Genetic and molecular characterization of mantle cell lymphoma

by

Prasath Pararajalingam

B.Sc., Western University, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2023

Declaration of Committee

Name:	**Prasath Pararajalingam**
Degree:	**Doctor of Philosophy (Molecular Biology and Biochemistry)**
Title:	**Genetic and molecular characterization of mantle cell lymphoma**
Committee:	**Chair: Jonathan Choy** Professor, Molecular Biology and Biochemistry
	Ryan D. Morin Supervisor Associate Professor, Molecular Biology and Biochemistry
	Sharon M. Gorski Committee Member Professor, Molecular Biology and Biochemistry
	David W. Scott Committee Member Associate Professor, Medicine
	Sophie Sneddon Examiner Lecturer, Molecular Biology and Biochemistry
	Ash Alizadeh External Examiner Professor, Medicine - Oncology Stanford

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Despite advances in treatment and disease management, Mantle Cell Lymphoma (MCL) remains an incurable disease where patients typically succumb to the disease within 5 years of diagnosis. MCL is clinically and genetically a heterogenous disease which is reflected by the fact that there are two recognized subtypes: conventional MCL (cMCL) and non-nodal leukemic MCL (nnMCL), each thought to represent distinct evolutionary paths to lymphoma. Conventional MCL is also genetically and clinically heterogeneous where varying frequencies of simple somatic mutations, copy number variations and structural variations are observed across tumours. Conventional MCL can be robustly divided into three risk groups based on patient and histopathological parameters using the MIPI score, or by measuring gene expression of proliferation associated genes. In both cases, however, the underlying genetic cause for these clinical differences is largely unknown. In this thesis, I will begin by describing the integrative analysis where I used targeted sequencing, whole-genome sequencing, and RNA-sequencing to uncover novel driver genes. From analysing newly sequenced exomes along with published exomes, I uncovered recurrent mutations in *HNRNPH1*, *EWSR1* and *DAZAP1*, thus implicating RNA-metabolism in MCL carcinogenesis. Furthermore, we validated the hypothesis that *HNRNPH1* mutations serve to increase *HNRNPH1* protein concentration by disrupting splicing of its own transcript. Next, using whole-genomes and RNA-sequencing, I found recurrent gains/amplifications and structural variations on chromosome 10 that served to increase *ABI1* expression. With this expanded catalogue of the genetic landscape of MCL, I applied non-negative matrix factorization to dissect cMCL and nnMCL cases into more refined subgroups. This resolved five genetic clusters that were prognostic for overall survival, progression-free survival, and disease-specific survival. This included the observations that within cMCL, patients with aberrant somatic hypermutation exhibited superior prognosis. In contrast, patients with harboured *TERT* amplifications, *TP53* alterations or *CCND1* 3'UTR mutations had worse prognosis. This work has uncovered new genes that may form the basis for therapeutic intervention and delineates genetically-defined clinical subsets that could form the basis for applying personalized medicine in MCL.

Keywords: Mantle cell lymphoma; cancer genomics; driver genes; whole-genome; transcriptome; NMF clustering

Acknowledgements

This thesis wouldn't have been possible without the support from my colleagues in the Morin lab and my supervisor.

First, I am profoundly grateful to my supervisor, Dr. Ryan D. Morin, for his guidance, expertise, and support. His feedback and dedication have been invaluable to my academic growth.

I would also like to sincerely thank the members of my research and defense committee, Dr. Sharon M. Gorski, and Dr. David W. Scott. Their insights and suggestions have been instrumental to my growth as a researcher. I would also like to send my thanks to Dr. Jonathan Choy, Dr. Sophie Sneddon, and Dr. Ash Alizadeh for taking the time to participate in my thesis examination.

My heartfelt appreciation goes towards my family and friends who have given me nothing up unconditional support and encouragement through this difficult journey. Their belief in my abilities have given me the motivation to succeed in graduate school. I also must give special thanks to my grandfather who instilled in me the passion for learning.

Lastly, I would like to express my gratitude to the many members of the Morin lab, past and present, who, without their knowledge and generosity, could not have completed this thesis. The great people I have met while here are some of the most intelligent people I have ever met and served as an inspiration for my own development.

Table of Contents

Declaration of Committee ii
Ethics Statement iii
Abstract iv
Acknowledgements vi
Table of Contents vii
List of Tables x
List of Figures xi
List of Acronyms xii

Chapter 1. Introduction 1
1.1. Histological characteristics of Mantle Cell Lymphoma 1
1.1.1. Cell-of-origin of MCL 1
1.1.2. MCL subtypes 3
1.2. Epidemiology and Aetiology 5
1.3. Genetics of Mantle Cell Lymphoma 6
1.3.1. Primary Aberration: *CCND1*::*IGH* 6
1.3.2. MCL Drivers Beyond *CCND1*::*IGH* 7
1.4. Current Treatment Strategies 10
1.5. Risk stratification 11
1.5.1. Prognostic indices 11
1.5.2. Gene expression-based classifications 12
1.6. Problem statement 14

Chapter 2 15
Coding and non-coding drivers of mantle cell lymphoma identified through exome and genome sequencing 15
2.1. Abstract 15
2.2. Introduction 16
2.3. Results 16
2.3.1. Resolving the frequency of SSMs and recurrently mutated genes in MCL 16
2.3.2. Novel mutation patterns in MCL 19
2.3.3. *HNRNPH1* intronic mutations disrupt HNRNPH1 binding motifs 22
2.3.4. *HNRNPH1* splicing is associated with inferior outcomes in MCL 25
2.3.5. Mutations and alternative splicing influence HNRNPH1 protein expression in MCL 28
2.3.6. Common *HNRNPH1* mutations disrupt productive splicing and translation 29
2.4. Discussion 32
2.5. Methods 35
2.5.1. Study design and sequencing 35
2.5.2. Exome data analysis 35
2.5.3. Recurrence analysis 36

2.5.4. Targeted sequencing, whole genome sequencing and data consolidation 36
2.5.5. Whole genome sequencing library construction and sequencing 37
2.5.6. Ribosomal RNA depletion RNA sequencing library construction and sequencing 37
2.5.7. *HNRNPH1* variant calling in Rule et al.[178], Agarwal et al.[128] and Khodadoust et al.[129] 39
2.5.8. RNA-seq analysis of MCL 40
2.5.9. Protein-RNA interactions of HNRNPH1 40
2.5.10. Cell-based experiments 41
2.5.11. Digital PCR 41
2.5.12. Plasmids 41
2.5.13. HEK transfections 42
2.5.14. Tissue microarray and immunohistochemistry 42
2.5.15. Statistical analysis 42

Chapter 3 44
Landscape of copy number and structural variations in MCL 44
3.1. Introduction 44
3.2. Results 46
3.2.1. Clinical and molecular characteristics of MCL tumours 46
3.2.2. Recurrent structural variations in B-cell lymphomas 48
3.2.3. *CCND1*::*IGH* rearrangements in MCL 50
3.2.4. *ABI1* frequently altered by structural variations in MCL 54
3.2.5. Integrative analysis of copy number variations 58
Deregulation of RB1 via the CCND1-CDK4/6 axis 60
Integrative analysis of CNVs and expression 62
3.3. Methods 68
3.3.1. Read mapping 68
3.3.2. Copy number variation analysis 68
3.3.3. Structural variation analysis 69
3.3.4. Quantification of gene expression 70
3.3.5. Statistical analyses 71

Chapter 4 72
Consensus clustering reveals MCL genetic subgroups 72
4.1. Introduction 72
4.2. Results 75
4.2.1. Somatic hypermutation in MCL 75
4.2.2. Genetic clustering in MCL 79
Characteristics of MCL Cluster 1 81
Characteristics of MCL Cluster 2 82
Characteristics of MCL Cluster 3 83
Characteristics of MCL Cluster 4 83
Characteristics of MCL Cluster 5 83

4.3. Methods 84
4.3.1. Detecting Simple Somatic Mutations 84
4.3.2. SSM mutational signature analysis 85
4.3.3. Immunoglobulin gene usage 85
4.3.4. Consensus clustering using non-negative matrix factorisation 86
4.3.5. Survival Analyses 87

Chapter 5 88
General Discussion 88
5.1. Summary of Research Findings 88
5.2. RNA-splicing dysregulation 91
5.3. Towards a new classification scheme for MCL 92
5.4. Future Directions 94
5.4.1. From Clusters to a Classification System 94
5.4.2. CRISPR screens for essential genes 95
5.4.3. RNA-binding proteins in MCL and B-cell lymphomas 96
5.4.4. Indolent MCL 97

References 99

Appendix A. Chapter 2 Supplemental Figures 115

Appendix B. Supplemental Data File 117

List of Tables

Table 1-1 Clinical and pathological differences between nnMCL and cMCL 4
Table 2-1 Recurrently mutated genes in MCL exomes. 18
Table 3-1 Clinical and molecular characteristics of cMCL in the BC cohort 47
Table 3-2 Recurrently amplified and deleted regions in MCL 60

List of Figures

Figure 1-1 History of classification in MCL from 1970 to 2016. Created with Biorender.com 3
Figure 1-2 Schematic diagram of MCL subtypes and model of tumour development 5
Figure 2-1 Overview of patient samples utilized in individual samples. 19
Figure 2-2 Recurrent mutations in combined MCL cohort. 20
Figure 2-3 Spatial distribution of mutations in DLBCL and MCL of EWSR1, and DAZAP1. 21
Figure 2-4 *HNRNPH1* mutations in MCL cluster near exon 4 in poly-G motifs..... 23
Figure 2-5 *HNRNPH1* mutations influence expression of HNRNPH1. 25
Figure 2-6 *HNRNPH1* splicing is independently associated with survival in MCL. ... 27
Figure 2-7 Evaluating HNRNPH expression by tissue microarray. 29
Figure 2-8 Mutations in *HNRNPH1* prevent negative regulation via nonsense-mediated decay..... 32
Figure 3-1 Q-Q plots of model fits for BL, DLBCL, FL, CLL, MM and MCL 49
Figure 3-2 Oncoplot of recurrently altered bins across DLBCL, FL, CLL, MCL, BL and MM 50
Figure 3-3 *CCND1*::*IGH* rearrangements in BC cMCL tumours..... 51
Figure 3-4 aSHM counts in relation to *CCND1*::*IGH* rearrangements in BC and Barcelona MCL tumours 52
Figure 3-5 *CCND1* and *SOX11* expression of B-cell lymphomas and naïve B cells 54
Figure 3-6 Chromosomal alterations affecting *ABI1* on chromosome 10..... 57
Figure 3-7 CNV incidence rates in MCL 59
Figure 3-8 Structural variations overlapping *CCND1* 3'UTR in BC and Barcelona cohorts..... 61
Figure 3-9 Variance-stabilized expression differences between amp(3q26.1) mutated, unmutated and normal samples. 63
Figure 3-10 Variance-stabilizied expression differences between amp(7p22.2) mutated, unmutated and normal samples. 64
Figure 3-11 Variance-stabilizied expression differences between amp(8q24.21) mutated, unmutated and normal samples. 65
Figure 3-12 Copy number gains and amplifications overlapping 15q21.26 GISTIC peak..... 66
Figure 4-1 Algorithmic view of SSM variant calling pipeline for MCL genomes..... 76
Figure 4-2 Putative aSHM in MCL 78
Figure 4-3 NMF clustering schematic and rank selection metrics..... 79
Figure 4-4 Survival impact of NMF clusters..... 80
Figure 4-5 Heatmap of tumour and mutational features of NMF genetic clusters 82
Figure 5-1 MCL NMF clustering described by Yi et al.[242] 93

List of Acronyms

ABC	Activated B cell
AIC	Akaike information criterion
AID	Activation-induced cytidine deaminase. Also known as AICDA.
ALP	Alkaline phosphatase
Amp	Amplification
ASCT	Autologous stem cell transplantation
aSHM	Aberrant somatic hypermutation
BAM	Binary alignment map
BC	British Columbia
BCR	B-cell receptor
BL	Burkitt's Lymphoma
BND	Breakpoint
BTK	Bruton's tyrosine kinase
BWA	Burrows-Wheeler aligner
CCL	Centrocytic lymphoma
CCND1	Cyclin D1
CLL	Chronic Lymphocytic Leukemia
cMCL	Conventional Mantle Cell Lymphoma
CNV	Copy Number Variation
COO	Cell-of-origin
COSMIC	Catalogue of somatic mutations in cancer
ddPCR	Digital droplet polymerase chain reaction
Del	Deletion
DLBCL	Diffuse Large B-cell Lymphoma
DSS	Disease-specific survival
ECOG	Eastern Cooperative Oncology Group
EGFP	Enhanced green fluorescent protein
ENCODE	Encyclopedia of DNA elements
FF	Fresh frozen
FFPE	Formalin-fixed paraffin embedded
FISH	Fluorescent in-situ hybridisation
FL	Follicular Lymphoma
FLIPI	Follicular lymphoma International Prognostic Index
GAMBL	Genomic Analysis of Mature B-cell Lymphoma
GATK	Genome analysis toolkit
GC	Germinal centre
GC content	Guanine-Cytosine content
GCB	Germinal centre B cell
GLM	Generalized linear model
gnomAD	Genome aggregation database
HA	Hemagglutinin
HEK	Human embryonic kidney cells. Also called HEK293.
HSCT	Hematopoietic stem-cell transplantation
IDL	Lymphocytic lymphoma of intermediate differentiation
IGH	Immunoglobulin Heavy chain
IGK	Immunoglobulin Light chain Kappa
IGL	Immunoglobulin Light chain Lambda
IPI	International Prognostic Index

