-cell stages. Due to factors that are still unknown, PAX5 might aberrantly bind to the MCL-specific accessible site of the enhancer and eventually lead to *SOX11* aberrant expression in conjunction to chromatin modifiers. In this context, PAX5 has been previously described to induce a permissive genome organization, as this TF has been reported to modify the genome topology even in the absence of transcription (Johanson et al., 2018). Interestingly, PAX5 seems to interact with CTCF in B-cell

progenitors. Nonethless, the contact between PAX5 and other B cell TFs with architectural proteins to mediate long-range interactions between regulatory regions crucial for B-cell differentiation and perhaps altered in B-cell neoplasms is still unknown (Azagra et al., 2020). Even though, these findings represent one step ahead towards understanding why *SOX11* becomes overexpressed in MCL, but more studies are needed to ellucidate the precise mechanisms by which lineage-specific and architectural TFs may cooperate to induce SOX11 expression and the MCL phenotype.

In summary, three-dimensional chromatin organization studies are revolutionizing our understanding of epigenome regulation. I believe that comprehensive analyses of all the omic layers are bringing new insights onto molecular mechansims influencing immune function, essential to construct bridges between genome form and function. Even thought the future is uncertain, the advent of novel molecular techniques allowing to manipulate the genome as well as next-generation sequencing technologies, including single-cell analyses which permits to dissect cellular heterogeneity are leading to the new picture of biology. It seems that on the next years a great amount of data is going to be generated with the intention to critically understand the normal cell function and consequently its fatal alterations contribuiting the progression of cancer.

GENERAL CONCLUSIONS

In this doctoral thesis, four studies are presented in four different chapters, each one containing the specific conclusions. Here, however, I present the general conclusions that can be drawn from all four studies:

1. The generation of genome-wide maps of the 3D genome architecture integrated with other epigenomic marks during human B-cell differentiation as well as in chronic lymphocytic leukemia and mantle cell lymphoma provides a rich multi-omics resource for the biomedical community.

2. Beyond the conventional active and inactive 3D genome compartents, a novel categorization of the genome into three compartments reveals the presence of an intermediate compartment enriched with H3K27me3, a histone mark related to the polycomb repressive complex.

3. A 28% of the genome undergoes compartment shift during normal B cell differentiation, which is mostly characterized by widespread chromatin activation in germinal center B cells and a reversion in memory B cells to a 3D architecture similar to naive B cells.

4. CLL and MCL show entity and subtype-specific changes in their 3D genome architecture, which frequently span large blocks of the DNA and deregulate genes involved in the respective diseases such as inactivation of *EBF1* in CLL and activation of *SOX11* in MCL.

5. High-throughput chromosomal conformation capture (Hi-C) allows the accurate detection of chromosomal rearrangements in CLL and MCL, including some alterations not observed by conventional karyotyping and whole-genome sequencing.

6. An integrative analysis of DNA methylation, chromatin accessibility, histone modifications, transcriptomes and three-dimensional genome interactions represents a powerful strategy to identify potential epigenetic drivers at distant regulatory elements in cancer, as a superenhancer region looping to the *SOX11* gene in conventional MCL.

7. A detailed characterization of the distant *SOX11* superenhancer reveals a complex structure, including MCL-specific and lineage-related accessible sites. Binding of PAX5 at the MCL-specific accessible site suggest that this transcription factor may be associated in the early steps of SOX11 expression in conventional MCLs.

REFERENCES

Achinger-Kawecka, J., Taberlay, P.C., and Clark, S.J. (2016). Alterations in Three-Dimensional Organization of the Cancer Genome and Epigenome. Cold Spring Harb. Symp. Quant. Biol. *81*, 41–51.

Adams, D., Altucci, L., Antonarakis, S.E., et al. (2012). BLUEPRINT to decode the epigenetic signature written in blood. Nat. Biotechnol. *30*, 224–226.

Agathangelidis, A., Darzentas, N., Hadzidimitriou, A., et al. (2012). Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: A molecular classification with implications for targeted therapies. Blood *119*, 4467–4475.

Agathangelidis, A., Vardi, A., Baliakas, P., et al. (2014). Stereotyped B-cell receptors in chronic lymphocytic leukemia. Leuk. Lymphoma *55*, 2252–2261.

Agirre, X., Meydan, C., Jiang, Y., et al. (2019). Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. Nat. Commun. *10*, 821.

Aguilar, C.A., and Craighead, H.G. (2013). Micro- and nanoscale devices for the investigation of epigenetics and chromatin dynamics. Nat. Nanotechnol. *8*, 709–718.

Akashi, K., Traver, D., Miyamoto, T., et al. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193–197.

Akkaya, M., Kwak, K., and Pierce, S.K. (2019). B cell memory: building two walls of protection against pathogens. Nat. Rev. Immunol.

Albero, R., Campo, E., and Jares, P. (2018). Cyclin D1 overexpression induces global transcriptional downregulation in lymphoid neoplasms. J. Clin. Invest.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Almamun, M., Levinson, B.T., Gater, S.T., et al. (2014). Genome-wide DNA methylation analysis in precursor B-cells. Epigenetics *9*, 1588–1595.

Amano, T., Sagai, T., Tanabe, H., et al. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

Amin, N.A., Seymour, E., Saiya-Cork, K., et al. (2016). A quantitative analysis of subclonal and clonal gene mutations before and after therapy in chronic lymphocytic leukemia. Clin. Cancer Res. *22*, 4525–4535.

Anantharaman, A., Lin, I.-J., Barrow, J., et al. (2011). Role of Helix-Loop-Helix Proteins during Differentiation of Erythroid Cells. Mol. Cell. Biol. *31*, 1332–1343.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

Annunziato, A. (2008). DNA Packaging: Nucleosomes and ChromatinAnnunziato. Nat. Educ. *1*, 26.

Argatoff, L.H., Connors, J.M., Klasa, R.J., et al. (1997). Mantle cell lymphoma: A clinicopathologic study of 80 cases. Blood *89*, 2067–2078.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., et al. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Augusto, L., Nagai, E., Park, S., et al. (2019). Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphoma. *11*.

Azagra, A., Marina-Zárate, E., Ramiro, A.R., et al. (2020). From Loops to Looks: Transcription Factors and Chromatin Organization Shaping Terminal B Cell Differentiation. Trends Immunol. *41*, 46–60.

Babu, D., and Fullwood, M.J. (2015). 3D genome organization in health and disease: Emerging opportunities in cancer translational medicine. Nucleus *6*, 382–393.

Balsas, P., Palomero, J., Eguileor, Á., et al. (2017). SOX11 promotes tumor protective microenvironment interactions through CXCR4 and FAK regulation in mantle cell lymphoma. Blood *130*, 501–513.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barros, S.P., and Offenbacher, S. (2009). Epigenetics: Connecting Environment and Genotype to Phenotype and Disease. J. Dent. Res. *88*, 400–408.

Barski, A., Cuddapah, S., Cui, K., et al. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837.

Barutcu, A.R., Lajoie, B.R., McCord, R.P., et al. (2015). Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol. *16*.

Bastonini, E., Jeznach, M., Field, M., et al. (2014). Chromatin barcodes as biomarkers for melanoma. Pigment Cell Melanoma Res. *27*, 788–800.

Baù, D., and Marti-Renom, M.A. (2012). Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. Methods *58*, 300–306.

Bea, S., Valdes-Mas, R., Navarro, A., et al. (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc. Natl. Acad. Sci. *110*, 18250–18255.

Beà, S., Ribas, M., Hernández, J.M., et al. (1999). Increased number of chromosomal imbalances and high-level DNA amplifications in mantle cell lymphoma are associated with blastoid variants. Blood *93*, 4365–4374.

Beagan, J.A., Gilgenast, T.G., Kim, J., et al. (2016). Local genome topology can exhibit an incompletely rewired 3D-folding state during somatic cell reprogramming. Cell Stem Cell *18*, 611–624.

Beagrie, R.A., Scialdone, A., Schueler, M., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519–524.

Becker, P.B., and Hörz, W. (2002). ATP-Dependent Nucleosome Remodeling. Annu. Rev. Biochem. *71*, 247–273.

Beekman, R., Chapaprieta, V., Russiñol, N., et al. (2018a). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat. Med. *24*, 868–880.

Beekman, R., Amador, V., and Campo, E. (2018b). SOX11, a key oncogenic factor in mantle cell lymphoma. Curr. Opin. Hematol. *25*, 299–306.

Béguelin, W., Rivas, M.A., Calvo Fernández, M.T., et al. (2017). EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. Nat. Commun. *8*.

Belver, L., Yang, A.Y., Albero, R., et al. (2019). GATA3-controlled nucleosome eviction drives MYC enhancer activity in T-cell development and leukemia. Cancer Discov. *9*, 1774–1791.

Benaglia, T., Chauveau, D., Hunter, D.R., et al. (2009). mixtools : An R Package for Analyzing Finite Mixture Models. J. Stat. Softw. *32*.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Berger, S.L., Kouzarides, T., Shiekhattar, R., et al. (2009). An operational definition of epigenetics. Genes Dev. *23*, 781–783.

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., et al. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. J. Vis. Exp.

Bert, S.A., Robinson, M.D., Strbenac, D., et al. (2013). Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell *23*, 9–22.

Bhoi, S., Ljungström, V., Baliakas, P., et al. (2016). Prognostic impact of epigenetic classification in chronic lymphocytic leukemia: The case of subset #2. Epigenetics *11*, 449–455.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA Demethylation Dynamics. Cell *146*, 866–872.

Bibikova, M. (2016). DNA Methylation Microarrays. In Epigenomics in Health and Disease, (Elsevier Inc.), pp. 19–46.

Bibikova, M., Le, J., Barnes, B., et al. (2009). Genome-wide DNA methylation profiling using Infinium ® assay. Epigenomics *1*, 177–200.

Bibikova, M., Barnes, B., Tsan, C., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Binet, J.L., Auquier, A., Dighiero, G., et al. (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer *48*, 198–206.

Bintu, B., Mateo, L.J., Su, J.H., et al. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science (80-.). *362*.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

de Boer, C.J., van Krieken, J.H., Kluin-Nelemans, H.C., et al. (1995). Cyclin D1 messenger RNA overexpression as a marker for mantle cell lymphoma. Oncogene *10*, 1833–1840.

Boettiger, A.N., Bintu, B., Moffitt, J.R., et al. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Bonev, B., Mendelson Cohen, N., Szabo, Q., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572.e24.

Bosch, F., Jares, P., Campo, E., et al. (1994). PRAD-1/cyclin D1 gene overexpression in chronic lymphoproliferative disorders: a highly specific marker of mantle cell lymphoma. Blood *84*, 2726–2732.

Bosch, F., López-Guillermo, A., Campo, E., et al. (1998). Mantle cell lymphoma: Presenting features, response to therapy, and prognostic factors. Cancer *82*, 567–575.

Bouwman, B.A.M., and de Laat, W. (2015). Getting the genome in shape: the formation of loops, domains and compartments. Genome Biol. *16*, 154.

Boveri, T. (1909). Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität. Arch. Für Zellforsch. *3*, 181–268.

Branco, M.R., and Pombo, A. (2006). Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. PLoS Biol. *4*, e138.

Brecher, M., and Banks, P.M. (1990). Hodgkin's disease variant of Richter's syndrome. Report of eight cases. Am. J. Clin. Pathol. *93*, 333–339.

Brennan, D.J., Ek, S., Doyle, E., et al. (2009). The transcription factor Sox11 is a prognostic factor for improved recurrence-free survival in epithelial ovarian cancer. Eur. J. Cancer *45*, 1510–1517.

Brescia, P., Schneider, C., Holmes, A.B., et al. (2018). MEF2B Instructs Germinal Center Development and Acts as an Oncogene in B Cell Lymphomagenesis. Cancer Cell *34*, 453-465.e9.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., et al. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bunting, K.L., Soong, T.D., Singh, R., et al. (2016). Multi-tiered Reorganization of the Genome during B Cell Affinity Maturation Anchored by a Germinal Center-Specific Locus Control Region. Immunity *45*, 497–512.

Burton, J.N., Adey, A., Patwardhan, R.P., et al. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125.

Cahill, N., and Rosenquist, R. (2013). Uncovering the DNA methylome in chronic lymphocytic leukemia. Epigenetics *8*, 138–148.

Cahill, N., Bergh, A.C., Kanduri, M., et al. (2013). 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. Leukemia *27*, 150–158.

Callan, H.G. (1986). Lampbrush chromosomes. Mol. Biol. Biochem. Biophys. *36*, 1–252.

Campo, E., and Rule, S. (2015). Mantle cell lymphoma: Evolving management strategies. Blood *125*, 48–55.

Campo, E., Raffeld, M., and Jaffe, E.S. (1999). Mantle-cell lymphoma. Semin. Hematol. *36*, 115–127.

Cavalli, G. (2007). Chromosome kissing. Curr. Opin. Genet. Dev. *17*, 443–450.

Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. Nat. Rev. Immunol. *11*, 478–488.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*.

Chantepie, S.P., Vaur, D., Grunau, C., et al. (2010). ZAP-70 intron1 DNA methylation status: Determination by pyrosequencing in B chronic lymphocytic leukemia. Leuk. Res. *34*, 800–808.

Chao, M.P., Seita, J., and Weissman, I.L. (2008). Establishment of a normal hematopoietic and leukemia stem cell hierarchy. Cold Spring Harb. Symp. Quant. Biol. *73*, 439–449.

Chaplin, D.D. (2010). Overview of the immune response. J. Allergy Clin. Immunol. *125*.

Chen, J. (2011). Hematopoietic stem cell development, aging and functional failure. In International Journal of Hematology, pp. 3–10.

Chen, Y., Zhang, Y., Wang, Y., et al. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol. *217*, 4025–4048.

Chiorazzi, N., and Ferrarini, M. (2011). Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. Blood *117*, 1781–1791.

Chiorazzi, N., Rai, K.R., and Ferrarini, M. (2005). Chronic lymphocytic leukemia. N. Engl. J. Med. *352*.

Choi, J., Ko, M., Jeon, S., et al. (2012). The SWI/SNF-like BAF Complex Is Essential for Early B Cell Development. J. Immunol. *188*, 3791–3803.

Christensen, J.L., and Weissman, I.L. (2001). Flk-2 is a marker in hematopoietic stem cell differentiation: A simple method to isolate long-term stem cells. Proc. Natl. Acad. Sci. U. S. A. *98*, 14541–14546.

Claus, R., Lucas, D.M., Stilgenbauer, S., et al. (2012). Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. J. Clin. Oncol. *30*, 2483–2491.

Cobaleda, C., Schebesta, A., Delogu, A., et al. (2007). Pax5: The guardian of B cell identity and function. Nat. Immunol. *8*, 463–470.

Consortium, T.I.C.G. (2010). International network of cancer genome projects. Nature *464*, 993–998.

Corcoran, M., Parker, A., Orchard, J., et al. (2005). ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. Haematologica *90*, 1078–1088.

Cramer, P., and Hallek, M. (2011). Prognostic factors in chronic lymphocytic leukemia—what do we need to know? Nat. Rev. Clin. Oncol. *8*, 38–47.

Cremer, T., and Cremer, M. (2010). Chromosome Territories. Cold Spring Harb. Perspect. Biol. *2*, a003889–a003889.

Cremer, T., Cremer, C., Schneider, T., et al. (1982). Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. Hum. Genet. *62*, 201–209.

Cullen, K.E., Kladde, M.P., and Seyfred, M.A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. Science (80-.). *261*, 203–206.

Cuneo, A., Bigoni, R., Rigolin, G.M., et al. (1999). Cytogenetic profile of lymphoma of follicle mantle lineage: correlation with clinicobiologic features. Blood *93*, 1372–1380.

Cutter, A.R., and Hayes, J.J. (2015). A brief review of nucleosome structure. FEBS Lett. *589*, 2914–2922.

Dalla-Favera, R., Martinotti, S., Gallo, R.C., et al. (1983). Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. Science (80-.). *219*, 963–967.

Dallosso, A.R., Hancock, A.L., Szemes, M., et al. (2009). Frequent Long-Range Epigenetic Silencing of

Protocadherin Gene Clusters on Chromosome 5q31 in Wilms' Tumor. PLoS Genet. *5*, e1000745.

Damle, R.N., Wasil, T., Fais, F., et al. (1999). Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood *94*, 1840–1847.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Decaestecker, E., Gaba, S., Raeymaekers, J.A.M., et al. (2007). Host-parasite "Red Queen" dynamics archived in pond sediment. Nature *450*, 870–873.

Deichmann, U. (2016). Epigenetics: The origins and evolution of a fashionable topic. Dev. Biol. *416*, 249–254.

Dekker, J., Rippe, K., Dekker, M., et al. (2002a). Capturing chromosome conformation. Science (80-.). *295*, 1306–1311.

Dekker, J., Rippe, K., Dekker, M., et al. (2002b). Capturing chromosome conformation. Science (80-.). *295*, 1306–1311.

Delmore, J.E., Issa, G.C., Lemieux, M.E., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904–917.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. *30*, 1357–1382.

Dhalluin, C., Carlson, J.E., Zeng, L., et al. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.

Díaz, N., Kruse, K., Erdmann, T., et al. (2018). Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. Nat. Commun. *9*, 4938.

Dictor, M., Ek, S., Sundberg, M., et al. (2009). Strong lymphoid nuclear expression of SOX11 transcription factor defines lymphoblastic neoplasms, mantle cell lymphoma and Burkitt's lymphoma. Haematologica *94*, 1563–1568.

Le Dily, F., Baù, D., Pohl, A., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. *28*, 2151–2162.

Dixon, J.R., Selvaraj, S., Yue, F., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Döhner, H., Stilgenbauer, S., Benner, A., et al. (2000). Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia. N. Engl. J. Med. *343*, 1910–1916.

Donaldson-Collier, M.C., Sungalee, S., Zufferey, M., et al. (2019). EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. Nat. Genet. *51*, 517–528.

Dong, J., Panchakshari, R.A., Zhang, T., et al. (2015). Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature *525*, 134–139.

Dong, P., Tu, X., Chu, P.Y., et al. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. Mol. Plant *10*, 1497–1509.

Van Dongen, J.J.M., Van Der Velden, V.H.J., Brüggemann, M., et al. (2015). Minimal residual disease diagnostics in acute lymphoblastic leukemia: Need for sensitive, fast, and standardized technologies. Blood *125*, 3996–4009.

Doroshow, D.B., Eder, J.P., and LoRusso, P.M. (2017). BET inhibitors: A novel epigenetic approach. Ann. Oncol. *28*, 1776–1787.

Dostie, J., Richmond, T.A., Arnaout, R.A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Dowen, J.M., Fan, Z.P., Hnisz, D., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Dryden, N.H., Broome, L.R., Dudbridge, F., et al. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome Res. *24*, 1854–1868.

Dunham, I., Kundaje, A., Aldred, S.F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Duran-Ferrer, M., Beekman, R., and Martín-Subero, J.I. (2017). *In silico* deconvolution and purification of cancer epigenomes. Oncoscience *4*, 25–26.

Dy, P., Penzo-Méndez, A., Wang, H., et al. (2008). The three SoxC proteins - Sox4, Sox11 and Sox12 - Exhibit overlapping expression patterns and molecular properties. Nucleic Acids Res. *36*, 3101–3117.

Dykstra, B., and De Haan, G. (2008). Hematopoietic stem cell aging and self-renewal. Cell Tissue Res. *331*, 91–101.

Ecker, S., Chen, L., Pancaldi, V., et al. (2017). Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types. Genome Biol. *18*, 18.

Eckhardt, F., Lewin, J., Cortese, R., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. *38*, 1378–1385.

Ek, S., Dictor, M., Jerkeman, M., et al. (2008). Nuclear expression of the non-B-cell lineage Sox11 transcription factor identifies mantle cell lymphoma. Blood *111*, 800–805.

Eltsov, M., MacLellan, K.M., Maeshima, K., et al. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc. Natl. Acad. Sci. U. S. A. *105*, 19732–19737.

Enjuanes, A., Fernàndez, V., Hernández, L., et al. (2011). Identification of Methylated Genes Associated with Aggressive Clinicopathological Features in Mantle Cell Lymphoma. PLoS One *6*, e19736.

Enjuanes, A., Albero, R., Clot, G., et al. (2013). Genome-wide methylation analyses identify a subset of mantle cell lymphoma with a high number of methylated CpGs and aggressive clinicopathological features. Int. J. Cancer *133*, n/a-n/a.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Ernst, T., Chase, A.J., Score, J., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat. Genet. *42*, 722–726.

Escaramís, G., Docampo, E., and Rabionet, R. (2015). A decade of structural variants: Description, history and methods to detect structural variation. Brief. Funct. Genomics *14*, 305–314.

Fang, R., Yu, M., Li, G., et al. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Res. *26*, 1345–1348.

Farber, D.L., Yudanin, N.A., and Restifo, N.P. (2014). Human memory T cells: Generation, compartmentalization and homeostasis. Nat. Rev. Immunol. *14*, 24–35.

Fazi, C., Scarfò, L., Pecciarini, L., et al. (2011). General population low-count CLL-like MBL persists over time without clinical progression, although carrying the same cytogenetic abnormalities of CLL. Blood *118*, 6618–6625.

Fecteau, J.F., Côté, G., and Néron, S. (2006). A New Memory CD27 – IgG + B Cell Population in Peripheral Blood Expressing V H Genes with Low Frequency of Somatic Mutation . J. Immunol. *177*, 3728–3736.

Federico, C., Scavo, C., Cantarella, C.D., et al. (2006). Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. Chromosoma *115*, 123–128.

Feldmann, A., Ivanek, R., Murr, R., et al. (2013). Transcription Factor Occupancy Can Mediate Active

Turnover of DNA Methylation at Regulatory Regions. PLoS Genet. *9*, e1003994.

Fernandez, V., Salamero, O., Espinet, B., et al. (2010). Genomic and Gene Expression Profiling Defines Indolent Forms of Mantle Cell Lymphoma. Cancer Res. *70*, 1408–1418.

Fernàndez, V., Hartmann, E., Ott, G., et al. (2005). Pathogenesis of Mantle-Cell Lymphoma: All Oncogenic Roads Lead to Dysregulation of Cell Cycle and DNA Damage Response Pathways. J. Clin. Oncol. *23*, 6364–6369.

Ferraiuolo, M.A., Sanyal, A., Naumova, N., et al. (2012). From cells to chromatin: Capturing snapshots of genome organization with 5C technology. Methods *58*, 255–267.

Ferrando, A.A. (2013). SOX11 is a mantle cell lymphoma oncogene. Blood *121*, 2169–2170.

Finlan, L.E., Sproul, D., Thomson, I., et al. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. *4*.

Flavahan, W.A., Drier, Y., Liau, B.B., et al. (2015). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Foster, H.A., Abeydeera, L.R., Griffin, D.K., et al. (2005). Non-random chromosome positioning in mammalian sperm nuclei, with migration of the sex chromosomes during late spermatogenesis. J. Cell Sci. *118*, 1811–1820.

Franke, M., Ibrahim, D.M., Andrey, G., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.

Fraser, J., Williamson, I., Bickmore, W.A., et al. (2015a). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol. Mol. Biol. Rev. *79*, 347–372.

Fraser, J., Ferrai, C., Chiariello, A.M., et al. (2015b). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol. Syst. Biol. *11*, 852.

Frigola, J., Song, J., Stirzaker, C., et al. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. *38*, 540–549.

Fu, K. (2005). Cyclin D1-negative mantle cell lymphoma: a clinicopathologic study based on gene expression profiling. Blood *106*, 4315–4321.

Fullwood, M.J., Liu, M.H., Pan, Y.F., et al. (2009). An oestrogen-receptor-α-bound human chromatin interactome. Nature *462*, 58–64.

Furtado, M., and Rule, S. (2011). Indolent mantle cell lymphoma. Haematologica *96*, 1086–1088.

Fuster, C., Martín-Garcia, D., Balagué, O., et al. (2019). Cryptic insertions of the immunoglobulin light chain enhancer region near CCND1 in t(11;14)-negative mantle cell lymphoma. Haematologica haematol.2019.237073.

Fuxa, M., and Busslinger, M. (2007a). Reporter Gene Insertions Reveal a Strictly B Lymphoid-Specific Expression Pattern of Pax5 in Support of Its B Cell Identity Function . J. Immunol. *178*, 3031–3037.

Fuxa, M., and Busslinger, M. (2007b). Reporter gene insertions reveal a strictly B lymphoid-specific expression pattern of Pax5 in support of its B cell identity function. J. Immunol. *178*, 8221.2-8221.

Fyodorov, D. V., Zhou, B.-R., Skoultchi, A.I., et al. (2018). Emerging roles of linker histones in regulating chromatin structure and function. Nat. Rev. Mol. Cell Biol. *19*, 192–206.

Gabrea, A., Bergsagel, P.L., Chesi, M., et al. (1999). Insertion of Excised IgH Switch Sequences Causes Overexpression of Cyclin D1 in a Myeloma Tumor Cell. Mol. Cell *3*, 119–123.

Gaidano, G., Foà, R., and Dalla-Favera, R. (2012). Molecular pathogenesis of chronic lymphocytic leukemia. J. Clin. Invest. *122*, 3432–3438.

Gaiti, F., Chaligne, R., Gu, H., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature *569*, 576–580.

Gall, J.G. (1956). On the submicroscopic structure of chromosomes. Brookhaven Symp. Biol. 17–32.

Gangaraju, V.K., and Bartholomew, B. (2007). Mechanisms of ATP dependent chromatin remodeling.

Mutat. Res. Mol. Mech. Mutagen. *618*, 3–17.

Gao, H., Lukin, K., Ramírez, J., et al. (2009). Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. Proc. Natl. Acad. Sci. U. S. A. *106*, 11258–11263.

Gelmann, E.P., Psallidopoulos, M.C., Papas, T.S., et al. (1983). Identification of reciprocal translocation sites within the c-myc oncogene and immunoglobulin µ locus in a Burkitt lymphoma. Nature *306*, 799–803.

Ghia, P., Guida, G., Stella, S., et al. (2003). The pattern of CD38 expression defines a distinct subset of chronic lymphocytic leukemia (CLL) patients at risk of disease progression. Blood *101*, 1262–1269.

Giorgio, E., Robyr, D., Spielmann, M., et al. (2015). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). Hum. Mol. Genet. *24*, 3143–3154.

Gnyszka, A., Jastrzebski, Z., and Flis, S. (2013). DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer Res. *33*, 2989–2996.

Goetze, S., Mateos-Langerak, J., Gierman, H.J., et al. (2007). The Three-Dimensional Structure of Human Interphase Chromosomes Is Related to the Transcriptome Map. Mol. Cell. Biol. *27*, 4475–4487.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Grasser, F., Neusser, M., Fiegler, H., et al. (2008). Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. J. Cell Sci. *121*, 1876–1886.

Greiner, T., Moynihan, M., Chan, W., et al. (1996). p53 mutations in mantle cell lymphoma are associated with variant cytology and predict a poor prognosis. Blood *87*, 4302–4310.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., et al. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell *157*, 369–381.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. Nature *389*, 349–352.

Guo, M., Price, M.J., Patterson, D.G., et al. (2018). EZH2 Represses the B Cell Transcriptional Program and Regulates Antibody-Secreting Cell Metabolism and Antibody Production. J. Immunol. *200*, 1039–1052.

Gutierrez, A., Tschumper, R.C., Wu, X., et al. (2010). LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. Blood *116*, 2975–2983.

Hadzidimitriou, A., Agathangelidis, A., Darzentas, N., et al. (2011). Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. Blood *118*, 3088–3095.

Haferlach, C., Dicker, F., Schnittger, S., et al. (2007). Comprehensive genetic characterization of CLL: A study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgVH status and immunophenotyping. Leukemia *21*, 2442–2451.

Hagege, H., Klous, P., Braem, C., et al. (2007). Quantitative analysis of chromosome conformation capture assays (3c-qpcr). Nat. Protoc. *2*, 1722–1733.

Hagman, J., and Lukin, K. (2006). Transcription factors drive B cell development. Curr. Opin. Immunol. *18*, 127–134.

Halldórsdóttir, A.M., Kanduri, M., Marincevic, M., et al. (2012). Mantle cell lymphoma displays a homogenous methylation profile: A comparative analysis with chronic lymphocytic leukemia. Am. J. Hematol. *87*, 361–367.

Hallek, M., Wanders, L., Strohmeyer, S., et al. (1992). Thymidine kinase: a tumor marker with prognostic value for non-Hodgkin's lymphoma and a broad range of potential clinical applications. Ann. Hematol. *65*, 1–5.

Hallek, M., Cheson, B.D., Catovsky, D., 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 *111*, 5446–5456.

Hamblin, T.J., Davis, Z., Gardiner, A., et al. (1999). Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood *94*, 1848–1854.

Hamblin, T.J., Orchard, J.A., Ibbotson, R.E., et al. (2002). CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. Blood *99*, 1023–1029.

Hansen, A.S., Cattoglio, C., Darzacq, X., et al. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20–32.

Harbour, J.W., and Dean, D.C. (2000). The Rb/E2F pathway: Expanding roles and emerging paradigms. Genes Dev. *14*, 2393–2409.

Harewood, L., Schütz, F., Boyle, S., et al. (2010). The effect of translocation-induced nuclear reorganization on gene expression. Genome Res. *20*, 554–564.

Harewood, L., Kishore, K., Eldridge, M.D., et al. (2017). Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol. *18*, 125.

Heintzman, N.D., Stuart, R.K., Hon, G., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heinz, S., Benner, C., Spann, N., et al. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hepperger, C., Mannes, A., Merz, J., et al. (2008). Three-dimensional positioning of genes in mouse cell nuclei. Chromosoma *117*, 535–551.

Hergeth, S.P., and Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep. *16*, 1439–1453.

Hernandez, L., Fest, T., Cazorla, M., et al. (1996). p53 gene mutations and protein overexpression are associated with aggressive variants of mantle cell lymphomas. Blood *87*, 3351–3359.

Heyn, H., and Esteller, M. (2012). DNA methylation profiling in the clinic: applications and challenges. Nat. Rev. Genet. *13*, 679–692.

Hide, T., Takezaki, T., Nakatani, Y., et al. (2009). Sox11 prevents tumorigenesis of glioma-initiating cells by inducing neuronal differentiation. Cancer Res. *69*, 7953–7959.

Hirt, C., Schüler, F., Dölken, L., et al. (2004). Low prevalence of circulating t(11;14)(q13;q32)–positive cells in the peripheral blood of healthy individuals as detected by real-time quantitative PCR. Blood *104*, 904–905.

Hitchins, M.P., Lin, V.A., Buckle, A., et al. (2007). Epigenetic inactivation of a cluster of genes flanking MLH1 in microsatellite-unstable colorectal cancer. Cancer Res. *67*, 9107–9116.

Hnisz, D., Weintraub, A.S., Day, D.S., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science (80-.). *351*, 1454–1458.

Van Holde, K.E. (1988). Chromatin: Springer Series in Molecular Biology (New York: Springer-Verlag).

Holliday, R. (2006). Epigenetics: A Historical Overview. Epigenetics *1*, 76–80.

Holliday, R., and Pugh, J. (1975). DNA modification mechanisms and gene activity during development. Science (80-.). *187*, 226–232.

Houseman, E.A., Accomando, W.P., Koestler, D.C., et al. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*.

Hu, G., Cui, K., Fang, D., et al. (2018). Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. Immunity *48*, 227-242.e8.

Hu, J., Zhang, Y., Zhao, L., et al. (2015). Chromosomal Loop Domains Direct the Recombination of Antigen

Receptor Genes. Cell *163*, 947–959.

Huang, Y., Pastor, W.A., Shen, Y., et al. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. PLoS One *5*, e8888.

Huebert, D.J., Kamal, M., O'Donovan, A., et al. (2006). Genome-wide analysis of histone modifications by ChIP-on-chip. Methods *40*, 365–369.

Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Imakaev, M., Fudenberg, G., McCord, R.P., et al. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Izban, K.F., Alkan, S., Singleton, T.P., et al. (2000). Multiparametric immunohistochemical analysis of the cell cycle proteins cyclin D1, Ki-67, p21(WAF1), p27(KIP1), and p53 in mantle cell lymphoma. Arch. Pathol. Lab. Med. *124*, 1457–1462.

Jaffe, A.E., and Irizarry, R.A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. *15*.

Janssen, A., Colmenares, S.U., and Karpen, G.H. (2018). Heterochromatin: Guardian of the Genome. Annu. Rev. Cell Dev. Biol. *34*, 265–288.

Jares, P., and Campo, E. (2008). Advances in the understanding of mantle cell lymphoma. Br. J. Haematol. *142*, 149–165.

Jares, P., Colomer, D., and Campo, E. (2007). Genetic and molecular pathogenesis of mantle cell lymphoma: Perspectives for new targeted therapeutics. Nat. Rev. Cancer *7*, 750–762.

Jares, P., Colomer, D., Campo, E., et al. (2012a). Molecular pathogenesis of mantle cell lymphoma. *122*, 3416–3423.

Jares, P., Colomer, D., and Campo, E. (2012b). Molecular pathogenesis of mantle cell lymphoma. J. Clin. Invest. *122*, 3416–3423.

Jares, P., Colomer, D., and Campo, E. (2012c). Molecular pathogenesis of mantle cell lymphoma. J. Clin. Invest. *122*, 3416–3423.

Javierre, B.M., Burren, O.S., Wilder, S.P., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369-1384.e19.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (80-.). *293*, 1074–1080.

Jethwa, A., Hüllein, J., Stolz, T., et al. (2013). Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. Br. J. Haematol. *163*, 496–500.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., et al. (2008). The 3D Structure of the Immunoglobulin Heavy-Chain Locus: Implications for Long-Range Genomic Interactions. Cell *133*, 265–279.