ISMCN	In-situ mantle cell neoplasia
LDH	Lactate dehydrogenase
LINE	Long interspersed nuclear element
LOH	Loss-of-heterozygosity
LTR	Long terminal repeat
MCL	Mantle Cell Lymphoma
MIPI	Mantle cell lymphoma International Prognostic Index
MM	Multiple Myeloma
mTOR	Mammalian target of rapamycin
NFkB	Nuclear factor kappa B
NHL	Non-Hodgkin lymphoma
NMD	Non-sense mediated decay
NMF	Non-negative matrix factorisation
nnMCL	Non-nodal, leukemic Mantle Cell Lymphoma
OS	Overall survival
PCA	Principle component analysis
PCR	Polymerase chain reaction
PFS	Progression-free survival
PI3K	Phosphatidylinositol-3 kinase
R	R programming language
R-CHOP	Rituximab plus cyclophosphamide, vincristine, doxorubicin, prednisone
R-DHAP	Rituximab plus dexamethasone, cytarabine, cisplatin
R-HyperCVAD	Rituximab plus hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone
R-MA	Rituximab plus methotrexate, cytarabine
RR	Relapsed-refractory
RNA-Seq	Ribonucleic acid sequencing
RSS	Recombination signal sequence
SHM	Somatic hypermutation
SINE	Short interspersed nuclear element
SNV	Single Nucleotide Variation
SSM	Simple Somatic Mutation
SV	Structural Variation
TMA	Tissue microarray
UTR	Untranslated region
V(D)J	Variable, diversity, joining segments
VAF	Variant allele frequency
WBC	White blood cell count
WGS	Whole-genome sequencing
WHO	World Health Organization

Chapter 1.

Introduction

Although the term Mantle Cell Lymphoma (MCL) was first coined in 1992 by the International Lymphoma Study Group, the disease was described in one form or another by various groups dating back to the 1970s[1]. Improvements in tools and techniques in the fields of pathology, histology, and genomics spanning five decades have brought into focus a clinically heterogeneous collection of mature B-cell cancers that are today called mantle cell lymphoma.

1.1. Histological characteristics of Mantle Cell Lymphoma

1.1.1. Cell-of-origin of MCL

Initial descriptions of MCL were made by Lennart et al, in the 1970s and was described as a non-Hodgkin lymphoma (NHL) comprising of small cells with irregular and cleaved nuclei with a diffuse morphology[2]. Termed diffuse germinocytoma and later centrocytic lymphoma (CCL), Lennart distinguished CCL from other non-Hodgkin lymphomas by the lack of “lymphoblasts and prolymphocytes” (centroblasts) which were common in follicular lymphomas, and the lack of proliferation centres commonly seen in chronic lymphocytic leukemia (CLL). Electron microscopy of CCL cells led Lennart to declare that these cells originate from “germinocytes” (centrocytes) owing to their similarity to cells within reactive germinal centres.

Berard et al[3] described a group of NHL tumours that could not easily be classified or were poorly differentiated leading to the researchers dubbing these lymphoma as “lymphocytic lymphoma of intermediate differentiation” (IDL). In lymph node sections, IDL resembled CCL with cells displaying either diffuse or nodular architecture around germinal centres comprising of small- to medium-sized cells with irregularly shaped or cleaved nuclei[4]. Immunologic studies of IDL and CCL showed a striking similarity between both diseases[5–7]. While most B-cell lymphomas show κ surface immunoglobulins, the majority of IDL cases expressed λ light chains[8]. IDL tumour cells

expressed pan B-cell markers typical of NHLs. Tumours also expressed T-cell marker CD5, a trait unique to IDL and CCL. Membrane alkaline phosphatase (ALP) activity was observed in 50% of IDL and CCL tumours. This enzyme is mostly expressed in normal lymphoid cells residing in primary follicles and the mantle zones of secondary follicles, which suggested that these tumours derive from a cell residing in this region, in contrast to the centrocytic origin proposed by Lennart.

Molecular studies revealed a common cytological characteristic within IDL and CCL tumours, which was the t(11;14)(q13;q32) translocation. This translocation was first discovered in the 1970s and was thought to be a rare occurrence in non-Hodgkin B-cell lymphomas but was discovered to recurrently occur in IDL and CCL tumours[9]. Similar rearrangements were discovered in other B-cell lymphomas which resulted in joining an oncogene and the highly active heavy chain immunoglobulin (IGH) locus on 14q32[10]. It was soon revealed the oncogene in question was a novel gene encoding a cyclin protein which was later dubbed cyclin D1 (*CCND1*)[11,12]. Constitutive activation by the IGH regulatory elements leads to high expression of *CCND1* mRNA and protein in both IDL and CCL[11].

Histological, immunological, and cytogenetic evidence all suggested commonality between IDL, CCL and other related B-cell malignancies as well as a possible cell-of-origin in the follicle mantle. This eventually led to the unification of these lymphomas into a single entity that was named mantle cell lymphoma with diagnostic criteria for MCL initially established in 1994[13]. Typical morphology/cytology or typical immunophenotype were required for making an MCL diagnosis. MCL cytological characteristics included small- or medium-sized cells with irregular nuclei, fine and condensed chromatin, small or absent nucleoli, pale or scant cytoplasm, rare or absent centroblasts and typically low mitotic rate. MCL morphology was highly variable and could include a nodular growth pattern with or without residual germinal centres (GC) or diffuse morphology with or without residual GC. Typical MCL immunophenotype included common B cell, T cell markers CD5 positivity, increased prevalence of λ light-chain compared to κ light chains, BCL2 positivity and BCL6 negativity. Owing to the t(11;14)(q13;q32) translocation, *CCND1* mRNA and protein overexpression is observed in >95% of MCLs. Notably, a small subset of MCL patients are negative for CCND1 and t(11;14)(q13;32), but are indistinguishable from *CCND1*::*IGH* positive MCL in terms of genetics, clinical presentation and evolution[14,15]. These cases are often observed with cyclin D2 (*CCND2*)

or cyclin D3 (*CCND3*) overexpression and an associated translocation event involving heavy chain or light chain immunoglobulin loci[14,16–20]. SOX11, a neuronal transcription factor, was found to be overexpressed in the nucleus and is unique to MCL but the expression of SOX11 varies within MCL[21–24].

1.1.2. MCL subtypes

It was recognized early on of the variable presentation of mantle cell lymphoma in terms of morphology and cytology. The apparent differences were later established to represent clinically important subsets of MCL[25,26] (Figure 1-1). Two aggressive MCL variants have been officially recognized by the World Health Organization (WHO) which are the blastoid and pleomorphic variants[27–29]. Tumours of both variants exhibit high proliferation rates but differ in terms of morphology and cytology. Blastoid cells resemble lymphoblasts with dispersed chromatin. Pleomorphic variant cells vary in shape, size, and staining (pleomorphism), but many are large cells with oval to irregular nuclei, pale cytoplasm, and prominent nucleoli in a small subset of the cells.

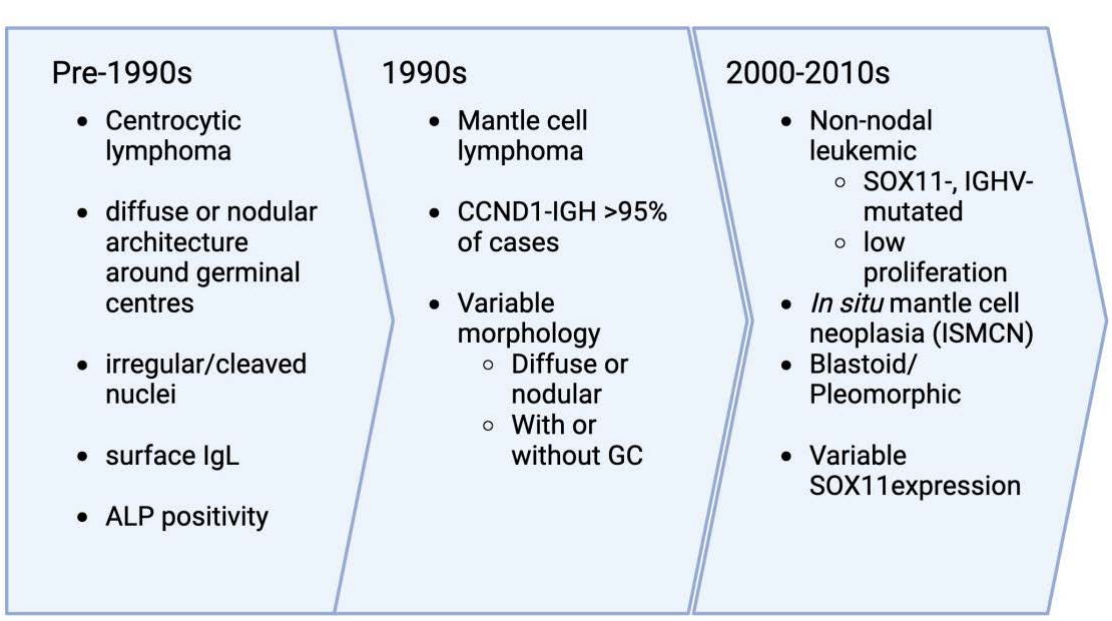

Figure 1-1 History of classification in MCL from 1970 to 2016. Created with Biorender.com

In contrast to the aggressive MCL variants, the recently defined subtype of non-nodal leukemic MCL (nnMCL) is often associated with low proliferation and indolent disease[23,30]. Patients with nnMCL present with peripheral blood, bone marrow and occasionally splenic involvement without significant adenopathy, and remain localized to the mantle zone of lymphoid follicles[31,32] (Table 1-1). *CCND1*::*IGH* translocation is observed in nearly all cases of nnMCL but the genomes of these lack many of the other common genetic mutations or aberrations characteristic of MCL[33]. They also generally lack expression of SOX11 and are more likely to be CD5 negative[23,33]. In contrast to conventional MCL (cMCL), nnMCL is often observed with somatic hypermutation of the immunoglobulin heavy chain variable regions, suggesting germinal centre experience of the originating cell population[34]. While nnMCL is associated with longer overall survival and progression-free survival, these patients may acquire further genetic mutations resulting in transformations to more aggressive variants[35,36].

Table 1-1 Clinical and pathological differences between nnMCL and cMCL

Clinical and pathological variable	nnMCL	cMCL
B-symptoms[33,37]	Rare	Common
Performance status[33] ≥ 2	Uncommon	Common
Nodal presentation[23,33,38]	Uncommon	Common
CD5[32]	Positive	Positive
SOX11[23]	Low-Negative	Positive
IGHV[33,34]	Mutated	Unmutated

nnMCL, non-nodal leukemic MCL; cMCL, conventional MCL; IGHV, immunoglobulin heavy chain variable region

Defining these morphological variants have provided a better understanding in MCL oncogenesis and have produced a plausible model of MCL development (Figure 1-2). Rearrangement of CCND1 and IGH commonly occurs early in B-cell development during the pro-B cell stage within the bone marrow. Neoplastic cells may or may not express SOX11 and this expression is dependent on whether the cells remain in the mantle zone, resulting in the conventional MCL variant, or enter the germinal centre to form nnMCL. In the evolutionary path to cMCL, additional genetic aberrations are acquired whereas the non-nodal leukemic variant has a more restricted landscape of secondary driver mutations.