Jiang, X.X., Nguyen, Q., Chou, Y.C., et al. (2011). Control of B Cell Development by the Histone H2A Deubiquitinase MYSM1. Immunity *35*, 883–896.

Jin, F., Li, Y., Dixon, J.R., et al. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Johanson, T.M., Lun, A.T.L., Coughlan, H.D., et al. (2018). Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. Nat. Immunol. *19*, 1257–1264.

Johanson, T.M., Chan, W.F., Keenan, C.R., et al. (2019). Genome organization in immune cells: unique challenges. Nat. Rev. Immunol. *19*, 448–456.

Jongen-Lavrencic, M., Grob, T., Hanekamp, D., et al. (2018). Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N. Engl. J. Med. *378*, 1189–1199.

Kantidze, O.L., Luzhin, A. V., Nizovtseva, E. V., et al. (2019). The anti-cancer drugs curaxins target spatial genome organization. Nat. Commun. *10*, 1–11.

Karolchik, D., Hinrichs, A., Furey, T., et al. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids

Res. *32*, 493D – 496.

Kasar, S., Kim, J., Improgo, R., et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat. Commun. *6*.

Kavyanifar, A., Turan, S., and Lie, D.C. (2018). SoxC transcription factors: multifunctional regulators of neurodevelopment. Cell Tissue Res. *371*, 91–103.

Kemeny, S., Tatout, C., Salaun, G., et al. (2018). Spatial organization of chromosome territories in the interphase nucleus of trisomy 21 cells. Chromosoma *127*, 247–259.

Kenter, A.L., and Feeney, A.J. (2019). New insights emerge as antibody repertoire diversification meets chromosome conformation. F1000Research *8*, 347.

Khan, A., Fornes, O., Stigliani, A., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*, D260–D266.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507–1520.

Kipps, T.J., Stevenson, F.K., Wu, C.J., et al. (2017). Chronic lymphocytic leukaemia. Nat. Rev. Dis. Prim. *3*, 16096.

Klein, U., Tu, Y., Stolovitzky, G.A., et al. (2001). Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J. Exp. Med. *194*, 1625–1638.

Klein, U., Tu, Y., Stolovitzky, G.A., et al. (2003). Transcriptional analysis of the B cell germinal center reaction. Proc. Natl. Acad. Sci. *100*, 2639–2644.

Klein, U., Lia, M., Crespo, M., et al. (2010). The DLEU2/miR-15a/16-1 Cluster Controls B Cell Proliferation and Its Deletion Leads to Chronic Lymphocytic Leukemia. Cancer Cell *17*, 28–40.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Kloetgen, A., Thandapani, P., Ntziachristos, P., et al. (2020). Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat. Genet. *52*, 388–400.

Klopocki, E., Ott, C.E., Benatar, N., et al. (2008). A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. J. Med. Genet. *45*, 370–375.

Kolar, G.R., Mehta, D., Pelayo, R., et al. (2007). A novel human B cell subpopulation representing the initial germinal center population to express AID. Blood *109*, 2545–2552.

Korbel, J.O., and Campbell, P.J. (2013). Criteria for inference of chromothripsis in cancer genomes. Cell *152*, 1226–1236.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Kridel, R., Meissner, B., Rogic, S., et al. (2012). Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood *119*, 1963–1971.

Krijger, P.H.L., Di Stefano, B., De Wit, E., et al. (2016). Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. Cell Stem Cell *18*, 597–610.

Kulinski, J.M., Tarakanova, V.L., and Verbsky, J. (2013). Regulation of Antiviral CD8 T-Cell Responses.

Kulis, M., Heath, S., Bibikova, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. *44*, 1236–1242.

Kulis, M., Merkel, A., Heath, S., et al. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat. Genet. *47*, 746–756.

Kumar, B. V., Connors, T.J., and Farber, D.L. (2018). Human T Cell Development, Localization, and Function throughout Life. Immunity *48*, 202–213.

Kumari, G., and Sen, R. (2015). Chromatin Interactions in the Control of Immunoglobulin Heavy Chain Gene Assembly. In Advances in Immunology, (Academic Press Inc.), pp. 41–92.

Kundu, S., Ji, F., Sunwoo, H., et al. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted

Domains that Change during Differentiation. Mol. Cell *65*, 432-446.e5.

Kuo, P.Y., Leshchenko, V. V., Fazzari, M.J., et al. (2015). High-resolution chromatin immunoprecipitation (ChIP) sequencing reveals novel binding targets and prognostic role for SOX11 in mantle cell lymphoma. Oncogene *34*, 1231–1240.

Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. Nat. Rev. Cancer *5*, 251–262.

Küppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene *20*, 5580–5594.

Kurosaki, T., Shinohara, H., and Baba, Y. (2010). B Cell Signaling and Fate Decision. Annu. Rev. Immunol. *28*, 21–55.

Laird, P.W. (2010). Principles and challenges of genome-wide DNA methylation analysis. Nat. Rev. Genet. *11*, 191–203.

Lakadamyali, M., and Cosma, M.P. (2015). Advanced microscopy methods for visualizing chromatin structure. FEBS Lett. *589*, 3023–3030.

Landau, D.A., Carter, S.L., Stojanov, P., et al. (2013). Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. Cell *152*, 714–726.

Landau, D.A., Carter, S.L., Getz, G., et al. (2014). Clonal evolution in hematological malignancies and therapeutic implications. Leukemia *28*, 34–43.

Landau, D.A., Tausch, E., Taylor-Weiner, A.N., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. Nature *526*, 525–530.

Lander, E.S., Linton, L.M., Birren, B., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., et al. (2014). Chromatin state dynamics during blood formation. Science (80-.). *345*, 943–949.

Lardelli, P., Bookman, M.A., Sundeen, J., et al. (1990). Lymphocytic Lymphoma of Intermediate Differentiation. Am. J. Surg. Pathol. *14*, 752–763.

Larson, A.G., Elnatan, D., Keenen, M.M., et al. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Lawrence, M., Huber, W., Pagès, H., et al. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*, e1003118.

Lecluse, Y., Lebailly, P., Roulland, S., et al. (2009). t(11;14)-positive clones can persist over a long period of time in the peripheral blood of healthy individuals. Leukemia *23*, 1190–1193.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lee, C.-J., Appleby, V.J., Orme, A.T., et al. (2002). Differential expression of SOX4 and SOX11 in medulloblastoma. J. Neurooncol. *57*, 201–214.

Lee, D.S., Luo, C., Zhou, J., et al. (2019). Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat. Methods *16*, 999–1006.

Lee, E.-J., Luo, J., Wilson, J.M., et al. (2013). Analyzing the cancer methylome through targeted bisulfite sequencing. Cancer Lett. *340*, 171–178.

Lee, S.-T., Xiao, Y., Muench, M.O., et al. (2012). A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. Nucleic Acids Res. *40*, 11339–11351.

Leshchenko, V. V., Kuo, P.Y., Shaknovich, R., et al. (2010). Genomewide DNA methylation analysis reveals novel targets for drug development in mantle cell lymphoma. Blood *116*, 1025–1034.

Lettice, L.A., Daniels, S., Sweeney, E., et al. (2011). Enhancer-adoption as a mechanism of human developmental disease. Hum. Mutat. *32*, 1492–1499.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707–719.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, R., Liu, Y., Hou, Y., et al. (2018). 3D genome and its disorganization in diseases. Cell Biol. Toxicol. *34*, 351–365.

Liang, Z., Li, G., Wang, Z., et al. (2017). BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. Nat. Commun. *8*.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). *326*, 289–293.

Limpens, J., de Jong, D., van Krieken, J.H., et al. (1991). Bcl-2/JH rearrangements in benign lymphoid tissues with follicular hyperplasia. Oncogene *6*, 2271–2276.

Lin, Y.C., Benner, C., Mansson, R., et al. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. *13*, 1196–1204.

Lister, R., Pelizzola, M., Dowen, R.H., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Locke, W.J., Guanzon, D., Ma, C., et al. (2019). DNA Methylation Cancer Biomarkers: Translation to the Clinic. Front. Genet. *10*, 1150.

Lohan, S., Spielmann, M., Doelken, S.C., et al. (2014). Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. Clin. Genet. *86*, 318–325.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lovec, H., Grzeschiczek, A., Kowalski, M.B., et al. (1994). Cyclin D1/bcl-1 cooperates with myc genes in the generation of B-cell lymphoma in transgenic mice. EMBO J. *13*, 3487–3495.

Lovén, J., Hoke, H.A., Lin, C.Y., et al. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell *153*, 320–334.

Luc, S., Buza-Vidas, N., and Jacobsen, S.E.W. (2008). Delineating the cellular pathways of hematopoietic lineage commitment. Semin. Immunol. *20*, 213–220.

Lucas, J.S., Zhang, Y., Dudko, O.K., et al. (2014). 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. Cell *158*, 339–352.

Lupiáñez, D.G., Kraft, K., Heinrich, V., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell *161*, 1012–1025.

Lyu, H., Liu, E., and Wu, Z. (2019). Comparison of normalization methods for Hi-C data. Biotechniques btn-2019-0105.

Ma, W., Ay, F., Lee, C., et al. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat. Methods *12*, 71–78.

Maass, P.G., Barutcu, A.R., and Rinn, J.L. (2019). Interchromosomal interactions: A genomic love story of kissing chromosomes. J. Cell Biol. *218*, 27–38.

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol. *13*, R44.

Malek, S.N. (2013). The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. Oncogene *32*, 2805–2817.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat. Struct. Mol. Biol. *10*, 882–891.

Mallm, J., Iskar, M., Ishaque, N., et al. (2019). Linking aberrant chromatin features in chronic lymphocytic leukemia to deregulated transcription factor networks. 1–44.

Mao, Z., Quintanilla-Martinez, L., Raffeld, M., et al. (2007). IgVH mutational status and clonality analysis of Richter's transformation: Diffuse large B-cell lymphoma and hodgkin lymphoma in association with B-cell chronic lymphocytic leukemia (B-CLL) represent 2 different pathways of disease evolution. Am. J. Surg. Pathol. *31*, 1605–1614.

Marco-Sola, S., Sammeth, M., Guigó, R., et al. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods *9*, 1185–1188.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Marti, G.E., Rawstron, A.C., Ghia, P., et al. (2005). Diagnostic criteria for monoclonal B-cell lymphocytosis. Br. J. Haematol. *130*, 325–332.

Martín-Garcia, D., Navarro, A., Valdés-Mas, R., et al. (2019). CCND2 and CCND3 hijack immunoglobulin light-chain enhancers in cyclin D1– mantle cell lymphoma. Blood *133*, 940–951.

Martin-Subero, J.I., and Oakes, C.C. (2018). Charting the dynamic epigenome during B-cell development. Semin. Cancer Biol. *51*, 139–148.

Martín-Subero, J.I. (2011). How Epigenomics Brings Phenothype into Being. Pediatr. Endocrinol. Rev. *9*, 506–510.

Martin, P., McGovern, A., Orozco, G., et al. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat. Commun. *6*, 1–7.

Mas, G., Blanco, E., Ballaré, C., et al. (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nat. Genet. *50*, 1452–1462.

Mateo, L.J., Murphy, S.E., Hafner, A., et al. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature *568*, 49–54.

Matthias, P., and Rolink, A.G. (2005). Transcriptional networks in developing and mature B cells. Nat. Rev. Immunol. *5*, 497–508.

Matutes, E., Oscier, D., Garcia-Marco, J., et al. (1996). Trisomy 12 defines a group of CLL with atypical morphology: Correlation between cytogenetic, clinical and laboratory features in 544 patients. Br. J. Haematol. *92*, 382–388.

Maurano, M.T., Wang, H., John, S., et al. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195.

Mayer, R., Brero, A., von Hase, J., et al. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. BMC Cell Biol. *6*.

McCall, M.N., Bolstad, B.M., and Irizarry, R.A. (2010). Frozen robust multiarray analysis (fRMA). Biostatistics *11*, 242–253.

McHeyzer-Williams, M., Okitsu, S., Wang, N., et al. (2012). Molecular programming of B cell memory. Nat. Rev. Immunol. *12*, 24–34.

McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics *11*, 165.

McManus, S., Ebert, A., Salvagiotto, G., et al. (2011). The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J. *30*, 2388–2404.

McNairn, A.J., and Gerton, J.L. (2008). Cohesinopathies: One ring, many obligations. Mutat. Res. - Fundam. Mol. Mech. Mutagen. *647*, 103–111.

Meaburn, K.J., and Misteli, T. (2007). Cell biology: Chromosome territories. Nature *445*, 379–781.

Medina, K.L., Pongubala, J.M.R., Reddy, K.L., et al. (2004). Assembling a gene regulatory network for specification of the B cell fate. Dev. Cell *7*, 607–617.

Mehra, P., and Kalani, A. (2018). What's in the "fold"? Life Sci. *211*, 118–125.

Mehta, I.S., Kulashreshtha, M., Chakraborty, S., et al. (2013). Chromosome territories reposition during DNA damage-repair response. Genome Biol. *14*.

Melchers, F. (2015). Checkpoints that control B cell development. J. Clin. Invest. *125*, 2203–2210.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet. *17*, 17–43.

Mieczkowski, J., Cook, A., Bowman, S.K., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*, 11485.

Migliazza, A., Martinotti, S., Chen, W., et al. (1995). Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. *92*, 12520–12524.

Mikeska, T., and Craig, J. (2014). DNA Methylation Biomarkers: Cancer and Beyond. Genes (Basel). *5*, 821–864.

Mockridge, C.I., Potter, K.N., Wheatley, I., et al. (2013). Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by V H -gene mutational status Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by V H -gene mutational status. *109*, 4424–4431.

Mohammad, H.P., and Baylin, S.B. (2010). Linking cell signaling and the epigenetic machinery. Nat. Biotechnol. *28*, 1033–1038.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. Science (80-.). *211*, 393–396.

Mohanty, A., Sandoval, N., Das, M., et al. (2016). CCND1 mutations increase protein stability and promote ibrutinib resistance in mantle cell lymphoma. Oncotarget *7*, 73558–73572.

Mohanty, A., Sandoval, N., Phan, A., et al. (2019). Regulation of SOX11 expression through CCND1 and STAT3 in mantle cell lymphoma. Blood *133*, 306–318.

Montefiori, L., Wuerffel, R., Roqueiro, D., et al. (2016). Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. Cell Rep. *14*, 896–906.

Montserrat, E. (2006). New prognostic markers in CLL. Hematology Am. Soc. Hematol. Educ. Program 279–284.

Morin, R.D., Johnson, N.A., Severson, T.M., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat. Genet. *42*, 181–185.

Moros, A., Rodríguez, V., Saborit-Villarroya, I., et al. (2014). Synergistic antitumor activity of lenalidomide with the BET bromodomain inhibitor CPI203 in bortezomib-resistant mantle cell lymphoma. Leukemia *28*, 2049–2059.

Morrison, S.J., and Weissman, I.L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity *1*, 661–673.

Morton, L.M., Turner, J.J., Cerhan, J.R., et al. (2007). Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). Blood *110*, 695–708.

Mozos, A., Royo, C., Hartmann, E., et al. (2009). SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. Haematologica *94*, 1555–1562.

Mukhopadhyay, S., Ramadass, A.S., Akoulitchev, A., et al. (2014). Formation of distinct chromatin conformation signatures epigenetically regulate macrophage activation. Int. Immunopharmacol. *18*, 7–11.

Müller, J.R., Janz, S., Goedert, J.J., et al. (1995). Persistence of immunoglobulin heavy chain/c-myc

recombination-positive lymphocyte clones in the blood of human immunodeficiency virus-infected homosexual men. Proc. Natl. Acad. Sci. U. S. A. *92*, 6577–6581.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., et al. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. *36*, 889–893.

Muzio, M., Apollonio, B., Scielzo, C., et al. (2008). Constitutive activation of distinct BCR-signaling pathways in a subset of CLL patients: a molecular signature of anergy. Blood *112*, 188–195.

Nadeu, F., Delgado, J., Royo, C., et al. (2016). Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. Blood *127*, 2122–2130.

Nadeu, F., Clot, G., Delgado, J., et al. (2018). Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. Leukemia *32*, 645–653.

Nadeu, F., Martin-Garcia, D., Clot, G., et al. (2020). Genomic and epigenomic insights into the origin, pathogenesis and clinical behavior of mantle cell lymphoma subtypes. Blood.

Nagano, T., Lubling, Y., Stevens, T.J., et al. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Nakashima, M.O., Durkin, L., Bodo, J., et al. (2014). Utility and diagnostic pitfalls of SOX11 monoclonal antibodies in mantle cell lymphoma and other lymphoproliferative disorders. Appl. Immunohistochem. Mol. Morphol. *22*, 720–727.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: Its Roles and Mechanisms. Annu. Rev. Genet. *43*, 525–558.

Nativio, R., Wendt, K.S., Ito, Y., et al. (2009). Cohesin Is Required for Higher-Order Chromatin Conformation at the Imprinted IGF2-H19 Locus. PLoS Genet. *5*, e1000739.