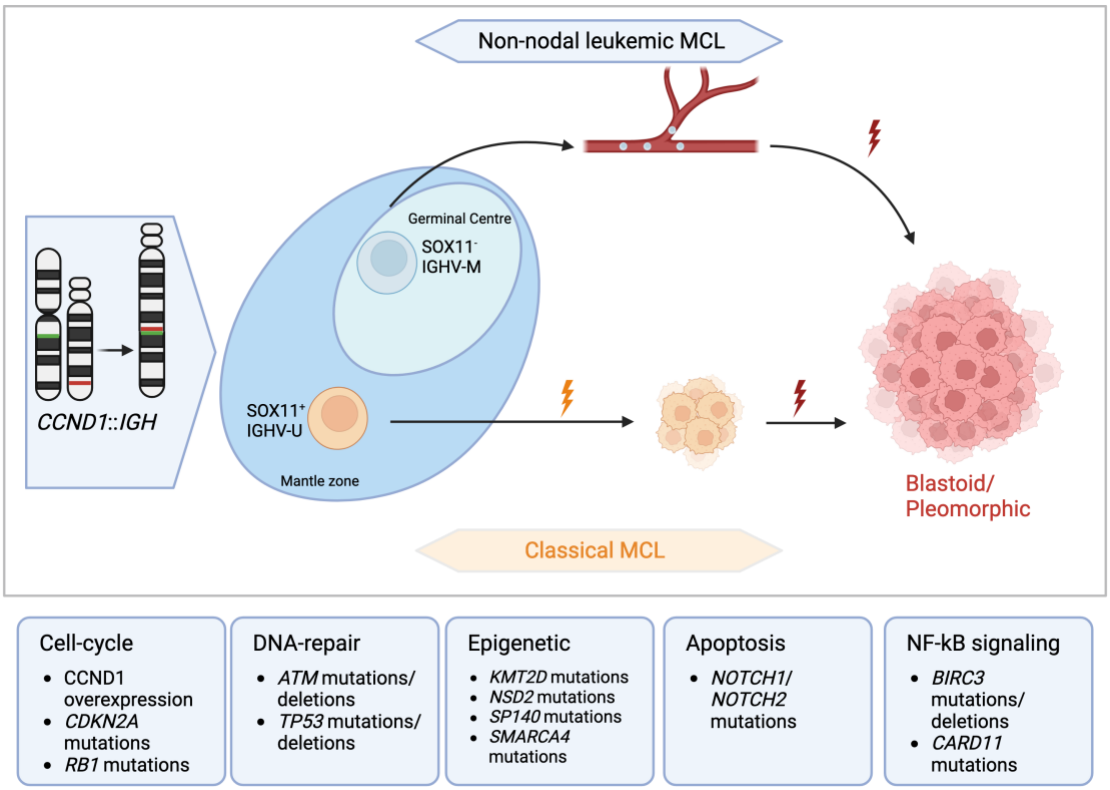

Figure 1-2 Schematic diagram of MCL subtypes and model of tumour development

Abbreviations: IGHV-U, unmutated immunoglobulin heavy chain variable region; IGHV-M, mutated immunoglobulin heavy chain variable region. Created with BioRender.com

1.2. Epidemiology and Aetiology

MCL accounts for 3-10% of NHL diagnoses[25] with an age-standardized incident rate of 0.8 per 100,000 individuals in the United States between 2011 and 2012[39]. At diagnosis, patients range from between 60 and 70 years old with the median age of 65 years[25]. MCL disproportionately affects male individuals with a male-to-female ratio of 3:1[25].

Evidence of non-molecular risk factors in MCL is either inconclusive or lacking compared to other B-cell lymphomas partly due to the relative rarity of the disease[40]. Factors related to occupation, lifestyle, infectious agents, or immune suppression have not been confidently established as risk factors in MCL[40]. Family history may contribute to increased NHL risk including MCL[41]. In a pooled case-control analysis of 10,211 NHLs

and 11,905 controls, Wang et al[41] showed a four-fold increase in MCL risk if any first-degree male relative was diagnosed with any NHL and two-fold risk if either parent was diagnosed with NHL.

1.3. Genetics of Mantle Cell Lymphoma

1.3.1. Primary Aberration: *CCND1*::*IGH*

CCND1::*IGH* translocation is the primary driver event in MCL and is thought to occur from aberrant breakages during VDJ recombination[42]. B-cell development is a multi-step process involving the bone marrow and lymphoid organs and each stage is associated with progressive completion towards a working immunoglobulin protein[43,44]. Immunoglobulins are comprised of two heavy chain (IGH) and two light chain (IGK or IGL) peptides. Heavy chains are further subdivided into variable (V), diversity (D) and joining (J) segments. Multiple copies of segments are present on chromosome 14 and must be joined together into a single gene prior to transcription. This rearrangement process, called V(D)J recombination, is the basis for generating antibody diversity of the immunoglobulin protein.

Beginning in the bone marrow, hematopoietic stems cells differentiate into early pro-B cells which involves rearrangement of D and J heavy chain segments followed by the formation of the late pro-B cell whereby a V segment is joined with the DJ segments to produce the final IGH gene. A similar process occurs involving light chain V and J segments to produce a viable light chain. Completion of the light chain marks the progression from pro-B cell to pre-B cell. Lastly, the rearranged chains are associated with the IgM constant domain gene and expressed on the B cell surface waiting for antigen activation. Immature B cells that fail to undergo auto-antigenic reactions exit the bone marrow and enter circulation where they are now classified as a mature, naïve B cell.

Mistakes during VDJ recombination provide a unique opportunity for aberrant rearrangements due to the formation of double stranded breaks in DNA. The precise moment when *CCND1* and IGH are juxtaposed in MCL cases can be inferred by the examining which V, D or J segments are involved in the rearrangement[45]. Cases were

often found with either D or J gene segments next to *CCND1* suggesting translocation occurring at the pro-B cell stage during D-J joining.

The proto-oncogene *CCND1* was first discovered in part due to its involvement in *CCND1*::*IGH* translocation observed in MCL[46,47]. Characterization of the proto-oncogene revealed its nature as an important cyclin protein involved in regulating transition from G1 to S-phase. CCND1 is not expressed in normal B cells but is found overexpressed in virtually all MCL tumours due to the juxtaposition of *CCND1* gene to the constitutive action of the IGH enhancer elements[48–50].

CCND1 role in regulating cell cycle progression involves binding to kinases CDK4/6 in a dose-dependent fashion[51]. The cyclin D1 kinase complex serves as a positive regulator of the transcription factor E2F by phosphorylating and releasing retinoblastoma (RB1) protein from binding E2F. Upon releasing from the inhibitory effect of RB1, E2F enters the nucleus to control the transcription of a variety of S-phase genes. Additionally, CDK4/6 and CCND1 further promotes transition into the S-phase by increasing the activity of cyclin E/CDK2.

1.3.2. MCL Drivers Beyond *CCND1*::*IGH*

While *CCND1*::*IGH* translocation is the most frequently observed mutation in MCL, the rearrangement alone is insufficient in inducing B-cell lymphomas according to transgenic mice studies whereby CCND1 is overexpressed[52]. This suggested that other genetic alterations were required for complete transformation and progression.

Mutations affecting *TP53* had been discovered in other B-cell malignancies and were often associated with progression into aggressive variants and poorer survival in CLL[53] and FL[54]. Mutations and over-expression of TP53 were soon found associated with blastoid MCL morphology and were predictive of shorter overall survival[55]. Given the importance of CCND1 in MCL, it was hypothesized that other cell-cycle control genes were involved. Recurrent 9p12 deletions suggested the involvement of CDKN2A in MCL[56]. Deletions and mutations of both *TP53* and *CDKN2A* are thought to act in conjunction with CCND1 to inhibit apoptosis and bypass cell-cycle checkpoints[57].

Involvement of other tumour suppressors were suspected as evidenced by recurrent deletions on chromosome 11q and 13q[58].

FISH experiments localized deletions on 11q to a region less than 1Mbp containing the *ATM* gene (ataxia telangiectasia mutated)[59]. Mutations in *ATM* are causative for a familial disease ataxia telangiectasia. The role of ATM as the candidate tumour suppressor deleted in this region stems from the observation that patients with ataxia telangiectasia are more likely to develop lymphoid malignancies and was shown to be inactivated in a subset of T-prolymphocytic leukemia and B-cell chronic lymphocytic leukemia[60–62]. *ATM* encodes a phosphoprotein containing a PI3K kinase domain. Like other PI3K containing proteins, ATM functions in DNA damage repair, cell-cycle checkpoints, and apoptosis. TP53 becomes phosphorylated and activated by ATM in response to DNA strand breakages leading to cell-cycle arrest. It is therefore likely that ATM inactivation works synergistically with CCND1 overexpression to overcome cell-cycle checkpoints resulting in increased genomic instability and evasion of apoptosis[63].

The next generation sequencing revolution have provided greater insights into the full spectrum of mutated genes in MCL and beyond. By providing unbiased nucleotide-level resolution of genetic mutations, exome sequencing studies have implicated several other genes in MCL oncogenesis[64–66]. ATM was determined to be the most frequently mutated gene in MCL and often harboured truncating and deleterious type mutations and, consistent with recurrence of 11q deletions found previously, suggested a tumour suppressor role in MCL[64–66]. Sequencing studies have implicated another putative tumour suppressor in 11q: *BIRC3* which was observed to contain similarly inactivating single nucleotide variations and indels[66]. BIRC3 plays a pivotal role in regulating apoptosis and cellular survival as an inhibitor of apoptosis (IAP) protein family member. Specifically, BIRC3 functions by regulating caspase activation and NF-kB signaling[67]. Inactivating BIRC3 may promote tumour development by stabilizing BCL2, an anti-apoptotic target, and the primary NF-kB signaling kinase, MAP3K14[67]. The latter mechanism is highly relevant with regards to treatment resistance to inhibitors of BTK such as ibrutinib[68]. Mutations and deletions of *BIRC3* while infrequently observed in CLL, are predictive of poor prognosis a trait consistent with MCL cases[69].

Recurrent deletions within 9p are thought to inactivate the *CDKN2A* gene which uniquely encodes for two tumour suppressors: INK4A (p16) and ARF (p14)[70]. These proteins are involved in deregulating already important pathways in MCL involving TP53 and CCND1. INK4A encodes a potent cyclin D CDK inhibitor that, when lost, likely serves to increase CCND1-CDK4/6 activity[70]. Simultaneously, loss of ARF encoded protein p14 will cause destabilization of the p53 leading to evasion of cell-cycle arrest in response to DNA damage and evasion of apoptosis[70]. Alterations involving *CCND1*, *CDKN2A*, *ATM* and *TP53* represent an important axis by which aspects of normal cellular processes are circumvented leading to tumour transformation and progression.

Additional genes have been found altered in MCL, but an understanding of the effect on tumour cells is not entirely clear. One such gene is *S1PR1* encoding for a G-protein coupled receptor which binds sphingosine-1-phosphate[64]. Frequent losses of 1p locus containing *S1PR1* and recurrence of truncating mutations suggest a role of tumour suppressor in MCL. S1PR1 is described to play a role in cellular adhesion and the loss of this gene may promote migration and effacement of normal lymphoid structure.

KMT2D is the third most altered gene according to exome sequencing studies behind CCND1 and ATM. In the same studies, *NSD2* (WHSC1, MMSET) was found recurrently mutated in MCL[66]. Together *KMT2D* and *NSD2* represent the limited number of histone modifiers found mutated in MCL. *KMT2D* encodes a histone methyltransferase which add methyl groups to histone 3 lysine residues (H3K4) and is a marker of transcriptionally active chromatin. In MCL, and other B-cell lymphomas, *KMT2D* is predominately observed having truncating mutations and an infrequent number of missense mutations of unknown consequence[65,66]. *NSD2* mutations on the other hand were mostly missense mutations and affected two residues (p.E1099K and p.T1150A). *NSD2*, like *KMT2D*, encodes a histone methyltransferase but catalyzes methyl addition onto lysine 36 (H3K36). Of the B-cell lymphomas, *NSD2* mutations appear unique to MCL, but translocations involving *NSD2* and IGH have been observed in plasma cell myeloma. Overexpression of NSD2 due to this translocation results in increase of H3K36 and decrease in H3K27 methylation genome wide.

1.4. Current Treatment Strategies

No established standard of care exists for treating MCL and treatment strategies vary between institutions. With the recognition of nnMCL as a distinct entity, an important consideration when treating MCL is the identification of patients with nnMCL as these patients can safely be given a "watch-and-wait" approach or given less intensive therapeutic regimens[71–74]. Forgoing or delaying treatment may also be suitable for asymptomatic patients with low-volume disease and tumours which lack high-risk factors such as high proliferation rate[75,76].

After ruling out indolent disease or nnMCL and confirmation of symptomatic, high-volume disease and/or high-risk factors, treatment considerations include patient age, fitness, and serious co-morbidities. Younger and healthy patients are assessed for their eligibility to receive autologous stem cell transplantation (ASCT) and are given high-intensity induction therapy[77]. Induction therapies for these patients include rituximab plus hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone (R-HyperCVAD) alternating with rituximab plus methotrexate and cytarabine (R-MA)[78], or alternating therapies of R-CHOP or modified R-CHOP (rituximab plus maximum-strength cyclophosphamide, vincristine, doxorubicin, and prednisone) or R-DHAP (rituximab plus dexamethasone, cytarabine, cisplatin) followed by high-dose cytabarine treatment[79–81]. Induction therapy is consolidated with ASCT followed by optional rituximab maintenance.