Nativio, R., Sparago, A., Ito, Y., et al. (2011). Disruption of genomic neighbourhood at the imprinted IGF2-H19 locus in Beckwith–Wiedemann syndrome and Silver–Russell syndrome. Hum. Mol. Genet. *20*, 1363–1374.

Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. Immunity *33*, 12–24.

Navarro, A., Clot, G., Royo, C., et al. (2012). Molecular Subsets of Mantle Cell Lymphoma Defined by the IGHV Mutational Status and SOX11 Expression Have Distinct Biologic and Clinical Features. Cancer Res. *72*, 5307–5316.

Navarro, A., Clot, G., Martínez-Trillos, A., et al. (2017). Improved classification of leukemic B-cell lymphoproliferative disorders using a transcriptional and genetic classifier. Haematologica *102*, e360–e363.

Neph, S., Kuehn, M.S., Reynolds, A.P., et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics *28*, 1919–1920.

Nir, G., Farabella, I., Pérez Estrada, C., et al. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLOS Genet. *14*, e1007872.

Nora, E.P., Lajoie, B.R., Schulz, E.G., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.L., et al. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22.

Nordström, L., Sernbo, S., Eden, P., et al. (2014). SOX11 and TP53 add prognostic information to MIPI in a homogenously treated cohort of mantle cell lymphoma - a Nordic Lymphoma Group study. Br. J.

Haematol. *166*, 98–108.

Northcott, P.A., Buchhalter, I., Morrissy, A.S., et al. (2017). The whole-genome landscape of medulloblastoma subtypes. Nature *547*, 311–317.

Norton, A.J., Matthews, J., Pappa, V., et al. (1995). Mantle cell lymphoma: Natural history defined in a serially biopsied population over a 20-year period. Ann. Oncol. *6*, 249–256.

Novak, P., Jensen, T., Oshiro, M.M., et al. (2008). Agglomerative epigenetic aberrations are a common event in human breast cancer. Cancer Res. *68*, 8616–8625.

Nygren, L., Wennerholm, S.B., Klimkowska, M., et al. (2012). Prognostic role of SOX11 in a population-based cohort of mantle cell lymphoma. Blood *119*, 4215–4223.

Oakes, C.C., and Martin-Subero, J.I. (2018). Insight into origins, mechanisms, and utility of DNA methylation in B-cell malignancies. Blood *132*, 999–1006.

Oakes, C.C., Claus, R., Gu, L., et al. (2014). Evolution of DNA Methylation Is Linked to Genetic Aberrations in Chronic Lymphocytic Leukemia. Cancer Discov. *4*, 348–361.

Oakes, C.C., Seifert, M., Assenov, Y., et al. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat. Genet. *48*, 253–264.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. *17*, 520–527.

Ojha, J., Ayres, J., Secreto, C., et al. (2015). Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. Blood *125*, 492–498.

Olins, A.L., and Olins, D.E. (1974). Spheroid Chromatin Units (ngr Bodies). Science (80-.). *183*, 330–332.

Orchard, J., Garand, R., Davis, Z., et al. (2003). A subset of t(11;14) lymphoma with mantle cell features displays mutated IgVH genes and includes patients with good prognosis, nonnodal disease. Blood *101*, 4975–4981.

Ordulu, Z., Kammin, T., Brand, H., et al. (2016). Structural Chromosomal Rearrangements Require Nucleotide-Level Resolution: Lessons from Next-Generation Sequencing in Prenatal Diagnosis. Am. J. Hum. Genet. *99*, 1015–1033.

Orkin, S.H., and Zon, L.I. (2008). SnapShot: Hematopoiesis. Cell *132*, 712.e1-712.e2.

Ou, H.D., Phan, S., Deerinck, T.J., et al. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science (80-.). *357*.

Palomero, J., Vegliante, M.C., Rodr, M.L., et al. (2015). SOX11 promotes tumor angiogenesis through transcriptional regulation of PDGFA in mantle cell lymphoma. Am. Soc. Hematol. *124*, 2235–2248.

Palomero, J., Vegliante, M.C., Eguileor, A., et al. (2016). SOX11 defines two different subtypes of mantle cell lymphoma through transcriptional regulation of BCL6. Leukemia *30*, 1596–1599.

Parada, L.A., Sotiriou, S., and Misteli, T. (2004). Spatial genome organization. Exp. Cell Res. *296*, 64–70.

Parkin, J., and Cohen, B. (2001). An overview of the immune system. Lancet *357*, 1777–1789.

Pasqualucci, L., Neumeister, P., Goossens, T., et al. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341–346.

Penzo-Méndez, A.I. (2010). Critical roles for SoxC transcription factors in development and cancer. Int. J. Biochem. Cell Biol. *42*, 425–428.

Peric-Hupkes, D., Meuleman, W., Pagie, L., et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol. Cell *38*, 603–613.

Pestell, R.G. (2013). New roles of cyclin D1. Am. J. Pathol. *183*, 3–9.

Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev. *22*, 3089–3114.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Phillips, J.E., and Corces, V.G. (2009). CTCF: Master Weaver of the Genome. Cell *137*, 1194–1211.

Pieper, K., Grimbacher, B., and Eibel, H. (2013). Mechanisms of allergic diseases B-cell biology and development. J. Allergy Clin. Immunol. *131*, 959–971.

Pinyol, M., Bea, S., Plà, L., et al. (2007). Inactivation of RB1 in mantle-cell lymphoma detected by nonsense-mediated mRNA decay pathway inhibition and microarray analysis. Blood *109*, 5422–5429.

Del Poeta, G., Maurillo, L., Venditti, A., et al. (2001). Clinical significance of CD38 expression in chronic lymphocytic leukemia. Blood *98*, 2633–2639.

Pope, B.D., Ryba, T., Dileep, V., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Proudhon, C., Hao, B., Raviram, R., et al. (2015). Long-Range Regulation of V(D)J Recombination. In Advances in Immunology, (Academic Press Inc.), pp. 123–182.

Puente, X.S., Pinyol, M., Quesada, V., et al. (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature *475*, 101–105.

Puente, X.S., Beà, S., Valdés-Mas, R., et al. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature *526*, 519–524.

Puente, X.S., Jares, P., and Campo, E. (2018). Chronic lymphocytic leukemia and mantle cell lymphoma: crossroads of genetic and microenvironment interactions. Blood *131*, 2283–2296.

Queirós, a C., Villamor, N., Clot, G., et al. (2015). A B-cell epigenetic signature defines three biological subgroups of chronic lymphocytic leukemia with clinical impact. Leukemia 598–605.

Queirós, A.C., Beekman, R., Vilarrasa-Blasi, R., et al. (2016). Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. Cancer Cell *30*, 806–821.

Quesada, V., Conde, L., Villamor, N., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat. Genet. *44*, 47–52.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinodoz, S.A., Ollikainen, N., Tabak, B., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744-757.e24.

Rabl, C. (1885). Über Zelltheilung. Morphol. Jahrb. *10*, 214–330.

Rada-Iglesias, A., Ameur, A., Kapranov, P., et al. (2008). Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. Genome Res. *18*, 380–392.

Rada-Iglesias, A., Grosveld, F.G., and Papantonis, A. (2018). Forces driving the three-dimensional folding of eukaryotic genomes. Mol. Syst. Biol. *14*, e8214.

Radich, J.P. (2009). How I monitor residual disease in chronic myeloid leukemia. Blood *114*, 3376–3381.

Rafique, S., Thomas, J.S., Sproul, D., et al. (2015). Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. Genome Biol. *16*, 1–19.

Rahmatpanah, F.B., Carstens, S., Guo, J., et al. (2006). Differential DNA methylation patterns of small B-cell lymphoma subclasses with different clinical behavior. Leukemia *20*, 1855–1862.

Rai, K., Sawitsky, A., Cronkite, E., et al. (1975). Clinical staging of chronic lymphocytic leukemia. Blood *46*, 219–234.

Raineri, E., Dabad, M., and Heath, S. (2014). A Note on Exact Differences between Beta Distributions in Genomic (Methylation) Studies. PLoS One *9*, e97349.

Raineri, E., Serra, F., Beekman, R., et al. (2018). Inference of genomic spatial organization from a whole genome bisulfite sequencing sample. BioRxiv 384578.

Ramani, V., Deng, X., Qiu, R., et al. (2017). Massively multiplex single-cell Hi-C. Nat. Methods *14*, 263–266.

Rao, S.S.P., Huntley, M.H., Durand, N.C., et al. (2014). A 3D Map of the Human Genome at Kilobase

Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Raval, A., Tanner, S.M., Byrd, J.C., et al. (2007). Downregulation of Death-Associated Protein Kinase 1 (DAPK1) in Chronic Lymphocytic Leukemia. Cell *129*, 879–890.

Raviram, R., Rocha, P.P., Bonneau, R., et al. (2014). Interpreting 4C-Seq data: How far can we go? Epigenomics *6*, 455–457.

Rawstron, A.C., Bennett, F.L., O'Connor, S.J.M., et al. (2008). Monoclonal B-Cell Lymphocytosis and Chronic Lymphocytic Leukemia. N. Engl. J. Med. *359*, 575–583.

Reinius, L.E., Acevedo, N., Joerink, M., et al. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One *7*, e41361.

Ricci, M.A., Manzo, C., García-Parajo, M.F., et al. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell *160*, 1145–1158.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. *14*, 9–25.

Riggs, A.D., Russo, V.E.A., and Martienssen, R.A. (1996). Epigenetic Mechanisms of Gene Regulation (Plainview, N.Y.: Cold Spring Harbor Laboratory Press).

Roessler, S., Gyory, I., Imhof, S., et al. (2007). Distinct Promoters Mediate the Regulation of Ebf1 Gene Expression by Interleukin-7 and Pax5. Mol. Cell. Biol. *27*, 579–594.

Roschke, V., Kopantzev, E., Dertzbaugh, M., et al. (1997). Chromosomal translocations deregulating c-mye are associated with normal immune responses. Oncogene *14*, 3011–3016.

Rose-Zerilli, M.J.J., Forster, J., Parker, H., et al. (2014). ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia: Data from the UK LRF CLL4 trial. Haematologica *99*, 736–742.

Rose-Zerilli, M.J.J., Gibson, J., Wang, J., et al. (2016). Longitudinal copy number, whole exome and targeted deep sequencing of "good risk" IGHV-mutated CLL patients with progressive disease. Leukemia *30*, 1301–1310.

Rosenwald, A., Alizadeh, A.A., Widhopf, G., et al. (2001). Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. J. Exp. Med. *194*, 1639–1647.

Rosenwald, A., Wright, G., Wiestner, A., et al. (2003). The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell *3*, 185–197.

Rossi, D., and Gaidano, G. (2009). Richter syndrome: Molecular insights and clinical perspectives. Hematol. Oncol. *27*, 1–10.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19*, 789–800.

Le Roy, C., Deglesne, P.-A., Chevallier, N., et al. (2012). The degree of BCR and NFAT activation predicts clinical outcomes in chronic lymphocytic leukemia. Blood *120*, 356–365.

Royo, C., Salaverria, I., Hartmann, E.M., et al. (2011). The complex landscape of genetic alterations in mantle cell lymphoma. Semin. Cancer Biol. *21*, 322–334.

Royo, C., Navarro, A., Clot, G., et al. (2012). Non-nodal type of mantle cell lymphoma is a specific biological and clinical subgroup of the disease. Leukemia *26*, 1895–1898.

Rozwadowska, N., Kolanowski, T., Wiland, E., et al. (2013). Characterisation of Nuclear Architectural Alterations during In Vitro Differentiation of Human Stem Cells of Myogenic Origin. PLoS One *8*.

Rubio-Moscardo, F., Climent, J., Siebert, R., et al. (2005). Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. Blood *105*, 4445–4454.

Ruiz-Velasco, M., and Zaugg, J.B. (2017). Structure meets function: How chromatin organisation conveys functionality. Curr. Opin. Syst. Biol. *1*, 129–136.

Ryba, T., Hiratani, I., Lu, J., et al. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. *20*, 761–770.

Saba, N.S., Liu, D., Herman, S.E.M., et al. (2016). Pathogenic role of B-cell receptor signaling and canonical NF-κB activation in mantle cell lymphoma. Blood *128*, 82–92.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat. Rev. Mol. Cell Biol. *7*, 437–447.

Salaverria, I., Zettl, A., Beà, S., et al. (2007). Specific Secondary Genetic Alterations in Mantle Cell Lymphoma Provide Prognostic Information Independent of the Gene Expression–Based Proliferation Signature. J. Clin. Oncol. *25*, 1216–1222.

Salaverria, I., Royo, C., Carvajal-Cuenca, A., et al. (2013). CCND2 rearrangements are the most frequent genetic events in cyclin D1 - mantle cell lymphoma. Blood *121*, 1394–1402.

Sanda, T., Lawton, L.N., Barrasa, M.I., et al. (2012). Core Transcriptional Regulatory Circuit Controlled by the TAL1 Complex in Human T Cell Acute Lymphoblastic Leukemia. Cancer Cell *22*, 209–221.

Sattler, S. (2017). The role of the immune system beyond the fight against infection. In Advances in Experimental Medicine and Biology, (Springer New York LLC), pp. 3–14.

Saurin, A.J., Shiels, C., Williamson, J., et al. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J. Cell Biol. *142*, 887–898.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. *103*, 1412–1417.

Scharer, C.D., Barwick, B.G., Guo, M., et al. (2018). Plasma cell differentiation is controlled by multiple cell division-coupled epigenetic programs. Nat. Commun. *9*.

Schilham, M.W., Oosterwegel, M.A., Moerer, P., et al. (1996). Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature *380*, 711–714.

Schilham, M.W., Moerer, P., Cumano, A., et al. (1997). Sox-4 facilitates thymocyte differentiation. Eur. J. Immunol. *27*, 1292–1295.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., et al. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Schmitt, A.D., Hu, M., and Ren, B. (2016a). Genome-wide mapping and analysis of chromosome architecture. Nat. Rev. Mol. Cell Biol. *17*, 743–755.

Schmitt, A.D., Hu, M., Jung, I., et al. (2016b). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. *17*, 2042–2059.

Schoenfelder, S., Sexton, T., Chakalova, L., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat. Genet. *42*, 53–61.

Schoenfelder, S., Sugar, R., Dimond, A., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schubart, K., Massa, S., Schubart, D., et al. (2001). B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. Nat. Immunol. *2*, 69–74.

Schuh, A., Becq, J., Humphray, S., et al. (2012). Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood *120*, 4191–4196.

Schultze, J., Nadler, L.M., and Gribben, J.G. (1996). B7-mediated costimulation and the immune response. Blood Rev. *10*, 111–127.

Schütte, J., Reusch, J., Khandanpour, C., et al. (2019). Structural Variants as a Basis for Targeted Therapies in Hematological Malignancies. Front. Oncol. *9*.

Scott, D.W., Abrisqueta, P., Wright, G.W., et al. (2017). New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies. J. Clin. Oncol. *35*, 1668–1677.

Scrucca, L., Fop, M., Murphy, T.B., et al. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. *8*, 289–317.

Seifert, M., Sellmann, L., Bloehdorn, J., et al. (2012). Cellular origin and pathophysiology of chronic lymphocytic leukemia. *209*, 2183–2198.

Seng, T.J., Currey, N., Cooper, W.A., et al. (2008). DLEC1 and MLH1 promoter methylation are associated with poor prognosis in non-small cell lung carcinoma. Br. J. Cancer *99*, 375–382.

Seo, W., Ikawa, T., Kawamoto, H., et al. (2012). Runx1-Cbfβ facilitates early B lymphocyte development by regulating expression of Ebf1. J. Exp. Med. *209*, 1255–1262.

Serra, F., Baù, D., Goodstadt, M., et al. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput. Biol. *13*, 1–17.

Seto, M., Yamamoto, K., Iida, S., et al. (1992). Gene rearrangement and overexpression of PRAD1 in lymphoid malignancy with t(11;14)(q13;q32) translocation. Oncogene *7*, 1401–1406.

Sexton, T., Yaffe, E., Kenigsberg, E., et al. (2012a). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Sexton, T., Kurukuti, S., Mitchell, J.A., et al. (2012b). Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. Nat. Protoc. *7*, 1335–1350.

Shaffer, A.L., Rosenwald, A., and Staudt, L.M. (2002). Lymphoid malignancies: The dark side of B-cell differentiation. Nat. Rev. Immunol. *2*, 920–932.

Shearstone, J.R., Pop, R., Bock, C., et al. (2011). Global DNA Demethylation During Mouse Erythropoiesis in Vivo. Science (80-.). *334*, 799–802.

Shin, Y., Chang, Y.C., Lee, D.S.W., et al. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *175*, 1481-1491.e13.

Showe, L.C., Ballantine, M., Nishikura, K., et al. (1985). Cloning and sequencing of a c-myc oncogene in a Burkitt's lymphoma cell line that is translocated to a germ line alpha switch region. Mol. Cell. Biol. *5*, 501–509.

De Silva, N.S., and Klein, U. (2015a). Dynamics of B cells in germinal centres. Nat. Rev. Immunol. *15*, 137–148.

De Silva, N.S., and Klein, U. (2015b). Dynamics of B cells in germinal centres. Nat. Rev. Immunol. *15*, 137–148.

Simonis, M., Klous, P., Splinter, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Simonis, M., Kooren, J., and de Laat, W. (2007a). An evaluation of 3C-based methods to capture DNA interactions. Nat. Methods *4*, 895–901.

Simonis, M., Kooren, J., and de Laat, W. (2007b). An evaluation of 3C-based methods to capture DNA interactions. Nat. Methods *4*, 895–901.

Sims, G.P., Ettinger, R., Shirota, Y., et al. (2005). Identification and characterization of circulating human transitional B cells. Blood *105*, 4390–4398.

Sirito, M., Lin, Q., Maity, T., et al. (1994). Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. Nucleic Acids Res. *22*, 427–433.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*, 1–35.

Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. *3*, 1–25.

Soldini, D., Valera, A., Solé, C., et al. (2014). Assessment of SOX11 Expression in Routine Lymphoma Tissue Sections: Characterization of New Monoclonal Antibodies for Diagnosis of Mantle Cell Lymphoma. Am. J. Surg. Pathol. *38*, 86–93.

Soler-Vila, P., Cuscó, P., Farabella, I., et al. (2020). Hierarchical chromatin organization detected by TADpole. Nucleic Acids Res.

Solovei, I., Kreysing, M., Lanctôt, C., et al. (2009). Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. Cell *137*, 356–368.

Song, J., and Shao, H. (2015). SNP Array in Hematopoietic Neoplasms: A Review. Microarrays *5*, 1.

Song, L., and Crawford, G.E. (2010). DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb. Protoc. *5*.

Song, S., and Matthias, P.D. (2018). The Transcriptional Regulation of Germinal Center Formation. Front. Immunol. *9*, 2026.

Speicher, M.R., and Carter, N.P. (2005). The new cytogenetics: Blurring the boundaries with molecular biology. Nat. Rev. Genet. *6*, 782–792.

Spielmann, M., Brancati, F., Krawitz, P.M., et al. (2012). Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am. J. Hum. Genet. *91*, 629–635.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.

Splinter, E., Heath, H., Kooren, J., et al. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. *20*, 2349–2354.

Stadhouders, R., Vidal, E., Serra, F., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. *50*, 238–249.

Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. Nature *569*, 345–354.

Stamatopoulos, K., Belessi, C., Moreno, C., et al. (2007). Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood *109*, 259–270.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780–791.

Stevens, T.J., Lando, D., Basu, S., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59–64.

Stilgenbauer, S., Winkler, D., Ott, G., et al. (1999). Molecular Characterization of 11q Deletions Points to a Pathogenic Role of the ATM Gene in Mantle Cell Lymphoma. Blood *94*, 3262–3264.

Stock, J.K., Brookes, E., and Pombo, A. Phospho-sensitive chromatin immunoprecipitation of RNA Polymerase II (Prot 48).

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Stransky, N., Vallot, C., Reyal, F., et al. (2006). Regional copy number-independent deregulation of transcription in cancer. Nat. Genet. *38*, 1386–1396.

Strasser, A., Harris, A.W., Bath, M.L., et al. (1990). Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. Nature *348*, 331–333.

Summers, K.E., Goff, L.K., Wilson, A.G., et al. (2001). Frequency of the Bcl-2/IgH rearrangement in normal individuals: Implications for the monitoring of disease in patients with follicular lymphoma. J. Clin. Oncol. *19*, 420–424.

Swain, J.L., Stewart, T.A., and Leder, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. Cell *50*, 719–727.

Swerdlow, S.H., and Williams, M.E. (2002). From centrocytic to mantle cell lymphoma: A clinicopathologic and molecular review of 3 decades. Hum. Pathol. *33*, 7–20.

Swerdlow, S.H., Habeshaw, J.A., Murray, L.J., et al. (1983). Centrocytic lymphoma: A distinct clinicopathologic and immunologic entity. A multiparameter study of 18 cases at diagnosis and relapse.

Am. J. Pathol. *113*, 181–197.

Swerdlow, S.H., Utz, G.L., and Williams, M.E. (1993). Bcl-2 protein in centrocytic lymphoma; a paraffin section study. Leukemia *7*, 1456–1458.

Swerdlow, S.H., Campo, E., Pileri, S.A., et al. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood *127*, 2375–2390.

Swerdlow, S.H., Campo, E., Harris, N.L., et al. (2017). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. (Lyon: International Agency for Research on Cancer).

Szalaj, P., and Plewczynski, D. (2018). Three-dimensional organization and dynamics of the genome. Cell Biol. Toxicol. *34*, 381–404.

Taberlay, P.C., Achinger-Kawecka, J., Lun, A.T.L., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. *26*, 719–731.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. *99*, 3740–3745.

Tangye, S.G., Ma, C.S., Brink, R., et al. (2013). The good, the bad and the ugly-T FH cells in human health and disease. Nat. Rev. Immunol. *13*, 412–426.

Tatavosian, R., Kent, S., Brown, K., et al. (2019). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. J. Biol. Chem. *294*, 1451–1463.

Taub, R., Kirsch, I., Morton, C., et al. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc. Natl. Acad. Sci. U. S. A. *79*, 7837–7841.

Timár, B., Fülöp, Z., Csernus, B., et al. (2004). Relationship between the mutational status of VH genes and pathogenesis of diffuse large B-cell lymphoma in Richter's syndrome. Leukemia *18*, 326–330.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D.., et al. (2008). PcG Proteins, DNA Methylation, and Gene Repression by Chromatin Looping. PLoS Biol. *6*, e306.

Tost, J., and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. Nat. Protoc. *2*, 2265–2275.

Treiber, T., Mandel, E.M., Pott, S., et al. (2010). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714–725.

Tremethick, D.J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. Cell *128*, 651–654.

Trussart, M., Serra, F., Bau, D., et al. (2015). Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. Nucleic Acids Res. *43*, 3465–3477.

Unnikrishnan, A., Guan, Y.F., Huang, Y., et al. (2016). A quantitative proteomics approach identifies ETV6 and IKZF1 as new regulators of an *ERG* -driven transcriptional network. Nucleic Acids Res. *44*, 10644–10661.

Vaandrager, J.W., Schuuring, E., Zwikstra, E., et al. (1996). Direct visualization of dispersed 11q13 chromosomal translocations in mantle cell lymphoma by multicolor DNA fiber fluorescence in situ hybridization. Blood *88*, 1177–1182.

Vegliante, M.C., Royo, C., Palomero, J., et al. (2011). Epigenetic activation of SOX11 in Lymphoid Neoplasms by Histone modifications. PLoS One *6*.

Vegliante, M.C., Palomero, J., Perez-Galan, P., et al. (2013). SOX11 regulates PAX5 expression and blocks terminal B-cell differentiation in aggressive mantle cell lymphoma. Blood *121*, 2175–2185.

Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001). The Sequence of the Human Genome. Science (80-.). *291*, 1304–1351.

Vidal, E., le Dily, F., Quilez, J., et al. (2018). OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. Nucleic Acids Res. *46*, e49–e49.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., et al. (2015). Clustering of mammalian Hox genes with other

H3K27me3 targets within an active nuclear domain. Proc. Natl. Acad. Sci. U. S. A. *112*, 4672–4677.

Vinson, C., and Chatterjee, R. (2012). CG methylation. Epigenomics *4*, 655–663.

Volpi, E. V., and Bridger, J.M. (2008). FISH glossary: An overview of the fluorescence in situ hybridization technique. Biotechniques *45*, 385–409.

Waddington, C. (1942). The epigenotype. Endeavour 18–20.

Waddington, C.H. (1957). The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. Strateg. Genes. A Discuss. Some Asp. Theor. Biol. With an Append. by H. Kacser.

Wang, K.C., and Chang, H.Y. (2018). Epigenomics. Circ. Res. *122*, 1191–1199.

Wang, D.C., Wang, W., Zhang, L., et al. (2019). A tour of 3D genome with a focus on CTCF. Semin. Cell Dev. Biol. *90*, 4–11.

Wang, H., Maurano, M.T., Qu, H., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. *22*, 1680–1688.

Wang, L., Lawrence, M.S., Wan, Y., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N. Engl. J. Med. *365*, 2497–2506.

Wang, S., Lee, S., Chu, C., et al. (2020). HiNT: A computational method for detecting copy number variations and translocations from Hi-C data. Genome Biol. *21*, 73.

Wani, A.H., Boettiger, A.N., Schorderet, P., et al. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat. Commun. *7*, 10291.

Wasik, A.M., Lord, M., Wang, X., et al. (2013). SOXC transcription factors in mantle cell lymphoma: the role of promoter methylation in SOX11 expression. Sci. Rep. *3*, 1400.

Wästerlid, T., Nordström, L., Freiburghaus, C., et al. (2017). Frequency and clinical implications of SOX11 expression in Burkitt lymphoma. Leuk. Lymphoma *58*, 1760–1763.

Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. Genes Dev. *28*, 672–682.

Wegner, M. (2010). All purpose Sox: The many roles of Sox proteins in gene expression. Int. J. Biochem. Cell Biol. *42*, 381–390.

Weigle, B., Ebner, R., Temme, A., et al. (2005). Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas. Oncol. Rep. *13*, 139–144.

Welzel, N., Le, T., Marculescu, R., et al. (2001). Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: Implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res. *61*, 1629–1636.

van de Werken, H.J.G., Landan, G., Holwerda, S.J.B., et al. (2012a). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods *9*, 969–972.

van de Werken, H.J.G., de Vree, P.J.P., Splinter, E., et al. (2012b). 4C Technology: Protocols and Data Analysis. In Methods in Enzymology, (Elsevier Inc.), pp. 89–112.

Whyte, W.A., Orlando, D.A., Hnisz, D., et al. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wiestner, A., Tehrani, M., Chiorazzi, M., 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. Blood *109*, 4599–4606.

Wilker, P.R., Kohyama, M., Sandau, M.M., et al. (2008). Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. Nat. Immunol. *9*, 603–612.

Will, A.J., Cova, G., Osterwalder, M., et al. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat. Genet. *49*, 1539–1545.

Willis, T.G., and Dyer, M.J.S. (2000). The role of immunoglobulin translocations in the pathogenesis of B-

cell malignancies. Blood *96*, 808–822.

de Wit, E., and de Laat, W. (2012a). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

de Wit, E., and de Laat, W. (2012b). A decade of 3C technologies: Insights into nuclear organization. Genes Dev.

Witko-Sarsat, V., Rieu, P., Descamps-Latscha, B., et al. (2000). Neutrophils: Molecules, functions and pathophysiological aspects. Lab. Investig. *80*, 617–654.

Wojdacz, T.K., Amarasinghe, H.E., Kadalayil, L., et al. (2019). Clinical significance of DNA methylation in chronic lymphocytic leukemia patients: Results from 3 UK clinical trials. Blood Adv. *3*, 2474–2481.

Wolf, S.F., and Migeon, B.R. (1982). Studies of X chromosome DNA methylation in normal human cells. Nature *295*, 667–671.

Wolffe, A.P. (1999). Chromatin: Structure and Function (San Diego: Academic).

Woodcock, C.L.F., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. Exp. Cell Res. *97*, 101–110.

Woolhouse, M.E.J., Webster, J.P., Domingo, E., et al. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. *32*, 569–577.

Wu, P., Li, T., Li, R., et al. (2017). 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. Nat. Commun. *8*.

Würtele, H., and Chartrand, P. (2006). Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosom. Res. *14*, 477–495.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. *43*, 1059–1065.

Yamazaki, H., Suzuki, M., Otsuki, A., et al. (2014). A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer Cell *25*, 415–427.

Yan, K.-K., Yardımcı, G.G., Yan, C., et al. (2017). HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. Bioinformatics *33*, 2199–2201.

Ye, T., Ravens, S., Krebs, A.R., et al. (2014). Interpreting and Visualizing ChIP-seq Data with the seqMINER Software. In Methods in Molecular Biology, (Humana Press Inc.), pp. 141–152.

de Yébenes, V.G., and Ramiro, A.R. (2006). Activation-induced deaminase: light and dark sides. Trends Mol. Med. *12*, 432–439.

Ying, C.Y., Dominguez-Sola, D., Fabi, M., et al. (2013). MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat. Immunol. *14*, 1084–1092.

Zack, T.I., Schumacher, S.E., Carter, S.L., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

Zandi, S., Bryder, D., and Sigvardsson, M. (2010). Load and lock: the molecular mechanisms of B-lymphocyte commitment. Immunol. Rev. *238*, 47–62.

Zarrin, A.A., Del Vecchio, C., Tseng, E., et al. (2007). Antibody class switching mediated by yeast endonuclease-generated DNA breaks. Science (80-.). *315*, 377–381.

Zenz, T., Mertens, D., Küppers, R., et al. (2010). From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat. Rev. Cancer *10*, 37–50.

Zhang, J., Jima, D., Moffitt, A.B., et al. (2014). The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. Blood *123*, 2988–2996.

Zhang, L., Murray, F., Zahno, A., et al. (2008a). Cyclic nucleotide phosphodiesterase profiling reveals increased expression of phosphodiesterase 7B in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. U. S. A. *105*, 19532–19537.

Zhang, Y., Liu, T., Meyer, C.A., et al. (2008b). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*,

R137.

Zhang, Y., McCord, R.P., Ho, Y.J., et al. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908–921.

Zhao, S., Guo, Y., and Shyr, Y. (2015). KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway. R Packag. Version 1.10.0.

Zhao, Z., Tavoosidana, G., Sjölinder, M., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

Zuin, J., Dixon, J.R., Van Der Reijden, M.I.J.A., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001.

(1997). A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. Blood *89*, 3909–3918.