In older patients who are not eligible for ASCT or high-intensity therapies, a variety of regimens are available for treatment[82]. Induction therapy for these patients include R-CHOP followed by rituximab maintenance[83], or rituximab plus bendamustine (R-BR)[84–86], or modified R-CHOP with vincristine substituted for the proteasomal inhibitor bortezomib[87].

Advances in treatment have shown modest improvements in patient outcome and/or reduced toxicity[88], but most patients will eventually relapse or were refractory to initial treatments and will require salvage therapy. Several novel therapies have been investigated or approved for usage in relapsed-refractory (RR) MCL. These include bortezomib[89], mammalian target of rapamycin (mTOR) inhibitor temsirolimus[90], phosphatidylinositol 3-kinase (PI3K) inhibitors idelalisib[91], Bcl-2 inhibitor venetoclax, and Bruton's tyrosine kinase (BTK) inhibitors such as ibrutinib[92–94], zanubritinib[95,96],

acalabrutinib[97]. Although approved for use In RR MCL, these therapeutics are not curative with complete response rates ranging between 19-71%[98].

1.5. Risk stratification

1.5.1. Prognostic indices

Since its classification as a separate malignancy, various characteristics about MCL have been described and published. An important feature observed in clinical trials was the wide variation in clinical course among MCL patients, but also the general unsatisfactory results of treatment regimens[99–101]. A prognostic index for MCL was required for better stratification of patients during clinical trials. Application of International Prognostic Index (IPI) and Follicular Lymphoma International Prognostic Index (FLIPI) indices on MCL patients proved to contain serious limitations and therefore an MCL specific prognostication index was required[29,102]. The MCL International Prognostic Index (MIPI) was developed to fulfill this need and was found superior to IPI in predicting clinical outcome[103,104]. MIPI utilizes patient age, ECOG performance status, white blood cell count (WBC) and lactate dehydrogenase (LDH) to separate patients into three groups with differential survival: low-, intermediate- and high-risk. High proliferation rate was identified early on as a marker of poor prognosis[29,102,105]. Proliferation as measured by the proportion of Ki67 positively staining cells was found independently prognostic of MIPI score in anti-CD20 immunochemotherapy treated patients and is included during MCL patient stratifications forming a separate metric called biologic MIPI (MIPI-b)[106].

Although many genetic factors have been nominated for having prognostic impacts on MCL survival, *TP53* and *CDKN2A* aberrations have consistently shown to predict inferior survival. *TP53* aberrations such as 17p deletion, *TP53* mutation, a combination of both, or p53 expression have been evaluated for prognostic value[107–109]. In a study involving 365 patients, the European Mantle Cell Network separated patients based on p53 immunohistochemistry expression into low-, intermediate-, and high-expression categories[110]. In both univariate and multivariate analysis, high p53 expression was associated with poor overall survival and inferior time-to-treatment

failure compared to low p53 expression independent of MIPI score and Ki67 index. In addition to *TP53*, *CDKN2A* aberrations are consistently associated with poor prognosis[111–113]. TP53 deletions and/or CDKN2A deletions were additively prognostic for poor overall survival in young MCL patients treated with high-dosage cytarabine treatment prior to myeloablative radio-chemotherapy and ASCT[113].

1.5.2. Gene expression-based classifications

In 2003, Rosenwald et al[16] demonstrated that a proliferation gene expression signature was associated with overall survival in MCL. Using a supervised training approach, 48 genes were found linearly correlated with length of survival and were highly expressed within tumours of patients with the shortest survival. Of the 48 genes, 28 were proliferation related genes involved in DNA replication, cell-cycle progression and metabolic proteins required for proliferating cells. Focusing on the top one-third most variably expressed genes of the 48 gene set, 38 genes – 20 of which were proliferation-related – were used to construct a proliferation score by averaging the expression of these genes. Subsequent analyses demonstrated that tumours with high average proliferation signature expression were associated with higher *CCND1* expression, expression of the shorter *CCND1* isoform which lacked 3'UTR possibly containing an RNA-destabilizing element[114,115], and *CDKN2A* deletions.

Due to the difficulties related to using microarray-based technologies and requirement of fresh frozen (FF) tissues, gene expression-based classification has not been made routine within the clinic. To improve usability and expand the applicability of expression based classification in a clinical setting, Scott et al[116] developed the MCL35 assay for use in both FF and formalin-fixed paraffin-embedded (FFPE) tumours. Beginning with 80 FF tumour samples from Rosenwald et al[16], Affymetrix U133 plus 2.0 microarrays were used to find genes that correlated with both the proliferation signature and OS. Sixty-nine genes including 30 housekeeping genes were selected for subsequent assay development. Expression was quantified for these 99 genes within 47 FFPE samples using the NanoString platform and within 39 matching FF samples using Affymetrix microarrays. The final set of genes included 17 proliferation-associated genes and 18 housekeeping genes which were selected based on high correlation between

NanoString and Affymetrix platforms and high expression in the NanoString data for proliferative genes, and low-variance and moderate-to-high expression for housekeeping genes which are included for normalization. Patients were categorized into low-, intermediate-, and high-risk groups based on expression of the selected proliferation genes after normalizing for housekeeping gene expression. The MCL35 assay was prognostic in patients treated with R-CHOP with median OS reaching 8.6, 2.6, and 1.1 years for low-, intermediate-, and high-risk patient categories, respectively. Like the original proliferation signature, common high-risk features were associated with the high-risk MCL35 group including an enrichment of aggressive MCL variants (blastoid/pleomorphic), TP53 protein positivity, and 3'UTR truncated *CCND1*. Gene signature-based risk stratifications demonstrate a link between the clinical variability in MCL and an underlying genetic explanation. However, the genetic aberrations which contribute to the observed variability, especially with respect to low- and intermediate-risk MCLs is unknown.

1.6. Problem statement

In the decades since the recognition of MCL as an entity, significant progress has been made in understanding the MCL pathogenesis and have improved clinical prospects for many patients. However, the disease remains largely incurable. I will explore the use of genome-wide analysis to aid in enhancing our understanding of the molecular and genetic nature of MCL, its driver mutations and combinations thereof, and the relationship between these and outcome.

This thesis will be guided by two main hypotheses: (1) that driver mutations remain to be discovered in MCL, and (2) MCL can be subdivided into molecular (or genetic) groups beyond the current divisions of conventional MCL and non-nodal leukemic MCL. To address the first hypothesis, this thesis will aim to identify recurrent driver mutations using a varied and comprehensive bioinformatic approach, including the identification of SNVs, CNVs and SVs. These analyses and results will be detailed in chapters 2 and 3. In Chapter 4, I will address the second hypothesis by combining results from the previous chapters and perform un-supervised clustering using non-negative matrix factorization to identify potential MCL subgroups.

Chapter 2.

Coding and non-coding drivers of mantle cell lymphoma identified through exome and genome sequencing

Data presented in this chapter were previously published. Pararajalingam P*, Coyle KM*, Arthur SE, Thomas N, Alcaide M, Meissner B, Boyle M, Qureshi Q, Grande BM, Rushton C, Slack GW, Mungall AJ, Tam CS, Agarwal R, Dawson SJ, Lenz G, Balasubramanian S, Gascoyne RD, Steidl C, Connors J, Villa D, Audas TE, Marra MA, Johnson NA, Scott DW, Morin RD. Coding and noncoding drivers of mantle cell lymphoma identified through exome and genome sequencing. Blood. 2020 Jul 30;136(5):572-584. doi: 10.1182/blood.2019002385. PMID: 32160292; PMCID: PMC7440974. * Contributed equally.

Contributions: PP analyzed and interpreted the data; SEA, MA performed library preparation; KMC and NT analyzed CLIP-seq and RNA-sequencing data; KMC performed cell-based and FFPE-based assays; MA performed sequencing and analysis targeted sequencing; BMG provided bioinformatic support and interpreted the data; BM, MB, GS, AM, and QQ performed nucleic acid extractions and sample quality control; GWS performed TMA scoring; DV and DWS provided clinical data and reviewed the cases; DWS, TEA, MAM, NJ, RG, CS, JC, SB, GL, CST, RA and RDM interpreted data, designed the study, and with PP and KMC, wrote the manuscript.

2.1. Abstract

Mantle cell lymphoma (MCL) is an uncommon B-cell non-Hodgkin lymphoma (NHL) that is incurable with standard therapies. The genetic drivers of this cancer have not been firmly established and the features that contribute to differences in clinical course remain limited. To extend our understanding of the biological pathways involved in this malignancy, we performed a large-scale genomic analysis of MCL using data from 51 exomes and 34 genomes alongside previously published exome cohorts. To confirm our findings, we re-sequenced the genes identified in the exome cohort in 191 MCL tumours, each having clinical follow-up data. This allowed me to confirm the prognostic

association of *TP53* and *NOTCH1* mutations. Our sequencing revealed novel recurrent non-coding mutations surrounding a single exon of the *HNRNPH1* gene. In RNA-seq data from 103 of these cases, MCL tumours with these mutations had a distinct imbalance of *HNRNPH1* isoforms. This altered splicing of *HNRNPH1* was associated with inferior outcomes in MCL and showed a significant increase in protein expression by immunohistochemistry. We describe a functional role for these recurrent non-coding mutations in disrupting an auto-regulatory feedback mechanism, thereby deregulating HNRNPH1 protein expression. Taken together, these data strongly implicate a role for aberrant regulation of mRNA processing in MCL pathobiology.

2.2. Introduction

The present study describes driver mutations in MCL and nominates the perturbation of mRNA processing as an important feature of MCL biology. Specifically, I report novel recurrent mutations affecting genes that encode three RNA-binding proteins *HNRNPH1*, *DAZAP1*, and *EWSR1*, including intronic mutations affecting exon 4 splicing in *HNRNPH1*. We demonstrate that select mutations in *HNRNPH1* alter this splicing and are associated with higher HNRNPH protein expression in patient tissues. Our functional characterization in MCL patient samples and cell lines indicate that *HNRNPH1* splicing is regulated by the HNRNPH1 protein via a negative feedback loop leading to exclusion of exon 4 in the alternative transcript.

2.3. Results

2.3.1. Resolving the frequency of SSMs and recurrently mutated genes in MCL

Several genes have previously been implicated as recurrent targets of simple somatic mutations (SSM) in MCL[64–66,117], though the relevant genes and their mutation incidence has varied considerably between these studies[118]. This can be attributed both to genetic heterogeneity in this malignancy and the limiting cohort sizes included in each study. To address this, we sequenced paired tumour/normal exomes from 51 MCLs

diagnosed in Canada and analyzed these data alongside available paired exome data. Three of the 87 available samples exhibited significantly higher mutation burden (median 5112; range 1621-14959) and were excluded due to hypermutation on the detection of drivers. In the remaining “discovery cohort”, comprising 84 cases, tumour exomes harboured an average of non-silent SSMs affecting 76 genes (range 30-219; Supplemental Figure A.1).

Through our analysis of this cohort, 16 genes were deemed recurrently mutated by two or more algorithms used to identify driver genes. Three of the algorithms found each of *ATM*, *BIRC3*, *TP53*, *S1PR1*, and *B2M* to be significantly mutated, and each of *MEF2B* and *WHSC1* were identified by two methods (Table 2-1). Notably, *CCND1*, often affected by somatic hypermutation, was identified by OncodriveCLUST, which relies on spatial clustering of mutations. Of the candidate MCL genes, those frequently mutated were *ATM*, *CCND1*, *TP53*, *WHSC1*, and *KMT2D*, each gene having been previously nominated by other studies. Three genes not previously attributed to MCL were also identified by at least two methods, namely *HNRNPH1*, *DAZAP1*, and *EWSR1*. Each of these three genes encode RNA-binding proteins that play a role in regulating RNA maturation including alternative splicing[119,120].

Table 2-1 Recurrently mutated genes in MCL exomes.

Gene	# of Tools	Mutated cases (%)	MutSigCV	Oncodrive-FML	OncodriveFM	Oncodrive-CLUST
EWSR1	4	3 (3.6)	1	1	1	1
ATM	3	37 (44)	1	1	1	0
B2M	3	3 (3.6)	1	0	1	1
BIRC3	3	8 (9.5)	1	1	1	0
DAZAP1	3	3 (3.6)	1	1	1	0
KMT2D	3	10 (12)	1	1	1	0
NOTCH2	3	4 (4.8)	0	1	1	1
S1PR1	3	6 (7.1)	1	1	1	0
TP53	3	9 (11)	1	1	1	0
UBR5	3	6 (7.1)	1	1	1	0
HNRNPH1	2	3 (3.6)	1	0	1	0
MEF2B	2	8 (9.5)	1	0	1	0
SP140	2	5 (6)	1	0	1	0
SPEN	2	3 (3.6)	0	1	1	0
WHSC1	2	13 (15)	0	1	1	0

Mutated case percentages calculated from discovery cohort (N=84)

2.3.2. Novel mutation patterns in MCL

Based on these results and prior studies, we performed targeted sequencing of the coding exons of 18 genes in 191 additional MCLs and separately performed WGS on 34 cases to broadly resolve the exonic and intronic mutation patterns (Figure 2-1; Supplemental Data 1). I consolidated variants across all samples sequenced by more than one approach and used the resulting non-redundant variants from 272 cases for subsequent analyses. Mutation patterns and prevalence in established MCL genes were largely consistent with prior reports (Figure 2-2). Each of *NOTCH1*, *MEF2B* and *CCND1* have been shown to have mutation hot spots in MCL and other cancers but the pattern of *MEF2B* mutations in MCL was distinct from that seen in other cancers[121–124] (Supplemental Figure A.2).

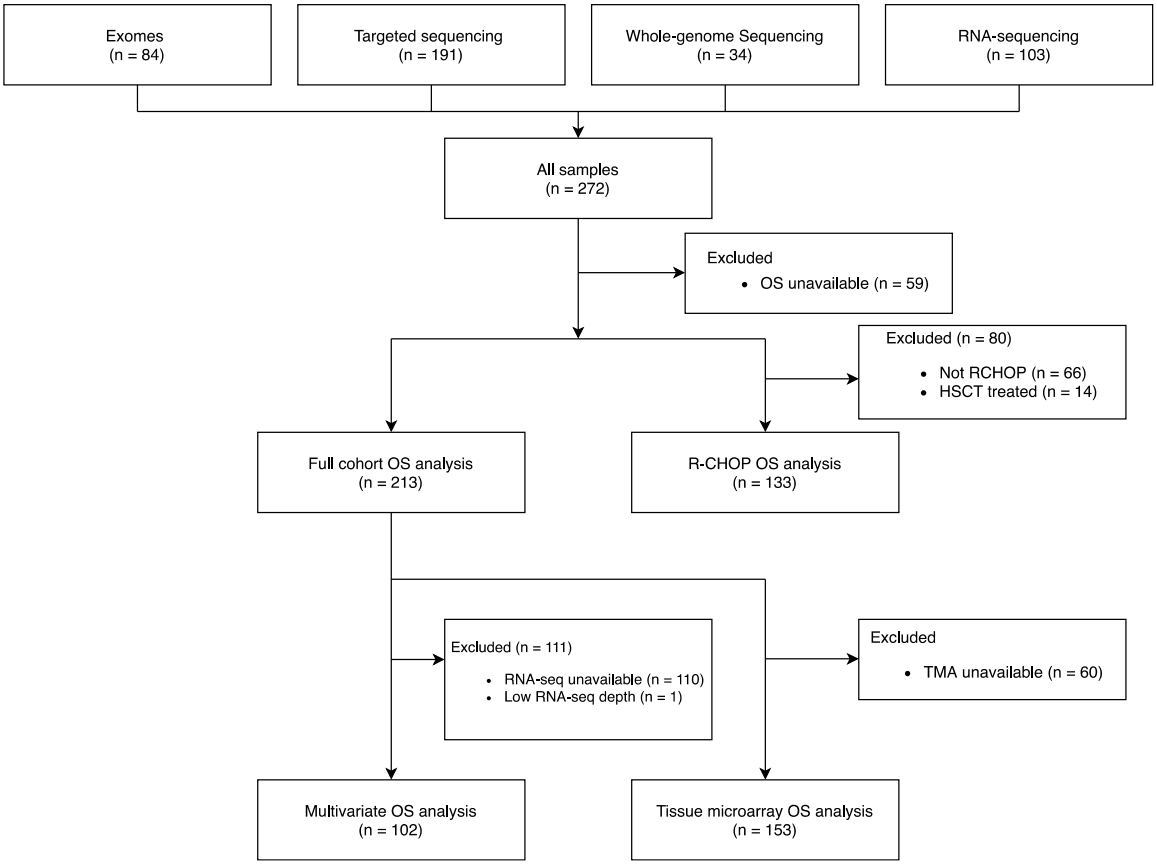

Figure 2-1 Overview of patient samples utilized in individual samples.

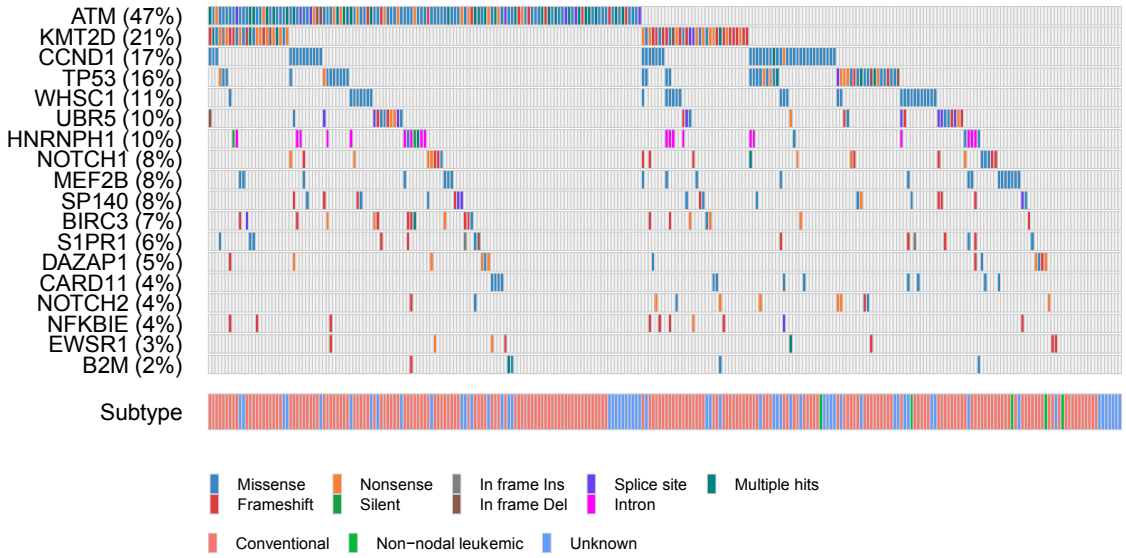

Figure 2-2 Recurrent mutations in combined MCL cohort.
Mutations observed across 272 MCL samples in 18 candidate MCL genes. Mutations shown here are limited to non-silent mutations for all genes except for *HNRNPH1*. For this gene, intronic and silent mutations affecting or immediately surrounding exon 4 are included.

Unsurprisingly, the incidence of non-silent mutations in newly identified genes was generally lower than those of established MCL genes. *EWSR1* was mutated in 8 cases (3%) and *DAZAP1*, in 13 cases (5%) (Figure 2-2). *EWSR1* predominantly harboured frameshift or nonsense mutations in MCL and exhibited a similar pattern at a lower prevalence in a larger compendium of DLBCLs (0.3%) (Figure 2-3A). This suggests that *EWSR1* has an unappreciated tumour suppressor function in MCL and possibly DLBCL. *DAZAP1* had a distinctive pattern with mutations clustered near the C-terminus in a region containing a nuclear localization signal (p.G383-R407)[125] and proline-rich protein-binding domain (Figure 2-3B)[126,127]. Nine cases harboured putative truncating mutations with each predicted to remove or disrupt the nuclear localization signal while leaving most of the open reading frame intact. Non-synonymous mutations in this region mainly affected highly conserved residues (i.e., p.F402, p.R406, p.R407). Previous work indicates that substitution of these residues causes cytoplasmic accumulation of DAZAP1 in human kidney epithelial (293T) and simian (COS7) cell lines[125].

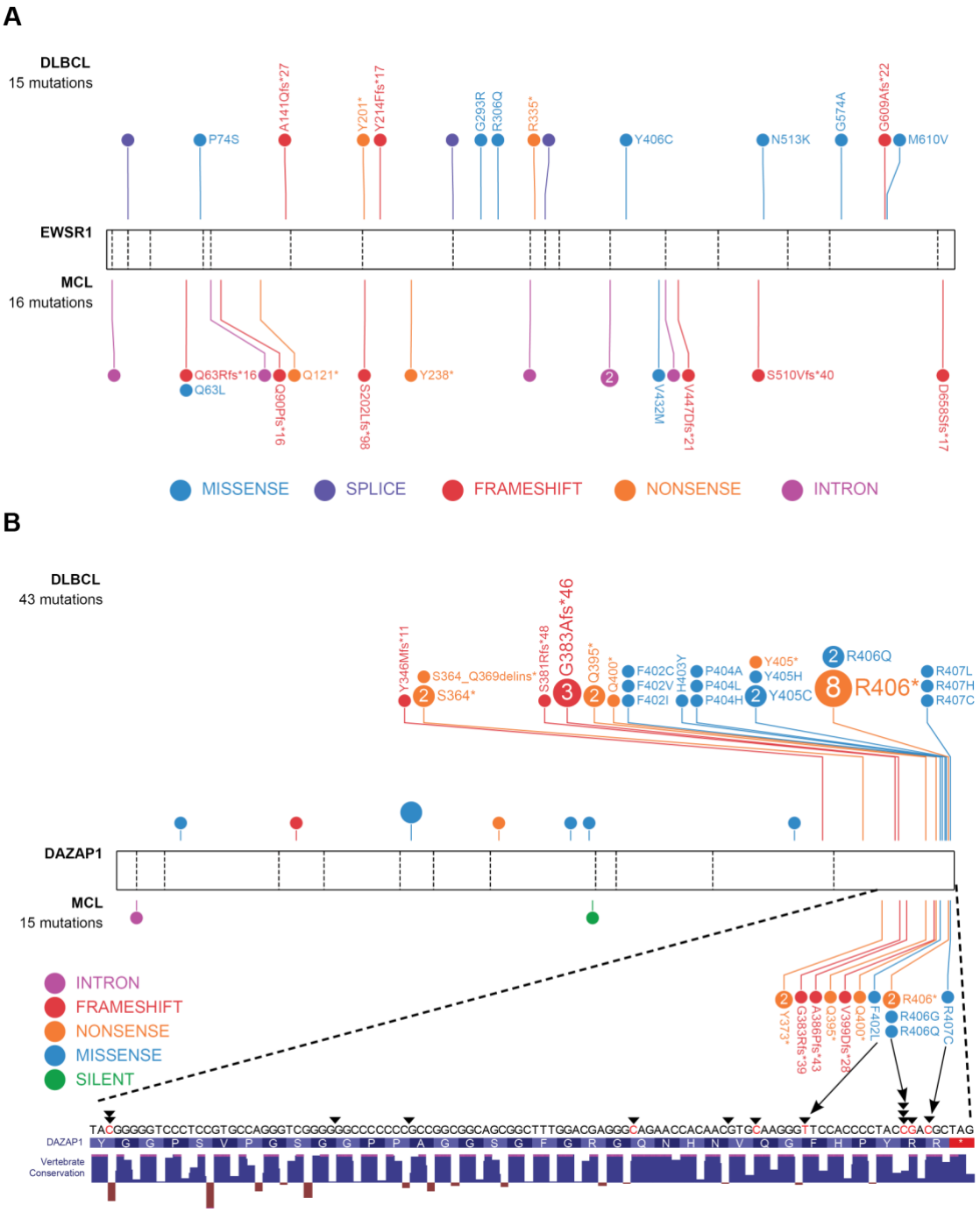

Figure 2-3 Spatial distribution of mutations in DLBCL and MCL of EWSR1, and DAZAP1.

(A) EWSR1 and (B) DAZAP1 amino acid changes and corresponding genetic mutation variant classification. Vertebrate conservation of DAZAP1 amino acids where changes alter protein cellular localization (F402, R406, R407).

2.3.3. *HNRNPH1* intronic mutations disrupt HNRNPH1 binding motifs

HNRNPH1 was mutated in 26 cases (10%) when both coding and non-coding mutations are considered, placing it as the eighth most mutated gene overall (Figure 2-2). Despite limited coverage of introns by our sequencing assay, intronic variants were the most common type of SSM detected in this gene, particularly in the regions surrounding exon 4 (Figure 2-4A). Paired tumour/normal sequencing confirmed that all recurrent variants were somatic and the WGS data confirmed the pattern was restricted to this exon and the immediate flanking regions (Figure 2-4B). HNRNP proteins are widely involved in regulating splicing by binding to pre-mRNA at specific motifs and either promoting or inhibiting usage of nearby splice sites. Distinct from other hnRNPs, HNRNPH1 (and its paralog HNRNPH2) preferentially binds RNA at poly-G motifs[128]. Strikingly, 73% (19/26) of patients with *HNRNPH1* mutations had mutations affecting a poly-G motif within or near this exon. Each of the affected bases are deeply conserved in the homologous region of all available vertebrate genomes, supporting their functional importance.