APPENDIX
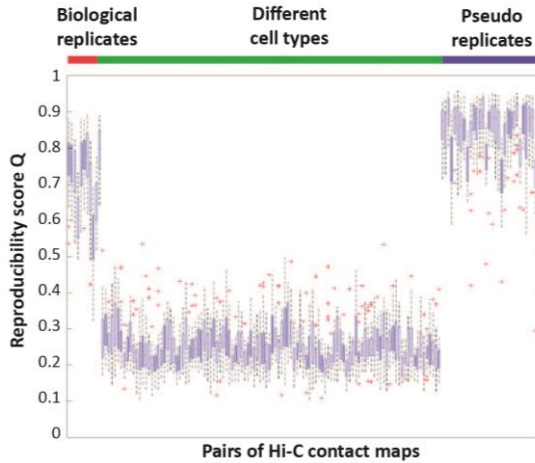
Appendix Figure 1. Reproducibility score for three sets of Hi-C contact maps pairs. Contact maps generated from Hi-C experiments performed in 11 cell lines are shown. Biological replicates referred to a pair of replicates of the same experiment. Pseudo replicates were obtained by pooling the reads from two replicates together performing down sampling. There were 11 biological replicates, 33 pairs of pseudo replicates, and 110 pairs of maps between different cell types. Each box shows for a pair the distribution of reproducibility score Q in 23 chromosomes, with crosses as the outliers. Analysis were performed using the HiC-spector, a matrix library for spectral and reproducibility analysis of Hi-C contact maps. Source: Yan et al., 2017.
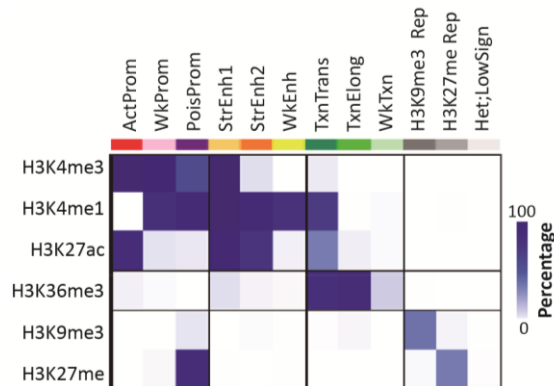
Appendix Figure 2. Emissions of the generated chromatin state model. Heatmap showing the percentages of regions assigned to a specific chromatin state (columns) that contain a specific histone mark (rows). ActProm, Active Promoter; WkProm, Weak Promoter; PoisProm, Poised Promoter; StrEnh1, Strong Enhancer 1; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; TxnTrans, Transcription Transition; TxnElong, Transcription Elongation; WkTxn, Weak Transcription; PolycombRepr, Polycomb-Repressed; Het;Repr, Heterochromatin-Repressed; Het;LowSign, Heterochromatin-Low Signal. Source: Beekman et al., 2018a.
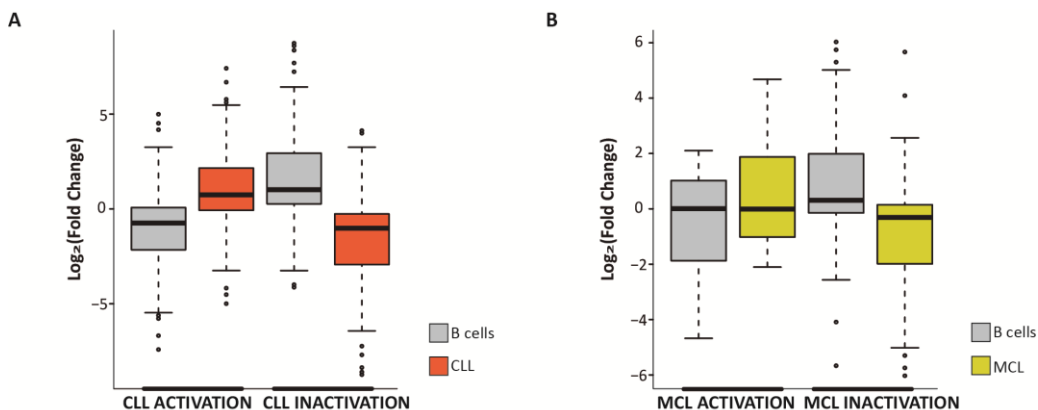
Appendix Figure 3. Gene expression of the dynamic compartments upon neoplasia transformation. A and **B.** Boxplots of the genes expressed in B cells and CLL (**A**) or MCL (**B**) in the significant dynamic compartments.
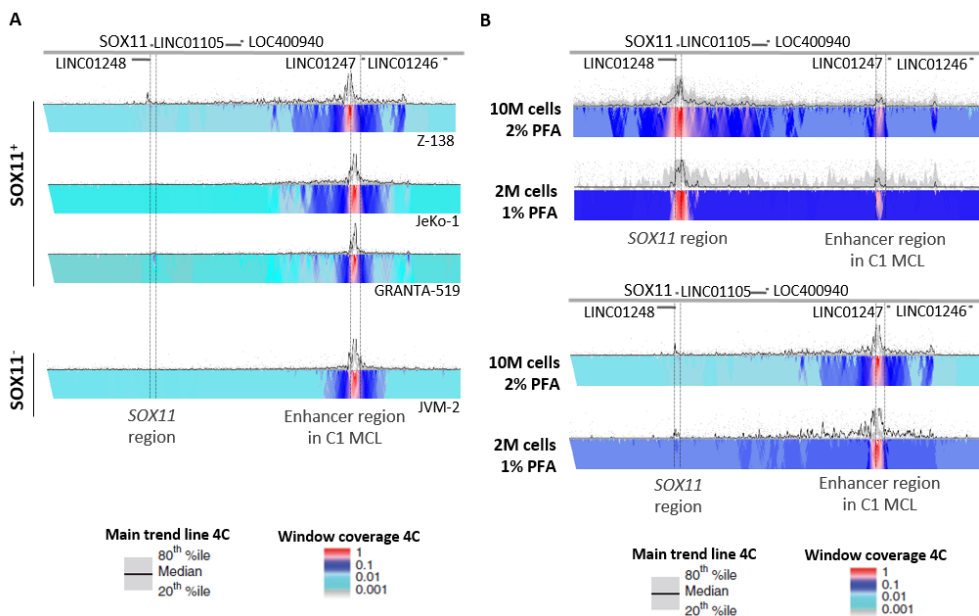
Appendix Figure 4. 4C-seq of SOX11 locus. A. Normalized 4C-seq intensities taking the enhancer in MCL C1 as viewpoint. **B.** Normalized 4C-seq intensities of Z-138 under two different conditions, 10M cells cross-linked with 2% formaldehyde or 2M cells cross-linked with 1% formaldehyde. 4C-seq were performed taking the *SOX11* region (*upper* panel) or the enhancer in MCL C1 (*lower* panel) as viewpoint.
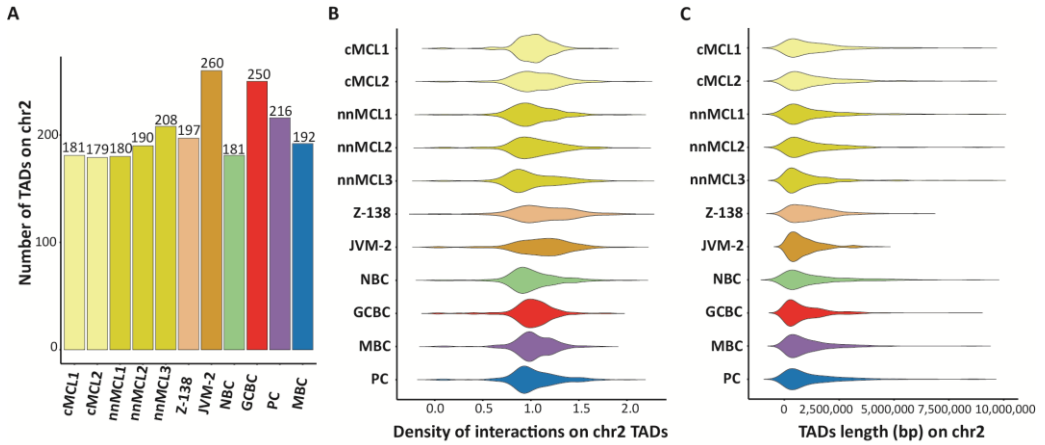
Appendix Figure 5. TAD features for merged samples. Number (**A**), density of interactions (**B**) and length distribution (**C**) of TADs at chromosome 2 for each studied sample.
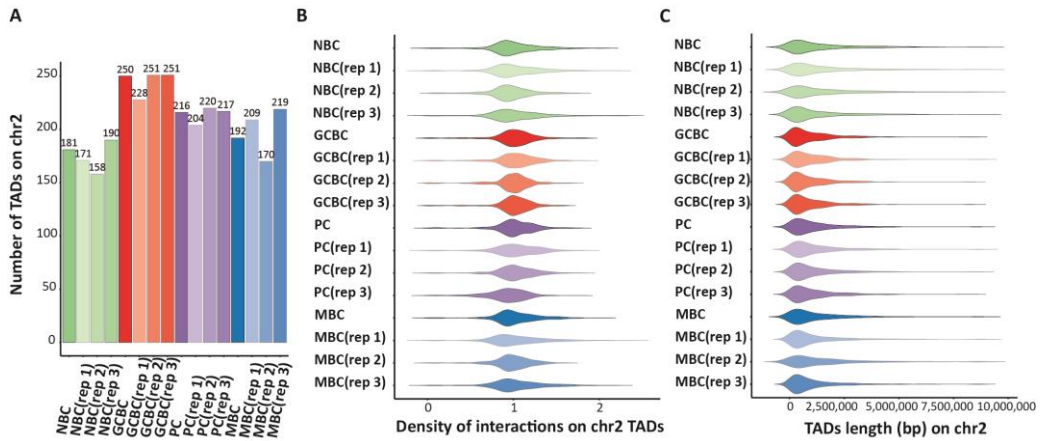
Appendix Figure 6. **TAD features for replicate samples.** Number (**A**), density of interactions (**B**) and length distribution (**C**) of TADs at chromosome 2 for each B-cell subpopulation (merged and replicates).

Sample	Case	B-cell neoplasia	Gender	Age at Diagnosis	IGHV status	IGHV % identity to germline	IGHV gene	SOX11
CLL	12	uCLL1	F	59	UNMUT	100	IGHV1-69	NE
	182	uCLL2	M	73	UNMUT	99,08	IGHV3-74	NE
	110	mCLL1	F	66	MUT	97,91	IGHV1-2	NE
	1228	mCLL2	M	69	MUT	90,41	IGHV1-69	NE
	1525	mCLL3	M	48	MUT	94,09	IGHV1-8	NE
	1532	mCLL4	M	84	MUT	96,8	IGHV3-21	NE
	3	mCLL5	M	45	MUT	97,25	IGHV4-61	NE
MCL	M019	cMCL1	M	82	UNMUT	99,25	IGHV1-18	Pos
	M001	cMCL2	M	63	UNMUT	98,80	IGHV3-21	Pos
	M009	nnMCL1	F	78	MUT	92,36	IGHV3-9	Neg
	M076	nnMCL2	M	66	MUT	93,69	IGHV4-59	Neg
	M004	nnMCL3	M	81	MUT	97,22	IGHV1-8	Neg

M = Male, F= Female
UNMUT = Unmutated, MUT = Mutated
NE = Not Evaluated, Pos = Positive, Neg = Negative

Appendix Table 1. Patient characteristics. The source was peripheral blood from all these samples.

B-cell subpopulation		Sequenced entries	Uniquely mapped		Mapped interactions	Filtering (% from mapped interactions)										Valid pairs	
Name	Code		Read1	Read2		SC	DE	CR	OR	D	E	RB	ED	S	L	Reads	% int
NBC (rep1)	EPI0000084	275.136.893	238.973.462	233.373.525	200.978.594	0,0	12,3	22,2	1,0	32,4	0,3	0,9	8,8	2,4	0,0	118.758.489	**59,1**
NBC (rep2)	EPI0000085	290.747.706	251.016.969	245.558.210	209.678.557	0,0	10,4	18,1	1,1	27,7	0,2	1,0	7,6	2,1	0,0	136.246.660	**65,0**
NBC (rep3)	EPI0000086	311.730.793	270.029.538	259.749.816	221.808.619	0,0	11,5	20,4	1,0	53,0	0,4	1,1	8,9	2,5	0,0	95.891.758	**43,2**
NBC		877.615.392	760.019.969	738.681.551	632.465.770	0,0	11,4	20,2	1,0	38,3	0,3	1,0	8,4	2,3	0,0	350.777.527	**55,5**
GCBC (rep1)	EPI0000087	292.945.674	256.039.997	247.446.057	211.887.433	0,0	7,6	19,8	1,0	38,9	0,2	0,3	8,1	2,3	0,0	123.937.761	**58,5**
GCBC (rep2)	EPI0000058	278.050.643	243.497.321	226.404.994	193.204.061	0,0	10,0	17,4	1,1	38,8	0,1	0,9	8,8	2,1	0,0	109.316.427	**56,6**
GCBC (rep3)	EPI0000059	276.647.300	241.208.574	217.477.480	186.033.233	0,0	10,7	17,3	1,1	34,7	0,2	1,0	8,6	2,0	0,0	110.222.741	**59,2**
GCBC		847.643.617	740.745.892	691.328.531	591.124.727	0,0	9,3	18,2	1,0	37,8	0,2	0,8	8,5	2,1	0,0	343.353.657	**58,1**
MBC (rep1)	EPI0000088	309.291.856	267.659.517	258.163.514	222.046.368	0,0	14,0	23,7	1,0	45,5	0,5	1,1	9,9	2,6	0,0	104.788.126	**47,2**
MBC (rep2)	EPI0000089	295.660.477	255.880.473	250.016.318	213.894.487	0,0	9,6	18,5	1,1	16,6	0,3	0,9	7,7	2,1	0,0	160.753.368	**75,2**
MBC (rep3)	EPI0000090	300.121.716	259.858.740	251.809.838	216.344.804	0,0	9,3	19,8	1,0	40,3	0,3	0,9	7,4	2,2	0,0	118.559.488	**54,8**
MBC		905.074.049	783.398.730	759.989.670	652.285.659	0,0	11,0	20,7	1,0	34,5	0,4	1,0	8,3	2,3	0,0	383.825.399	**58,8**
PC (rep1)	EPI0000091	281.020.068	241.636.939	232.447.744	199.010.722	0,0	11,5	21,9	1,1	57,8	0,5	0,5	8,9	2,4	0,0	77.274.162	**38,8**
PC (rep2)	EPI0000060	290.762.661	253.273.109	241.461.097	207.025.863	0,0	8,0	17,6	1,0	43,8	0,3	0,6	7,6	2,0	0,0	111.312.188	**53,8**
PC (rep3)	EPI0000071	287.145.402	251.288.847	239.017.227	205.749.632	0,1	8,6	16,6	1,0	45,9	0,5	0,1	10,6	2,4	0,0	102.024.282	**49,6**
PC		858.928.131	746.198.895	712.926.068	611.786.217	0,0	9,4	18,7	1,0	49,2	0,4	0,4	9,0	2,3	0,0	290.599.729	**47,5**

Appendix Table 2. *In situ* Hi-C experimental quality metrics from B-cell subpopulations. SC, self circle. DE, dangling-end. CR, close to RE. OR; over represented. D, duplicated. E, error. RB, random break. ED, extra dangling. S, too short. L, too Large. Int, intersection.

B-cell neoplasia		Sequenced entries	Uniquely mapped		Mapped interactions	Filtering (% from mapped interactions)										Valid pairs	
Name	Code		Read1	Read2		SC	DE	CR	OR	D	E	RB	ED	S	L	Reads	% int
uCLL1	EPI0000105	269.480.531	244.425.392	230.876.541	199.534.199	0,0	8,2	19,5	1,1	25,4	0,2	0,5	10,4	2,3	0,0	140.034.119	**70,2**
uCLL2	EPI0000109	213.574.350	184.564.968	175.170.081	149.572.584	0,0	7,6	18,0	1,2	42,9	0,3	0,2	9,6	2,1	0,0	82.139.685	**54,9**
mCLL1	EPI0000104	257.866.192	233.627.290	220.796.484	191.227.697	0,0	9,2	18,3	1,2	19,1	0,3	0,6	9,9	2,1	0,0	142.531.393	**74,5**
mCLL2	EPI0000106	312.038.119	274.834.835	267.726.968	228.559.899	0,0	6,2	17,1	1,2	37,6	0,2	0,2	8,3	2,0	0,0	139.276.566	**60,9**
mCLL3	EPI0000107	247.275.310	220.150.171	202.417.268	174.702.640	0,1	7,4	19,7	1,1	52,6	0,4	0,2	12,4	2,7	0,0	82.892.746	**47,4**
mCLL4	EPI0000108	303.890.427	268.116.573	260.473.860	221.958.771	0,0	6,5	17,4	1,3	40,9	0,2	0,2	8,9	2,0	0,0	128.695.501	**58,0**
mCLL5	EPI0000110	318.138.428	280.050.162	274.340.304	233.978.267	0,0	7,4	17,4	1,2	38,4	0,3	0,3	8,9	2,0	0,0	138.351.660	**59,1**
cMCL1	EPI0000083	246.342.237	219.343.291	204.974.310	177.833.804	0,0	9,9	21,5	1,2	45,4	0,3	0,5	10,2	2,4	0,0	90.821.913	**51,1**
cMCL2	EPI0000101	211.921.058	181.396.571	174.329.249	149.124.272	0,0	9,9	20,7	1,5	27,0	0,3	0,3	9,2	2,2	0,0	97.651.988	**65,5**
nnMCL1	EPI0000100	215.238.134	171.557.746	164.309.122	140.820.157	0,0	11,3	23,8	1,0	33,7	0,4	0,4	10,4	2,5	0,0	81.932.326	**58,2**
nnMCL2	EPI0000102	227.155.429	193.846.395	186.602.287	157.026.571	0,0	8,1	19,4	1,0	36,5	0,3	0,4	8,9	2,2	0,0	93.571.457	**59,6**
nnMCL3	EPI0000103	241.408.692	205.387.567	198.629.405	169.863.325	0,1	11,2	32,4	1,0	24,8	0,6	0,4	10,8	3,3	0,0	109.014.551	**64,2**
JVM-2	EPI0000111	276.932.165	237.258.005	224.827.659	195.427.028	0,1	13,6	32,7	1,0	49,8	0,7	0,3	13,1	3,4	0,0	80.150.476	**41,0**
Z-138	EPI0000112	364.659.479	274.880.176	250.863.196	213.438.903	0,0	11,1	34,0	1,1	61,6	0,4	0,2	13,6	3,5	0,0	72.090.208	**33,8**

Appendix Table 3. *In situ* Hi-C experimental quality metrics from B-cell neoplasias and two MCL cell lines (Z-138 and JVM-2). SC, self circle. DE, dangling-end. CR, close to RE. OR; over represented. D, duplicated. E, error. RB, random break. ED, extra dangling. S, too short. L, too Large. Int, intersection.

Barcode (ICGC)	H3K27ac	H3K4me1	H3K4me3	H3K36me3	H3K9me3	H3K27me3	Input_1	Input_2	Input_3	ATAC-seq	RNA-seq	WGBS	DNA methy.	Hist. mod/ATAC-seq/RNA-seq
NBCB1	S00X9SH1							-	-	-	-	-		4
NBCB2	S00XAQH1							-	-	-	-	-		4
NBCB3	S0159LH1							-	-	-	-	-		4
NBCB4	-	-	-	-	-	-	-	-	-	-	S019HQ11	-		4
NBCB5	-	-	-	-	-	-	-	-	-	-	S019JM11	-		4
NBCB6	-	-	-	-	-	-	-	-	-	-	S019KK11	-		4
NBCB7	-	-	-	-	-	-	-	-	-	S01ECG71	-	S01ECGA1	4	
NBCB8	-	-	-	-	-	-	-	-	-	-	-	NBC_NC11.42	1	
NBCB9	-	-	-	-	-	-	-	-	-	-	-	NBC_NC11.41	2	
NBCB10	-	-	-	-	-	-	-	-	-	S01SME71	-	-		4
NBCB11	-	-	-	-	-	-	-	-	-	S01SHO71	-	-		4
GCBC1	S00W0DH1							-	-	-	-	-		4
GCBC2	S00Y9OH1							-	-	-	S00Y9O11	S00Y9OA1	2	4
GCBC3	S013ARH1							-	-	-	S013AR11	S013ARA1	4	4
GCBC4	-	-	-	-	-	-	-	-	-	-	S019FU11	-		4
GCBC5	-	-	-	-	-	-	-	-	-	-	-	GCBC_T11.8	2	
GCBC6	-	-	-	-	-	-	-	-	-	S01DK471	-	-		4
GCBC7	-	-	-	-	-	-	-	-	-	S01SDW71	-	-		4
GCBC8	-	-	-	-	-	-	-	-	-	S01SGQ71	-	-		4
ncsMBC1	S015DDH1							-	-	-	-	-		4
ncsMBC2	-	-	-	-	-	-	-	-	-	-	S019NE11	-		4
ncsMBC3	-	-	-	-	-	-	-	-	-	-	-	ncsMBC_NC11.42	1	
ncsMBC4	-	-	-	-	-	-	-	-	-	S01EEC71	-	-		4
ncsMBC5	-	-	-	-	-	-	-	-	-	S01SKI71	-	-		4
ncsMBC6	-	-	-	-	-	-	-	-	-	S01SP871	-	-		4
csMBC1	S015BHH1							-	-	-	-	-		4
csMBC2	S015CFH1							-	-	-	-	-		4
csMBC3	-	-	-	-	-	-	-	-	-	-	S019MG11	-		4
csMBC4	-	-	-	-	-	-	-	-	-	-	S01TDS11	-		4
csMBC5	-	-	-	-	-	-	-	-	-	-	-	csMBC_NC11.42	1	
csMBC7	-	-	-	-	-	-	-	-	-	S01EDE71	-	-		4
csMBC8	-	-	-	-	-	-	-	-	-	S01SJK71	-	-		4
csMBC9	-	-	-	-	-	-	-	-	-	S01SNC71	-	-		4

Appendix Table 4. General overview of the omic layers analyzed with the reference if the data is published, (see continuation in next page).

| Barcode (ICGC) | Histone modifications | | | | | | | | | ATAC-seq | RNA-seq | WGBS | References | |
	H3K27ac	H3K4me1	H3K4me3	H3K36me3	H3K9me3	H3K27me3	Input_1	Input_2	Input_3				DNA methy.	Hist. mod/ ATAC-seq/RNA-seq
PCT1				S00VKEH1				-	-	-	-	-		4
PCT2				S00Y8QH1				-	-	-	S00Y8Q11	S00Y8QA1	4	4
PCT3				S0139TH1			S014BPH1	S01EBIH1		-	S0149T11	-		4
PCT4	-	-	-	-	-	-	-	-	-	-	S019GS11	-		4
PCT5	-	-	-	-	-	-	-	-	-	-	-	PBC_T14.1	2	
PCT6	-	-	-	-	-	-	-	-	-	-	-	PBC_T12.20	2	
PCT7	-	-	-	-	-	-	-	-	-	S01DJ671	-	-		4
PCT8	-	-	-	-	-	-	-	-	-	S01SCY71	-	-		4
PCT9	-	-	-	-	-	-	-	-	-	S01SFS71	-	-		4
uCLL1				S00B0NH1				-	-	S00B0N71	S00B0N11	S00B0NA1	4	4
uCLL2				S00B1LH1				-	-	S00B1L71	S00B1L11	S00B1LA1	4	4
mCLL1				S00AYXH1				-	-	S00AYX71	S00AYX11	S00AYXA1	4	4
mCLL2				S01FG4H1				-	-	S01FG471	S01FG411	S01FG4A1	4	4
mCLL3				S00B2JH1				-	-	S00B2J71	S00B2J11	S00B2JA1	4	4
mCLL4				S00GPRH1				-	-	S00GPR71	S00GPR11	S00GPRA1	4	4
mCLL5				S00AXZH1				-	-	S00AXZ71	S00AXZ11	3CLL	1	4
cMCL1				S01FH2H1				-	-	S01FH271	S01FH211	S01FH2A1		
cMCL2				S01FKXH1				-	-	S01FKX71	S01FKX11	S01FKXA1	3	3*
nnMCL1				S01FJZH1				-	-	S01FJZ71	S01FJZ11	S01FJZA1		
nnMCL2				S01FF6H1				-	-	S01FF671	S01FF611	S01FF6A1		
nnMCL3				S01FE8H1				-	-	S01FE871	S01FE811	S01FE8A1	3	3*

* H3K27ac, H3K4me3, H3K4me1

Appendix Table 4 (continuation). General overview of the omic layers analyzed with the reference if the data is published. 1, Kulis *et al.,* 2012; 2, Kulis *et al.,* 2015; 3, Queriós *et al.,* 2016; 4, Beekman *et al.,* 2018.

A

Motif ID	Transcription factor	*p*-value
MA0052.1	MEF2A	3.98e-07
MA0052.3	MEF2A	5.99e-06
MA0773.1	MEF2D	6.44e-06
MA0660.1	MEF2B	7.31e-06
MA0052.2	MEF2A	2.43e-05
MA0786.1	POU3F1	4.25e-05
MA0628.1	POU6F1	5.79e-05
MA0785.1	POU2F1	6.45e-05
MA0497.1	MEF2C	7.29e-05
MA0662.1	MIXL1	9.45e-05
MA0790.1	POU4F1	9.97e-05
MA0495.2	MAFF	0.000532
MA0507.1	POU2F2	0.000905
MA0867.1	SOX4	0.000978
MA0158.1	HOXA5	0.00127
MA1124.1	ZNF24	0.00205
MA0698.1	ZBTB18	0.00217
MA0124.1	NKX3-1	0.00217
MA0893.1	GSX2	0.00382
MA0488.1	JUN	0.00407
MA0895.1	HMBOX1	0.00539
MA0655.1	JDP2	0.00714
MA0914.1	ISL2	0.00733
MA0754.1	CUX1	0.0127
MA0903.1	HOXB3	0.0132
MA0834.1	ATF7	0.0132
MA0476.1	FOS	0.016

B

Motif ID	Transcription factor	*p*-value
MA0830.1	TCF4	4.11e-05
MA0522.2	TCF3	0.000149
MA0138.1	REST	0.000367
MA0106.2	TP53	0.000611
MA0103.2	ZEB1	0.0016
MA0103.3	ZEB1	0.00265
MA0871.1	TFEC	0.00366
MA0625.1	NFATC3	0.00647
MA0526.1	USF2	0.00663
MA0106.3	TP53	0.00912
MA0052.3	MEF2A	0.00934
MA0088.2	ZNF143	0.00956
MA1117.1	RELB	0.0098
MA0497.1	MEF2C	0.0118
MA0692.1	TFEB	0.0142
MA0464.2	BHLHE40	0.0159
MA1112.1	NR4A1	0.0182
MA0052.2	MEF2A	0.0186
MA0058.1	MAX	0.0186

Appendix Table 5. List of identified enriched binding motifs expressed in GCBC (**A**) and CLL-specific active compartments (**B**). The motif identification (ID), name of the transcription factor and *p*-value is shown.

Genes differentially upregulated at GCBC specific active compartments
AC096559.1, AICDA, APAF1, ARHGDIB, ASPH, BASP1, BCAT1, BMPR1A, BPTF, C21orf91, C3orf67, CACYBP, CADPS, CCDC171, CCDC18, CCT8, CDKN3, CHN2, CMPK2, CNTLN, COMMD10, CYP39A1, DARS2, DIAPH3, EFNB2, ENPP3, FAM184A, FAT3, FBXO30, FGD6, FNDC1, FRY, GINS3, GPR52, GPSM2, GRAMD1B, GTDC1, HELLS, HMHB1, KCNMB4, KCNQ5, KIAA1211, KNOP1P4, LINC00158, LRMP, MAP3K7CL, MAPK10, MED12L, METAP2, MMS22L, MRPL39, MTF2, MYO3B, MYO9A, NUDFA12, PDE8B, PID1, PIGN, pk, PRDX6, PRIM2, PRKCH, PRRG4, RABGAP1L, RAPGEF5, RCSD1, RFC3, RGS8, RP11-22B23.1, RP11-296O14.3, RP11-444D3.1, RP11-478C6.5, RP11-478C6.6, RP11-545I5.3, RSAD2, SFXN1, SGCB, SKAP2, SLC25A27, SMPDL3A, SORL1, SPAG16, SPC25, STK39, SUGCT, TBC1D4, TEX41, TIAM2, TMPO, TMPO-AS1, TNFSF11, TRIM9, TRIP13, TTK, VEZT, WASF3, ZBTB44, ZNF732, ZNF876P, ZRANB3

Appendix Table 6. List of genes differentially upregulated (FDR<0.05) at GCBC specific active compartments, alphabetically ordered.

Genes differentially downregulated in CLL and MCL shared inactivated compartments
AC023590.1, AC093818.1, ASPH, EXT1, MIR17HG, MKI67, PDE4D, PDK1, RP11-159H10.3, SAMD12, SH3RF1, SLCO5A1

Appendix Table 7. Genes differentially downregulated (FDR<0.05) in the compartments commonly losing interactions upon MCL and CLL transformation compared to normal B cells.

Case	Chr_1	Pos_1	Chr_2	Pos_2	SV
cMCL1	1	49608581	1	143197896	Del
cMCL1	1	115736231	1	120197525	Del
cMCL1	11	77464743	11	116972587	Del
cMCL1	12	51131547	12	51132583	Dup
cMCL1	12	51131734	12	51138783	Del
cMCL1	12	51137966	12	51139304	Del
cMCL1	12	51163699	12	44907303	Inv
cMCL1	14	106329450	11	69439367	Transl
cMCL1	17	6368338	17	6370107	Inv
cMCL1	5	136919459	5	136924754	Del
cMCL2	1	64431999	1	64432560	Del
cMCL2	10	22691861	10	26115463	Inv
cMCL2	11	69452394	14	106329438	Transl
cMCL2	11	108112540	11	108120925	Dup
cMCL2	14	46423184	10	25361922	Transl
cMCL2	15	38178485	21	25273323	Transl
cMCL2	15	40586199	15	78598624	Del
cMCL2	15	77296906	15	63254366	Dup
cMCL2	19	27806661	8	26884990	Transl
cMCL2	20	26097933	20	26108799	Del
cMCL2	21	31995316	21	36204992	Inv
cMCL2	3	161212445	3	161211909	Inv
cMCL2	4	22537629	4	22537119	Inv
cMCL2	5	4336832	15	38179881	Transl
cMCL2	6	80357199	6	79447177	Dup
cMCL2	8	26876182	8	33637677	Dup
cMCL2	8	33569569	8	33588309	Inv
cMCL2	8	33652545	8	26876182	Dup
cMCL2	8	34403864	8	33644553	Dup
nnMCL2	11	69346868	14	106349763	Transl
nnMCL3	11	69346207	14	106370363	Transl
nnMCL3	13	51463625	13	51848332	Del
nnMCL3	13	51495367	13	51510743	Dup
nnMCL3	13	51510352	13	51459740	Dup
nnMCL3	13	51510362	13	51510933	Inv
nnMCL3	13	51848283	13	51495256	Inv
nnMCL3	17	56249661	17	57753598	Del

Appendix Table 8. Structural variants in MCL cases. Characterization of the different structural variants per case. Chr_1 and Pos_1 refers to the chromosome and position where the first break takes place and Chr_2 and Pos_2 refers to the chromosome and position where the second break takes place. Chr, Chromosome. Pos, Position. SV, Structural Variant. Del, Deletion. Dup, Duplication. Transl, Translocation. Inv, Inversion. The reference genome used was GRCh37.

Sample	Age	Sex	IGHV ID (%)	SOX11	ST	Emp-SP (%)	Method (SP)	SP-met (%)	Sample Timer	Morphology	Nod	Mut	GE	Tre	Sta-tus	Series	Epi.
M001	63	1	98.7	1	LN	90.0%	VI-IHC	94.9%	At diagnosis	classic/blastic	1		1	1	1	BCN	C1 MCL
M003	75	1	93.72	0	PB	96.0%	FC	94.6%	At diagnosis	pleomorphic	0	TLR2, TP53	1	1	1	BCN	C2 MCL
M004	81	1	97.22	0	PB	99.4%	FC	99.0%	Pre-treat.	classic	0	TP53		1	0	BCN	C2 MCL
M009	78	0	92.36	0	PB	98.0%	FC	90.9%	Post-treat.	small cell	0	BIRC3	1	1	1	BCN	C2 MCL
M012	59	0	96.18	1	LN	82.0%	FC	76.6%	At diagnosis	classic	1	MEF2B	1	1	0	BCN	C1 MCL
M015	70	1	96.53	0	PB	85.2%	FC	80.4%	Untreated	small cell	0		1	0	0	BCN	C2 MCL
M016	50	0	92.04	0	PB	97.2%	FC	97.9%	Pre-treat.	small cell	0		1	0	0	BCN	C2 MCL
M021	73	1	94.44	0	PB	99.0%	FC	98.2%	At diagnosis	small cell	0	TLR2	1	0	0	BCN	C2 MCL
M026	63	1	95.09	1	PB	99.5%	FC	99.1%	At diagnosis	classic	1	TP53	1	1	1	BCN	C1 MCL
M027	63	1	96.88	0	PB	98.9%	FC	97.7%	Untreated	small cell	1	TP53	1	0	0	BCN	C2 MCL
M029	68	1	98.25	1	PB	99.3%	FC	97.6%	At diagnosis	blatoid/pleomrophic	0	NOTCH2, TP53	1	1	1	BCN	C1 MCL
M030	89	0	98.26	1	PB	99.7%	FC	98.7%	At diagnosis	classic	1		1	1	1	BCN	C1 MCL
M035	54	0	100	1	LN			79.7%	At diagnosis	classic	1			1	1	BCN	C1 MCL
M039	82	0	87.68	0	PB	95.0%		90.2%	At diagnosis	small cell	0	TP53	1	1	1	BCN	C2 MCL
M047	71	1	98.63	1	LN			85.1%	At diagnosis	blastoid	1			1	0	BCN	C1 MCL
M054	71	1	97.31	1	LN			78.0%	At diagnosis	pleomorphic	1			1	1	BCN	C1 MCL
M059	68	1	99.55	1	LN			88.9%	Post-treat.	classic	1			1	1	BCN	C1 MCL
M071	51	0	99.12	1	LN			63.8%		classic	1			1	1	BCN	C1 MCL
M075	58	1	99.09	0	mucosa			80.5%		classic						BCN	C1 MCL
M076	66	1	93.68	0	PB	68.0%	FC	69.9%	At diagnosis	small cell	0		1	0	0	BCN	C2 MCL
M078	68	0	99.55	1	LN			87.2%	Post-treat.	pleomorphic	1	TP53		1	1	BCN	C1 MCL
M088	66	1	100	1	spleen			85.9%	Post-treat.	classic	0	TP53, WHSC1	1	1	1	BCN	C1 MCL
M106	67	1	100	1	LN			75.4%	At diagnosis	classic	1			1	1	BCN	C1 MCL
M108	63	1	100	1	tissue			88.5%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M114	61	0	100	1	spleen			75.8%			1		1	1	1	BCN	C1 MCL
M122	70	1	99.55	1	LN			77.9%	At diagnosis	classic	1	MEF2B	1	1	0	BCN	C1 MCL
M163	43	1	92.04	0	LN			67.0%	At diagnosis	classic	1			0	0	BCN	C2 MCL
M195	69	1	93.75	0	PB	93.0%		95.2%	At diagnosis		1	TP53		1	1	BCN	C2 MCL
M197	64	1	96.53	0	PB	86.0%	FC	80.5%	At diagnosis	blastoid	0	TP53		1	1	BCN	C2 MCL
M198	49	0	100	1	PB	85.0%	FC	72.8%	At diagnosis	classic	1	WHSC1	1	1	1	BCN	C1 MCL
M199	65	1	94.44	1	PB	89.0%	FC	81.7%	Pre-treat.	classic	1	TP53, WHSC1	1	1	1	BCN	C1 MCL
M200	70	1	98.61	1	PB	90.0%	FC	81.3%	At diagnosis	blastoid	1	NOTCH2, TP53, WHSC1	1	1	1	BCN	C1 MCL
M201	85	1	97.19	0	PB	90.0%	FC	92.6%	Pre-treat.	classic	0		1	1	1	BCN	C2 MCL
M202	51	1	99.65	1	PB	72.0%	FC	67.7%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M203	58	1	99.33	1	PB	94.0%	FC	94.2%	At diagnosis	classic	0		1	1	0	BCN	C1 MCL
M205	64	1	100	1	PB	80.0%	FC	69.5%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M206	64	0	100	1	PB	92.0%	FC	97.7%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M207	69	1	98.95	1	PB	68.0%	FC	63.0%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M208	63	1	98.96	1	PB	62.0%	FC	63.9%	At diagnosis	classic	1			1	0	BCN	C1 MCL
M209	41	1	98.96	1	PB	56.0%	FC	53.6%	At diagnosis	blastoid	1	WHSC1		1	1	BCN	C1 MCL
M210		1		1	PB	81.0%	FC	80.3%	At diagnosis	classic	1		1		0	BCN	C1 MCL
M211	56	1	97.57	0	PB	86.0%	FC	84.2%	Pre-treat.	classic	0			1	0	BCN	C2 MCL
M212		1		1	PB	84.0%	FC	79.9%	At diagnosis		1	BIRC3	1	1	0	BCN	C1 MCL

Appendix Table 9. Clinical and biological features of the MCL patients included in **Study 3**, (see continuation and legend in next page).