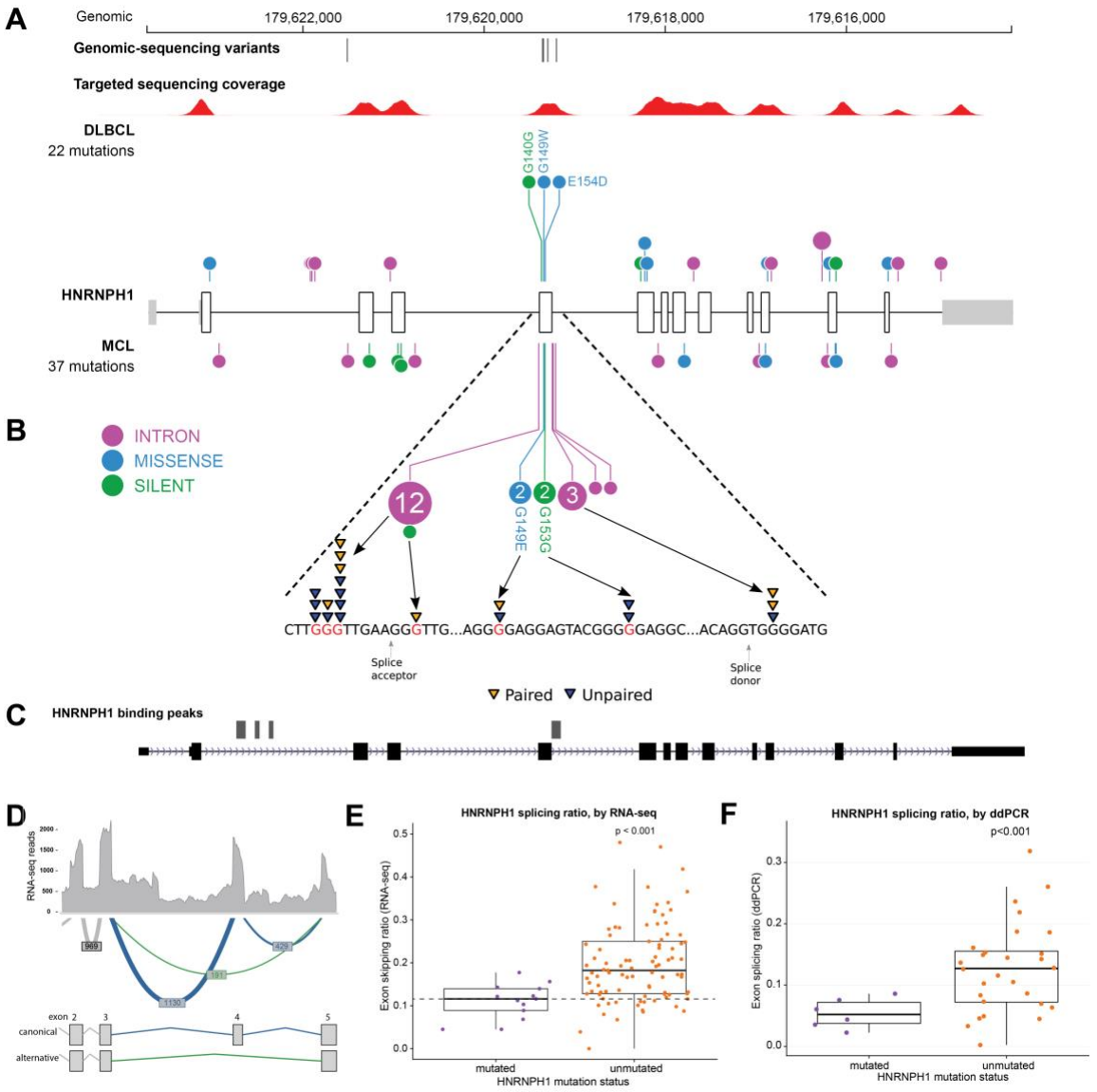

Figure 2-4 *HNRNPH1* mutations in MCL cluster near exon 4 in poly-G motifs.
(A) Somatic mutations found in genomic sequencing cases, and targeted sequencing coverage of a representative sample. The prevalence and pattern of mutations in *HNRNPH1* is compared between DLBCL (top) and MCL (bottom). (B) Splice site and intronic mutations observed surrounding exon 4 affecting poly-G motifs. (C) HNRNPH1 iCLIP binding peaks around exon 4 (Refseq isoform NM_001257293). (D) Sashimi plot of splicing events in *HNRNPH1*. Canonical splicing events shown in blue and exon 4 skipping shown in green. (E) RNA-seq splicing ratios of mutated and unmutated *HNRNPH1* cases. RNA-seq splicing ratio was calculated by the summing reads supporting the alternative event, divided by total reads supporting the canonical splicing event. (F) Digital PCR splicing ratios of mutated and unmutated HNRNPH1 cases.

Further confirming the recurrence of this event, I found additional mutations in this region of *HNRNPH1* by re-analyzing published and unpublished exomes from two

recent studies. Specifically, I noted mutations consistent with the same pattern in three of 16 (18.8%) relapsed/refractory MCL exomes sequenced from a recent clinical trial[129] and four of 24 (16.7%) exomes from another recent study[130] (Supplemental Data 2). We also sequenced this gene in diagnostic tumour tissue from 145 patients treated with either ibrutinib or temsirolimus on a recent clinical trial (NCT01646021)[94] and found mutations in 11 (7.5%) of these cases. Among the available WGS data from Burkitt lymphoma[131] (n = 106), DLBCL (n = 153), chronic lymphocytic leukemia (CLL; n = 144) and follicular lymphoma (n = 110), I only identified two DLBCL patients with *HNRNPH1* mutations in this region (1.3%), suggesting the potential for these to be driver mutations with a highly specific function in MCL biology[132]. This highly reproducible mutation pattern provides strong evidence that these mutations have a regulatory function, most likely affecting the expression and/or splicing of *HNRNPH1* mRNA.

There is a growing list of splicing regulators, including multiple HNRNP family members, that modulate splicing of their own mRNA to tightly regulate expression[133,134]. Taken in consideration along with the prevalence of *HNRNPH1* mutations in sequence contexts resembling HNRNP motifs led us to speculate that HNRNPH1 protein regulates its own expression by modulating the splicing of the *HNRNPH1* transcript. We re-analyzed HNRNPH1 iCLIP-seq data from *Uren et al*[128] and confirmed multiple sites of interaction between HNRNPH1 and its pre-mRNA including exon 4 (Figure 2-4C), supporting a model of direct association at the region affected by mutation.

HNRNPH1 has multiple alternative isoforms including several transcripts that result from skipping of exon 4, which are predicted to be targets of nonsense-mediated decay (NMD) (Figure 2-4D). Although they do not directly affect canonical splice signals, we hypothesized that the mutations in these poly-G motifs impact the splicing or skipping of exon 4. We analyzed RNA-seq data from 103 cases with known *HNRNPH1* mutation status to evaluate splicing differences between mutated (n = 15) and unmutated (n = 88) tumours. By comparing the number of reads supporting the exon skipping event to reads supporting inclusion of exon 4, we found mutated cases exhibited a ratio of isoforms that favours inclusion of exon 4 ($P = 1.13 \times 10^{-5}$, Wilcox rank sum; Figure 2-4E). We implemented a custom ddPCR assay to separately quantify canonical and alternative *HNRNPH1* transcripts. Using this assay, we corroborated these findings in selected cases ($P < 0.001$; Figure 2-4F), which showed a strong correlation ($R = 0.66$, $P < 0.01$) with splicing ratios determined from RNA-seq data from the corresponding cases (Figure

2-5A). These results support the notion that *HNRNPH1* mutations favour the inclusion of exon 4, or suppress the skipping of this exon, promoting the formation of the full-length transcript (Figure 2-5B). Based on our model, these mutations disrupt the binding of HNRNPH1 to poly-G motifs surrounding exon 4 and dampen the normal feedback inhibition (Figure 2-5C).

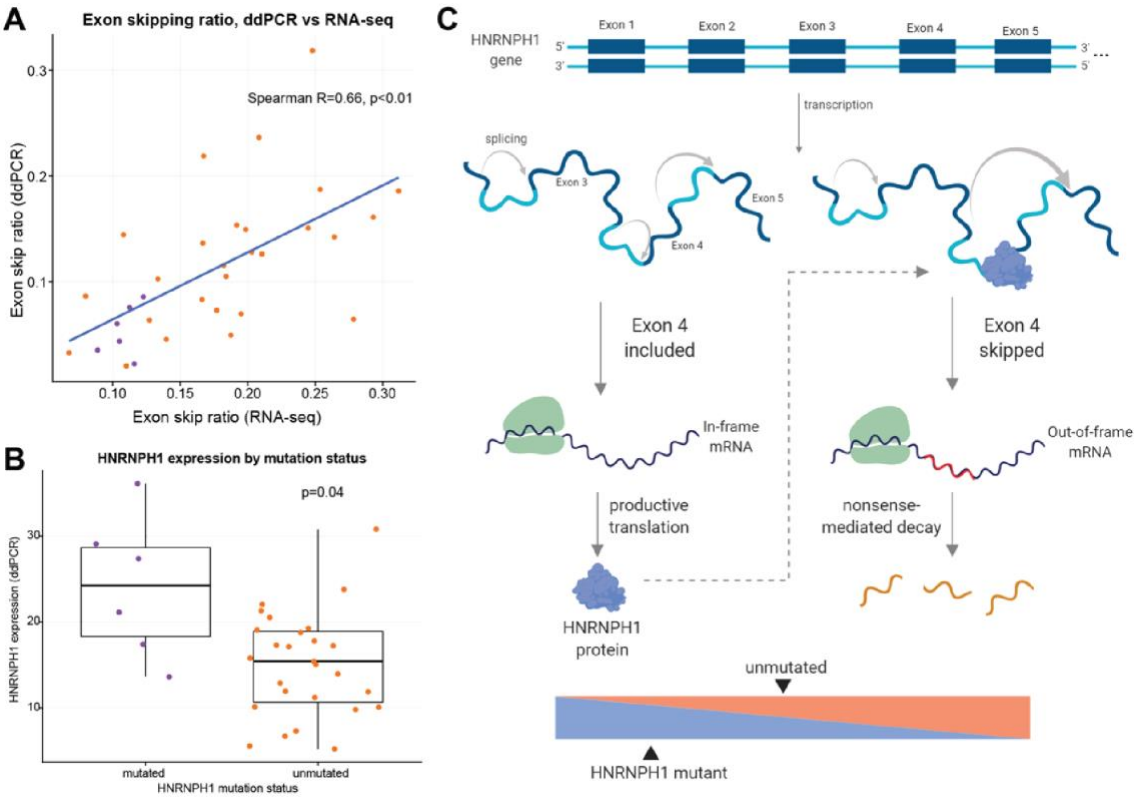

Figure 2-5 *HNRNPH1* mutations influence expression of HNRNPH1.
(A) Exon-skipping ratios between RNA-seq and a custom ddPCR assay. (B) Total HNRNPH1 mRNA expression in mutated and unmutated cases. (C) Proposed model for HNRNPH1 autoregulation suggests that productive splicing and translation of an mRNA including exon 4 generates HNRNPH1 protein, while the presence of excess HNRNPH1 causes exon 4 skipping and leads to nonsense-mediated decay. The observed mutations proximal to exon 4 disrupt this balance and lead to decreased rates of exon 4 skipping.

2.3.4. *HNRNPH1* splicing is associated with inferior outcomes in MCL

I next examined whether any mutations identified in this study were associated with patient outcome. This entailed two separate analyses, first using all cases with available survival data and separately within the subset of cases who received R-CHOP.

In univariate comparisons, mutations in *NOTCH1* (HR = 2.05; Q-value = 8.6 x 10^{-2}) or *TP53* (HR = 3.38; Q-value = 2.8 x 10^{-7}) were associated with shorter OS in the complete cohort. Consistent with previous reports, there was also a significant prognostic association of *NOTCH1* (HR = 2.38; Q-value = 3.6 x 10^{-2}) and *TP53* (HR = 3.53; Q-value = 8.4 x 10^{-6}) mutations in patients treated with R-CHOP. Notably, while *KMT2D* mutations have been recently implicated as a prognostic feature in MCL[135], our analysis did not reproduce this result (Supplemental Figure A.3). Additionally, in R-CHOP treated cases *EWSR1* was associated with shorter OS (HR = 8.71; Q-value = 3.5 x 10^{-3}), although the number of mutated cases was small. In contrast, a significant association was not observed when these patients were stratified by mutation status of other genes, including *HNRNPH1*. I separately evaluated the effect of *HNRNPH1* mutation status in the relapsed/refractory MCL patients treated with either ibrutinib or temsirolimus. Due to the limited number of mutated cases, patients on both arms of the trial were considered together in this analysis. In contrast to the R-CHOP cohort, patients with *HNRNPH1* mutations had significantly shorter PFS (Figure 2-6A) and may contribute to increased aggressiveness of the disease within relapsed/refractory MCL.

Given the strong association between *HNRNPH1* alternative splicing and mutation status, we rationalized that the proportion of *HNRNPH1* mRNAs containing exon 4 could be used as a proxy for HNRNPH1 protein expression. We selected a conservative threshold to assign cases with “mutant-like” exon skipping based on the median value in all cases with *HNRNPH1* mutations (as shown in Figure 2-4E). In patients with RNA-seq data available, this stratification revealed significantly shorter OS in patients with mutant-like splicing of *HNRNPH1* (Figure 2-6B; HR = 2.50; P = 0.00388). Supporting the utility of this information, the splicing ratio was also significantly associated with OS when treated as a continuous variable. In a multivariate analysis, *TP53* mutations, *HNRNPH1* mutant-like splicing, and blastoid morphology were each independently associated with shorter OS (Figure 2-6C). We conclude that the mutant-like splicing pattern of *HNRNPH1*, which favours the productive isoform, is a novel biomarker of inferior outcome in MCL and is independent of established prognostic genetic features and morphology.

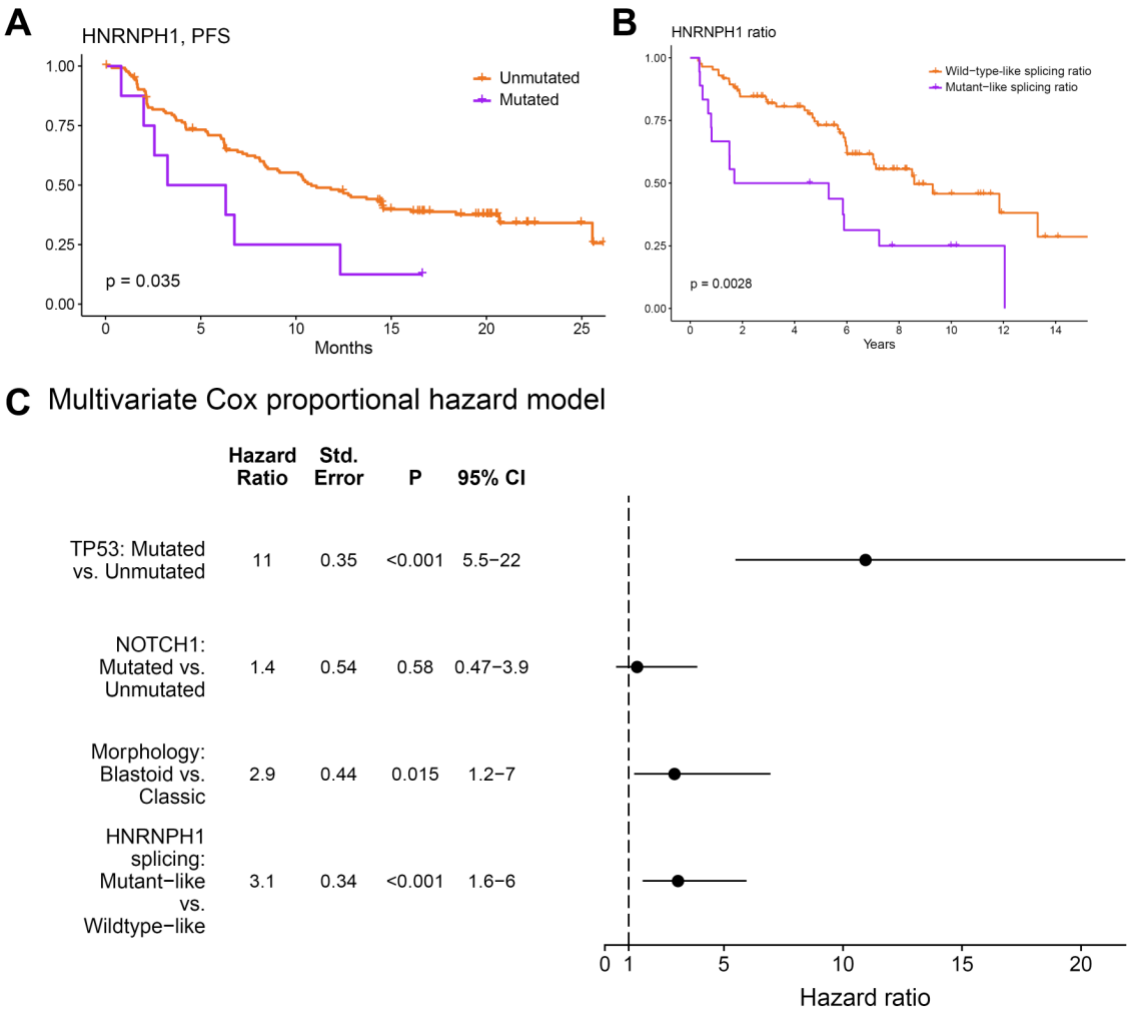

Figure 2-6 ***HNRNPH1*** **splicing is independently associated with survival in MCL.**

(A) Progression-free survival difference between mutated and unmutated HNRNPH1 cases. (B Overall survival difference between wlid-type and mutant-like splicing ratio HNRNPH1 cases. Splicing ratio was calculated for RNA-seq cases. (C) Multivariate survival analysis using Cox proportional hazard on RNA-seq MCL cases.

2.3.5. Mutations and alternative splicing influence HNRNPH1 protein expression in MCL

We hypothesized that *HNRNPH1* mutations in poly-G tracts disrupt an autoregulatory negative feedback loop, which predicts a higher HNRNPH protein expression in *HNRNPH1*-mutant tumours. To address this directly, we evaluated HNRNPH expression in 170 MCL tumours by immunohistochemistry (Figure 2-7A). Of these cases, all had at least one type of sequencing data available and 79 had RNA-seq performed. Consistent with our hypothesis, tissues with strong HNRNPH staining intensity were significantly enriched for *HNRNPH1* mutations (P = 0.0007214, Fisher's exact test). Tissues with strong staining were also enriched for cases with mutant-like splicing (P = 0.001251, Fisher's exact test) and the distribution of splicing ratios was significantly different between tissues with moderate and strong staining (Figure 2-7B). Based on the RNA-seq data, the total mRNA level of *HNRNPH1* was not significantly higher in cases with strong staining (Figure 2-7C). This suggests that the relative proportion of canonical transcripts, rather than the total mRNA abundance, is more directly related to HNRNPH1 protein expression. Our initial finding showing an association between productive splicing and survival is consistent with the trend observed here, namely the association between strong HNRNPH staining and shorter survival.

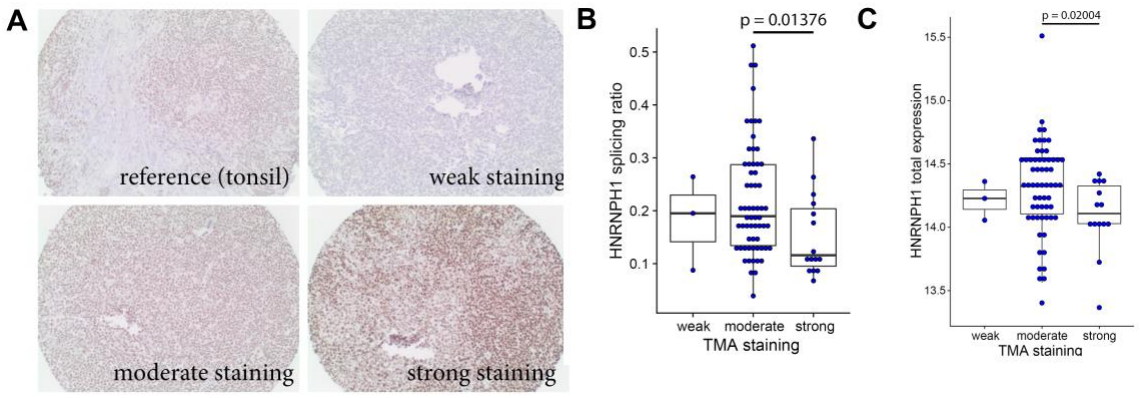

Figure 2-7 Evaluating HNRNPH expression by tissue microarray.
(A) Representative images of tumour cores scored as weak, moderate, or strong HNRNPH staining. (B) *HNRNPH1* splicing ratio stratified by TMA staining categories. (C) Total *HNRNPH1* mRNA expression stratified by TMA straining categories.

2.3.6. Common *HNRNPH1* mutations disrupt productive splicing and translation

The correlation between *HNRNPH1* isoform usage and increased protein levels only indirectly implicates NMD in this process. To substantiate the role of NMD *in vitro*, we inhibited this process using the eukaryotic translation inhibitor cycloheximide. Cycloheximide is a widely used indirect inhibitor of NMD, owing to the essential role of translation in the NMD process[136–139]. In 3 MCL cell lines (JVM2, REC-1, and Z-138) and in HEK cells, we demonstrate that cycloheximide treatment causes a significant and dose-dependent increase in the alternative non-productive *HNRNPH1* transcript compared to total *HNRNPH1* transcript (alternative plus canonical) (Figure 2-8A). This was consistent with the change in splicing pattern observed in SRSF3, which has alternative isoforms that are targeted to NMD due to inclusion of a poison exon (Figure 2-8B)[136]. These results suggest that the *HNRNPH1* isoforms lacking exon 4 are degraded in an NMD-dependent manner in MCL cells.

To functionally demonstrate that mutations found in MCL disrupt regulation of alternative splicing, we constructed a minigene containing the genomic sequence for *HNRNPH1* from exons 2 through 6 (HNRNPH1_ex2_6), including all intronic sequences. Productive splicing of the minigene creates a full-length in-frame peptide containing the hemagglutinin (HA) tag at the C terminus. Unproductive splicing, resulting from skipping

of exon 4, forces the C terminus out-of-frame and causing translated peptides to lack the terminal HA tag. We first transiently transfected the minigene in HEK cells along with a vector bearing the cDNA for HNRNPH1 tagged with EGFP. Ectopic expression of HNRNPH1 notably impaired expression of the HA tag, suggesting a HNRNPH1-dependent switch from productive to unproductive splicing of the minigene (Figure 2-8C). Subsequently, we separately generated three distinct mutants of this minigene by site-directed mutagenesis (Figure 2-8D), and transiently transfected each mutant into HEK cells. The presence of any one of the three mutations tested markedly increased abundance of the HA-tagged peptide, confirming a shift towards productive splicing (Figure 2-8E). This suggests that all three poly-G tracts are individually essential for proper regulation of *HNRNPH1* splicing and adds further support to a model in which these mutations disrupt the autoregulatory feedback cycle and favour HNRNPH1 protein expression.

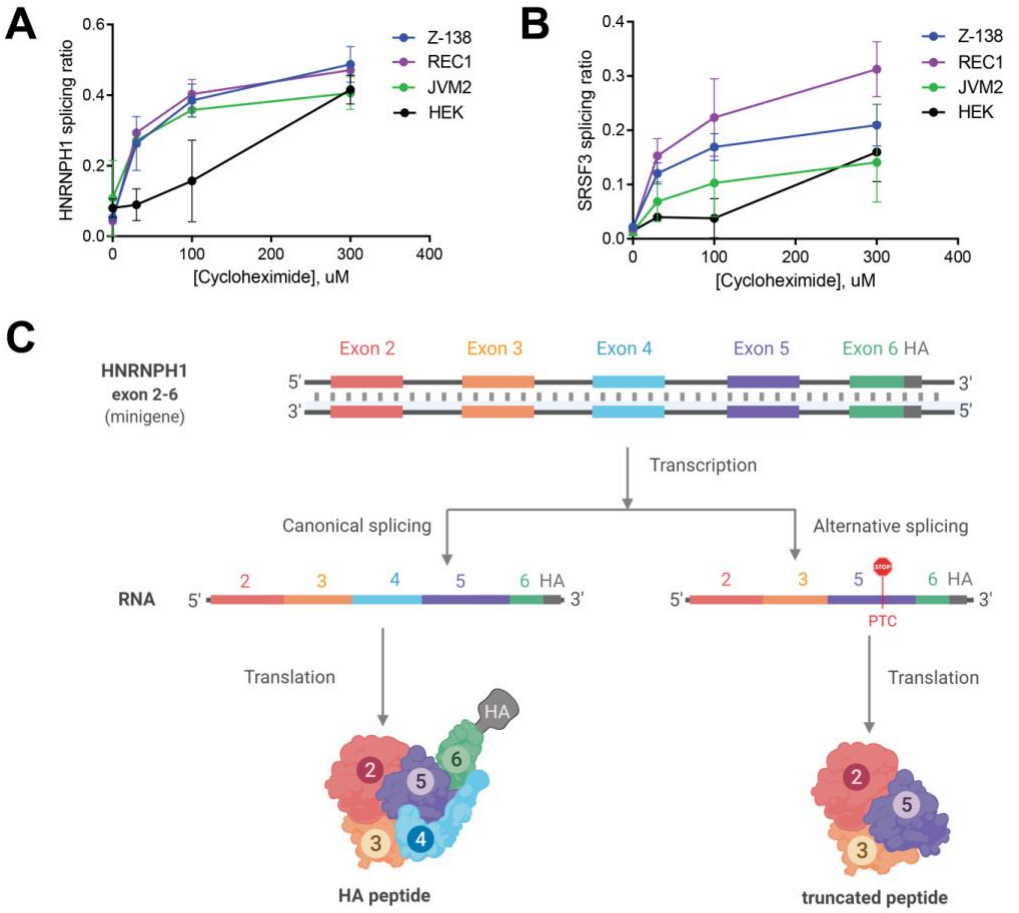
A
HNRNPH1 splicing ratio
0.6
0.4
0.2
0.0
0
100
200
300
400
[Cycloheximide], uM
Z-138
REC1
JVM2
HEK
B
SRSF3 splicing ratio
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
[Cycloheximide], uM
Z-138
REC1
JVM2
HEK
C
HNRNPH1
exon 2-6
(minigene)
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6 HA
5'
3'
3'
5'
Transcription
Canonical splicing
Alternative splicing
RNA
5'
2
3
4
5
6 HA
3'
5'
2
3
5
STOP
6 HA
3'
PTC
Translation
Translation
HA
2
6
5
3
4
HA peptide
2
5
3
truncated peptide
D
EGFP + -
EGFP-HNRNPH1 - +
minigene wt wt
EGFP-HNRNPH1
HNRNPH1
EGFP-HNRNPH1
EGFP
HA
Histone H3
E
CTTGGGTTGAAGGGTTG...AGGGGAGGAGTACGGGGGAGGC...
Splice acceptor
ΔA
ΔB
ΔC
F
minigene wt ΔA ΔB ΔC
HNRNPH
HA
Histone H3