Sample	Age	Sex	IGHV ID (%)	SOX11	ST	Emp-SP (%)	Method (SP)	SP-met (%)	Sample Timer	Morphology	Nod	Mut	GE	Tre	Status	Series	Epi.
M213	70	1	100	0	PB	68.0%	FC	51.1%	Pre-treat.	classic	0	TP53, WHSC1		0	0	BCN	C1 MCL
M214	57	1	97.57	1	PB	84.0%	FC	77.1%	At diagnosis	classic	0	WHSC1	1	1	0	BCN	C1 MCL
M215	54	1	100	1	PB	60.0%	FC	39.5%	At diagnosis	clasiic	1			1	0	BCN	C1 MCL
M218	69	1	94.74	0	PB	100%	FC	100.0%	At diagnosis		0		1	0	0	BCN	C2 MCL
M220	76	1	97.57	0	BM			71.1%	At diagnosis						1	BCN	C2 MCL
M221	54	1	100	1	BM	89.5%	FC	87.8%	At diagnosis			BIRC3			1	BCN	C1 MCL
M224	72	1	100	0				92.7%	At diagnosis			MEF2B			1	BCN	C1 MCL
M225	71	1	100	1	BM	90.0%	FC	68.3%	At diagnosis					1	1	BCN	C1 MCL
M227	61	1	97.76					43.4%	At diagnosis						0	BCN	C1 MCL
M228	68	1	91.67	0	BM			52.6%	At diagnosis						0	BCN	C2 MCL
V17724		1	98.7	1	LN	90.0%	VI-IHC	67.4%			1					Kiel	C1 MCL
V17782		0	100	1	LN	80.0%	VI-IHC	76.2%	At diagnosis		1					Kiel	C1 MCL
V17824	75	1	99.1	1	LN	90.0%	VI-IHC	62.7%	At diagnosis		1			1	1	Kiel	C1 MCL
V17832		1	92.3		LN	80.0%	VI-IHC	72.9%	At diagnosis		1					Kiel	C1 MCL
V17985		1	99.6	1	LN	90.0%	VI-IHC	80.1%	At diagnosis		1					Kiel	C1 MCL
V18094	62	1	97.5	1	LN	80.0%	VI-IHC	81.1%	At diagnosis		1			0	0	Kiel	C1 MCL
V18131		1	100	0	LN	70.0%	VI-IHC	59.3%			1					Kiel	C2 MCL
V18213		1	98.96		LN	70.0%	VI-IHC	68.6%	At diagnosis		1					Kiel	C1 MCL
V18258	69	1	100		LN	80.0%	VI-IHC	72.0%	At diagnosis		1			1	0	Kiel	C1 MCL
V18348		1	100	1	LN	90.0%	VI-IHC	73.8%			1					Kiel	C1 MCL
V18485		1	99.1	1	LN	90.0%	VI-IHC	80.5%			1					Kiel	C1 MCL
V18498	76	0	98.96	1	LN	90.0%	VI-IHC	77.8%	Pre-treat.		1			1	1	Kiel	C1 MCL
V18505		1	100	1	LN	90.0%	VI-IHC	79.5%			1					Kiel	C1 MCL
V18562	55	1	100	1	LN	90.0%	VI-IHC	82.8%	At diagnosis		1			1	1	Kiel	C1 MCL
V18573	56	0	100	1	LN	90.0%	VI-IHC	77.3%	At diagnosis		1			0	0	Kiel	C1 MCL
V18587	73	1	100		LN	90.0%	VI-IHC	70.8%	At diagnosis		1			1	0	Kiel	C1 MCL
V18630	84	1	94.3		LN	90.0%	VI-IHC	77.9%	At diagnosis		1			1	0	Kiel	C1 MCL
V18633		1	97.8	1	LN	70.0%	VI-IHC	76.2%	At diagnosis		1					Kiel	C1 MCL
V18663		1	100	1	LN	80.0%	VI-IHC	80.3%	Pre-treat.		1					Kiel	C1 MCL
V18711	34	1	100	1	LN	90.0%	VI-IHC	83.1%	Pre-treat.		1					Kiel	C2 MCL
V18719		1	98.7	1	LN	80.0%	VI-IHC	78.5%			1					Kiel	C1 MCL
V18726		1	99.1	0	LN	80.0%	VI-IHC	75.8%			1					Kiel	C2 MCL
V18737	60	1	100	1	LN	80.0%	VI-IHC	76.2%	At diagnosis		1				0	Kiel	C1 MCL
V18750		1	99.6	1	LN	90.0%	VI-IHC	67.6%			1					Kiel	C1 MCL
V18917	85	1	99.6	1	LN	80.0%	VI-IHC	74.4%	At diagnosis		1			0	1	Kiel	C1 MCL
V19080	57	0	99.1		LN	80.0%	VI-IHC	68.8%	At diagnosis		1			0	0	Kiel	C1 MCL
V19164		1	100	1	LN	90.0%	VI-IHC	88.9%	At diagnosis		1					Kiel	C1 MCL
V19165		1	99.6	0	LN	80.0%	VI-IHC	67.5%			1					Kiel	C1 MCL
V19203		1	99.1	1	LN	90.0%	VI-IHC	80.1%	Pre-treat.		1					Kiel	C1 MCL

Appendix Table 9 (continuation). Clinical and biological features of the MCL patients included in **Study 3**. Sex: 0-female, 1-male; SOX11: 0-not expressed, 1-expressed; Nodal: 0-non-nodal, 1-nodal; Treatment: 0-no treatment, 1-treated; Status: 0-alive, 1-dead, 2-unknown. IGVH ID, IGVH Identity; ST, Sample Type; Emp-SP, Empiric sample purity; SP-met, Sample purity based on DNA methylation; Nod, Nodal; Mut, Murations; GE, gene expression data; Tre, Treatment; Epi, epigenetic classification; PB, peripheral blood; LN, lymphoid node; BM, bone marrow; Pre-/Post-treat., Pre-/Post- treatment; VI-IHC, Visual inspection-IHC.

MCL or NBC samples	SOX11 status	chr2:6,477,577*	chr2:6,477,615*	chr:6,484,702*	chr2:6,484,925*
M027	SOX11-neg	96	82	94	89
M039	SOX11-neg	78	44	12	44
M016	SOX11-neg	96	77	90	59
M163	SOX11-neg	93	61	82	ND
M197	SOX11-neg	97	84	59	91
M021	SOX11-neg	59	21	22	26
M218	SOX11-neg	55	34	ND	ND
M220	SOX11-neg	95	78	91	94
M228	SOX11-neg	91	75	79	80
M003	SOX11-neg	95	75	91	92
M029	SOX11-pos	3	8	21	14
M030	SOX11-pos	1	7	2	3
M035	SOX11-pos	30	20	17	21
M071	SOX11-pos	32	24	ND	ND
M078	SOX11-pos	7	10	10	10
M088	SOX11-pos	12	13	9	ND
M106	SOX11-pos	27	27	ND	ND
M108	SOX11-pos	ND	ND	7	10
M114	SOX11-pos	14	17	14	17
M198	SOX11-pos	21	19	18	20
M207	SOX11-pos	66	51	25	30
M212	SOX11-pos	ND	ND	ND	16
NBC1	NA	95	85	96	95
NBC2	NA	95	84	96	96
NBC3	NA	85	68	73	69
NBC4	NA	91	80	37	97

*DNA methylation levels are expressed as percentage.

Appendix Table 10. DNA methylation of the CpGs located in the *SOX11* distal enhancer region in MCL primary cases detected by bisulfite pyrosequencing. SOX11-neg, SOX11-negative. SOX11-pos, SOX11-positive. NA, Non applicable.

Accession	Score	Coverage	Peptide	PSM in Peak 3	MW	Gene	Entry
Q15853-2	69.5	21.6	9	40	338	36.2	USF2
B1AQP1	27.57	22.7	6	29	282	31	USF1
B8ZZU6	31.1	23.39	6	15	389	42.3	ATF2
P20585	15.65	8.09	9	14	1137	127.3	MSH3
Q9H7L9	13.34	7.32	3	10	328	38.1	SUDS3
Q9Y6D9	7.96	13.23	7	8	718	83	MAD1L1
O00257	10.75	11.79	6	7	560	61.3	CBX4
A0A087WWZ6	8.57	5.18	5	7	1177	134.2	PRDM11
P35716	18.29	13.83	3	7	441	46.7	SOX11
Q5T7U1	6.4	12.25	5	7	457	52.3	GTF3C5
Q6NT76-3	13.97	12.92	4	6	418	47	HMBOX1
Q9HCK0	7.02	8.39	4	6	441	49.9	ZBTB26
Q9UJU2	6.48	10.53	4	5	399	44.2	LEF1
Q9UJU2-6	6.48	10.88	4	5	386	42.6	LEF1
A0A1B0GUV5	3.55	30.17	4	5	116	13.5	TCF7L2
P36402-9	5.14	13.42	4	5	365	39.7	TCF7
Q9HCE3	3.41	5.07	4	5	1301	141.6	ZNF532
Q16236-3	10.19	6.53	4	5	582	65.3	NFE2L2
P17535	10.65	10.66	3	4	347	35.2	JUND
Q8N9N5-7	9.69	4.72	2	4	508	55.5	BANP
O75190	10.09	13.8	3	4	326	36.1	DNAJB6
Q4LE39-3	8.39	3.82	3	4	1153	130.3	ARID4B
O14948-4	6.4	10.71	4	4	280	31.5	TFEC

Appendix Table 11. Proteins present at the specific chromatin accessible peak 3 of *SOX11* enhancer region. The proteins with more than three unique peptides are listed. Score, indicates the sum of the scores of the individual peptides. Coverage, indicate the percentage of protein sequences identified. Peptides, indicates the number of distinct peptide sequences identified per protein. PSMs. Peptide Spectrum Matches, indicates the total number of identified peptide sequences for the protein, including those redundantly identified. MW (KDa), molecular weight of the protein in kiloDalton. Gene, gene name and entry are indicated.

Accession	Score	Coverage	Peptide	PSM in Peak 1	MW	Gene	Entry
Q02548-2	65.2	29.28	7	26	362	38.9	PAX5
Q5T5X7	46.89	13.77	10	25	828	94.4	BEND3
Q9NZI7	35.51	15.56	8	22	540	60.5	UBP1
P51572	16.84	22.76	6	10	246	28	BCAP31
Q99592	24.49	10.92	3	10	522	58.3	ZBTB18
P28324-2	9.25	12.1	5	9	405	44.6	ELK4
Q9NX63	14.59	16.74	3	9	227	26.1	CHCHD3
B4DJ81	11.76	17.84	8	9	611	66.9	NDUFS1
Q9H0U3	17.58	8.06	3	8	335	38	MAGT1
Q8NHM5-4	8.35	6.96	7	8	1265	144.7	KDM2B
P10606	7.93	15.5	3	7	129	13.7	COX5B
Q4VC44-2	9.63	5.59	2	7	715	80	FLYWCH1
Q14165	12.9	14.73	3	7	292	32.2	MLEC
Q02447-4	10.04	3.55	2	7	479	52	SP3
Q15723-4	7.03	8.13	3	7	504	54.2	ELF2
Q9NVH2-4	2.22	2.96	2	6	913	101.1	INTS7
I3L072	8.37	7.31	3	6	547	60.8	C17orf80
P41162	10.08	7.23	2	6	512	57	ETV3
Q00577	7.96	13.98	3	5	322	34.9	PURA
Q9H9B4	8.86	13.66	3	5	322	35.6	SFXN1
P31930	8.21	8.13	4	5	480	52.6	UQCRC1
O94874	7.94	3.9	3	5	794	89.5	UFL1
O15062	10.09	6.5	4	5	677	74.2	ZBTB5
E7EVG2	8.24	4.11	5	5	1460	168.3	PBRM1
A0A140T9R1	8.43	12.89	4	5	357	39.8	FLOT1
Q86UP2-2	8.61	2.85	4	4	1300	149.5	KTN1
Q96QR8	10.23	7.69	2	4	312	33.2	PURB
Q8TCJ2	6.97	5.08	3	4	826	93.6	STT3B
Q9HD20-2	8.57	3.96	4	4	1086	121	ATP13A1
P53567	2.14	17.33	2	4	150	16.4	CEBPG
Q9HCM1	2.79	3.15	4	4	1747	194.7	KIAA1551
Q8IY37	3.92	2.16	2	4	1157	129.5	DHX37
O75477	11.17	13.01	3	4	346	38.9	ERLIN1
Q9H3K2	4.14	9.28	3	4	345	37.2	GHITM
Q96L73-2	2.54	2.6	4	4	2427	267.2	NSD1
P12755	1.76	4.81	3	4	728	80	SKI
A0A0G2JH46	5.94	10.92	2	4	229	25.8	HLA-DRA
Q9Y4X4-3	6.45	9.06	2	4	309	33.2	KLF12
J3QLD9	6.85	10.98	4	4	428	47.1	FLOT2

Appendix Table 12. Proteins present at the specific chromatin accessible peak 1 of *SOX11* enhancer region. The proteins with more than three unique peptides are listed. Score, indicates the sum of the scores of the individual peptides. Coverage, indicate the percentage of protein sequences identified. Peptides, indicates the number of distinct peptide sequences identified per protein. PSMs, Peptide Spectrum Matches, indicates the total number of identified peptide sequences for the protein, including those redundantly identified. MW (KDa), molecular weight of the protein in kilodalton. Gene, gene name and entry are indicated.

Motif ID	chr	start	end	Strand	*p*-value	Region	Chrom.accessible
MA0014.2	chr2	6351148	6351166	+	1.44E-06	MCL superenhancer	
MA0014.2	chr2	6326521	6326539	-	3.33E-06	MCL superenhancer	
MA0014.2	chr2	5690855	5690873	-	4.66E-06	Promoter	
MA0014.2	chr2	6333974	6333992	+	6.66E-06	MCL superenhancer	
MA0014.3	chr2	6333648	6333659	+	6.73E-06	MCL superenhancer	
MA0014.3	chr2	5696937	5696948	+	1.33E-05	Promoter	
MA0014.2	chr2	6329482	6329500	-	1.68E-05	MCL superenhancer	
MA0014.2	chr2	5696563	5696581	+	2.09E-05	Promoter	
MA0014.2	chr2	6354995	6355013	+	2.56E-05	MCL superenhancer	
MA0014.2	chr2	5691418	5691436	+	2.63E-05	Promoter	
MA0014.3	chr2	6350296	6350307	-	3.13E-05	MCL superenhancer	
MA0014.2	chr2	6351022	6351040	+	3.37E-05	MCL superenhancer	
MA0014.3	chr2	5693618	5693629	+	4.20E-05	Promoter	
MA0014.2	chr2	5691887	5691905	+	4.40E-05	Promoter	
MA0014.2	chr2	6345407	6345425	+	4.77E-05	MCL superenhancer	Peak 3
MA0014.3	chr2	6340392	6340403	+	4.96E-05	MCL superenhancer	
MA0014.3	chr2	6337270	6337281	+	5.34E-05	MCL superenhancer	Peak 1
MA0014.3	chr2	5693330	5693341	+	5.79E-05	Promoter	
MA0014.2	chr2	6356165	6356183	-	5.85E-05	MCL superenhancer	
MA0014.3	chr2	5697073	5697084	-	5.92E-05	Promoter	
MA0014.3	chr2	6353039	6353050	-	5.92E-05	MCL superenhancer	Peak 4
MA0014.2	chr2	5696707	5696725	+	6.39E-05	Promoter	
MA0014.3	chr2	6325864	6325875	+	7.59E-05	MCL superenhancer	
MA0014.2	chr2	5692797	5692815	-	8.20E-05	Promoter	
MA0014.3	chr2	6326518	6326529	-	8.54E-05	MCL superenhancer	
MA0014.3	chr2	6351158	6351169	+	8.54E-05	MCL superenhancer	

Appendix Table 13. Results from finding individual motif occurrences (FIMO) for the PAX5 transcription factor at the *SOX11* promoter (chr2:5,689,137-5,702,215; GRCh38), and SOX11-positive MCL superenhancer (chr2:6,325,427-6,356,576; GRCh38). The chromatin accessible peak targeted is indicated.