Figure 2-8 Mutations in *HNRNPH1* prevent negative regulation via nonsense-mediated decay.

We separately quantified canonical and alternative (A) *HNRNPH1* and (B) *SRSF3* transcripts by digital PCR in MCL cells (REC1, JVM2, Z138) and HEK cells cultured with cycloheximide (an indirect inhibitor of NMD). This revealed an increasing proportion of the alternative transcript (skipped exon 4 in *HNRNPH1*) with increasing concentrations of cycloheximide. (C) Output from the minigene reporter is represented schematically. Differential splicing of the transcribed pre-mRNA results in inclusion or exclusion of *HNRNPH1* exon 4. The out-of-frame transcript (on right) results in the introduction of a premature termination codon (PTC) in exon 5 and translation of a truncated peptide. Translated peptides are represented in cartoon form. (D) HA tag abundance from wildtype *HNRNPH1* minigene is detected by Western blotting. HA expression, representing productive splicing, is lost in the presence of HNRNPH1 overexpression. (E) The 3 independent G>T mutations introduced into the HNRNPH1 minigene are schematically represented in relation to the patient-identified mutations. (F) Expression of the HA tag from wildtype and mutant HNRNPH1 minigenes are detected by Western blotting. HA expression, again representing productive splicing, is substantially higher when mutated minigenes are transfected into HEK cells as compared to the wildtype minigene.

2.4. Discussion

Using 272 MCL cases, I have described the incidence and pattern of recurrence of mutations in genes with known relevance to MCL, including *ATM*, *KMT2D*, *TP53*, *CCND1*, and *NOTCH1,* focusing on those which could be identified from exome sequencing. Using clinical data available for the bulk of these cases, I confirmed the prognostic association of mutations in both *TP53* and *NOTCH1*. *NOTCH1* was not independently prognostic in a multivariate model that included *TP53* mutations. Notably, in the current data, *WHSC1* mutations were also not associated with OS, in contrast to another study[140], which may attributed to the limited sample size of that study or differences in the patient cohorts. Motivated by the putative role of mutations in *HNRNPH1*, I also identified a strong association between *HNRNPH1* splicing and outcome. Although the likely consequence of an imbalance of *HNRNPH1* isoforms is an increase in HNRNPH1 protein abundance, I found a stronger association between the splicing ratio and patient outcome. This can be attributed to a limited dynamic range available for scoring HNRNPH expression by immunohistochemistry. In the absence of other methods for quantifying this protein in tissues, our results indicate that direct measurement of splicing may be a robust biomarker for HNRNPH activity.

Although the incidence was low, I consider the mutation pattern of *EWSR1* as a notable finding. *EWSR1* is an established cancer gene that is typically discussed in the

context of the *FLI1*::*EWSR1* fusion oncoprotein that drives Ewing sarcoma[141]. The pattern of mutations observed here implies a separate tumour suppressor role of this gene in MCL. Notably, EWSR1 has been implicated in regulating CCND1 by promoting formation of the less oncogenic CCND1a isoform relative to the shorter CCND1b isoform[142]. Although the targets of EWSR1 have not been established in MCL, our data are consistent with the notion that loss of EWSR1 activity alters RNA metabolism and splicing of genes relevant to MCL.

The *DAZAP1* mutations described here are similar to previous reports in a subset of DLBCLs[132,143]. Based on previous mutagenesis experiments[125], I hypothesize that the more common *DAZAP1* mutations cause reduced nuclear occupancy and affect interactions with other proteins. This could disrupt several processes, including transcription, alternative splicing, mRNA transport and translation[119,127,144].

While HNRNPH1 has been identified as overexpressed in other cancer types[145–147], our description of regulatory mutations in *HNRNPH1* is novel and suggests an unappreciated role for HNRNPH1 in B-cell development and/or lymphomagenesis. HNRNPH1 is a member of the HNRNPH/F family of heterogeneous nuclear ribonucleoproteins[148] and binds to various cis-regulatory elements that, depending on the sequence context and interacting proteins, can promote or suppress the use of nearby splice sites[149]. Our data supports a model wherein HNRNPH1 protein normally limits its own accumulation by favouring the skipping of exon 4, thus directing its mRNA to NMD. Self-regulation by modulating unproductive splicing and translation is an emerging theme among other RNA binding proteins, including HNRNPA2B1, HNRNPL, and SRSF3[134,150,151,133,152]. We show using RNA-seq and ddPCR that tissues with mutations near exon 4 have a biased representation of the productive isoform containing this exon. The effect of this on protein expression was confirmed through immunohistochemical analysis of tumour tissue. Similar to the predicted effects of other RNA binding proteins with a multiplicity of targets[153–155], increased expression of HNRNPH1 is expected to have widespread effects on the splicing landscape in MCL[128,149]. The relative paucity of mutations in this region in other B-cell NHL is consistent with a more important role of HNRNPH1 in MCL biology. This warrants further exploration of the suite of genes and splicing events regulated by HNRNPH1 in MCL.

The concurrent identification of three novel MCL-related genes (*EWSR1*, *DAZAP1*, and *HNRNPH1*) with related function is compelling as it may implicate mRNA maturation, splicing and/or trafficking as a general feature of lymphomagenesis in MCL. Accordingly, there is growing evidence relating alterations in RNA-binding proteins and splice factors in numerous cancers, including other B-cell lymphomas, to various aspects of cancer cell biology[156–158]. Specifically, small changes in RNA-binding proteins can have large downstream effects on gene expression and can thus impact multiple hallmarks of cancer[159]. For example, the splicing factor SF3B1 was identified as recurrently mutated in CLL[160–162] and further detailed investigations have identified widespread alternative splicing affecting multiple cellular pathways[163,164]. The identification of pleiotropic effects downstream of SF3B1, including DNA damage response, apoptosis, and Notch signaling, indicate that widespread disruptions to RNA processing can enhance cancer cell survival by multiple pathways[163,165]; further work will identify these downstream effects in MCL.

In summary, through genomic analysis of 273 MCL tumours, I delineated several novel recurrently mutated genes with a range of mutation incidences. This result implicates an important role for RNA-binding proteins and RNA processing in MCL as compared to other B-cell lymphomas, suggesting that RNA metabolism and splicing have a specific role in MCL pathology. Our functional data attributes the common mutations in *HNRNPH1* to disruptions in HNRNPH1 autoregulation, leading to increased HNRNPH1 protein expression in MCL. Further work which links these mutations to dysregulation of specific RNA molecules will highlight the relevance of RNA processing in MCL.

2.5. Methods

2.5.1. Study design and sequencing

I assembled a discovery cohort of paired fresh-frozen (FF) tumour-normal exome sequencing from 51 novel Canadian cases and 33 previously published cases (Bea, Wu). Our validation cohort consisted of targeted sequencing performed on Formalin-Fixed Paraffin-Embedded (FFPE) material representing 191 diagnostic tumour samples from BC (170 unique cases). Sixteen validation samples and 18 additional FF biopsies from BC underwent whole genome sequencing (WGS). We performed RNA-seq on a subset of the BC cases (103 total). This study was approved by the BC Cancer REB and all patients provided informed consent unless the requirement for consent was waived by the REB.

2.5.2. Exome data analysis

For in-house exomes, we used the Agilent SureSelect Human All Exon kits for library preparation and HiSeq2000 instruments (Illumina) for sequencing. We separately obtained exome data from 29 patients described in *Bea et al*[66], which we downloaded from the European Genome-Phenome Archive (EGAS00001000510) in BAM format, extracted to FASTQ and validated for data integrity using BamHash (v1.1)[66,166]. FASTQ files for tumour and matched normal exomes (7 patients) described in *Wu et al* were kindly provided by the authors[64]. For cases with sequence data from more than one tumour biopsy, I included the earliest sample in our analysis. External cases were only included if a matching normal sample was available. Default settings were used for all software unless otherwise stated. I used BWA (v0.7.6a) to map reads to the GRCh38 human reference lacking alternate contigs[167] and subjected these BAM files to soft-clipping of overlapping read pairs using bamUtils clipoverlap (v1.0.13)[168]. For each BAM file, I applied GATK mark duplicates (v3.4.0), and adjusted alignments for putative indels using GATK indel realigner (v3.4.0)[169]. I utilized Strelka (v1.0.14)[170] to detect simple somatic mutations (SSMs) including single nucleotide variants (SNVs) and indels and annotated these variants using Variant Effect Predictor (Ensembl release 83)[171] and vcf2maf.

2.5.3. Recurrence analysis

As genes important for lymphomagenesis can exhibit a variety of mutational patterns, I employed a voting strategy involving four separate algorithms to identify recurrently mutated genes and hot spots. Several algorithms to infer genes that are recurrently mutated in cancer cohorts have been described and each of these relies on a variety of features such as predicted functional impact[172,173], spatial clustering[174], and mutation rates[175]. I identified significantly mutated genes using a combination of MutSigCV[175], OncodriveFM[172], OncodriveFML[173] and OncodriveCLUST[174] using a false discovery rate threshold of 0.1 for each algorithm. I nominated genes identified by two or more methods for further sequencing in an extended cohort. Additionally, *NOTCH1*, *CARD11*, *NFKBIE* were included due to their importance in MCL and other B-cell NHL [64,122,176].

2.5.4. Targeted sequencing, whole genome sequencing and data consolidation

DNA extracted from 191 formalin-fixed paraffin-embedded (FFPE) diagnostic MCL tumour biopsies was used to generate libraries that were enriched for exons corresponding to a set of putative MCL-related genes using a hybridization-based capture approach involving complementary DNA oligonucleotides[132]. These were pooled and sequenced using a MiSeq (Illumina). Paired 150-nt reads were generated, demultiplexed, and mapped to the GRCh38 human reference using Geneious (v9.1.5). Variants with minor allele frequencies greater than 0.0001 in any gnomAD population were considered germline variants and removed[177]. I consolidated variants from cases sequenced by more than one method using a tiered approach. Variants found in exomes and genomes were combined per patient. In targeted sequenced tumours for which a normal exome or genome data was available, I considered variants with more than one read support in the normal to be germline variants and removed from analysis. Variants from targeted sequencing cases that did not overlap with exome or genome cases were included in the final variant set.

2.5.5. Whole genome sequencing library construction and sequencing

Whole genome sequencing (WGS) libraries were constructed from MCL fresh frozen tumour and constitutive DNA collected from patients in British Columbia, Canada. To minimize library bias and coverage gaps associated with PCR amplification of high GC or AT-rich regions we have implemented a version of the TruSeq DNA PCR-free kit (E6875-6877B-GSC, New England Biolabs), automated on a Microlab NIMBUS liquid handling robot (Hamilton). Briefly, 500 ng of genomic DNA was arrayed in a 96-well microtitre plate and subjected to shearing by sonication (Covaris LE220). Sheared DNA was end-repaired and size selected using paramagnetic PCRClean DX beads (C-1003-450, Aline Biosciences) targeting a 300-400 bp fraction. After 3' A-tailing, full length TruSeq adapters were ligated. Libraries were purified using paramagnetic (Aline Biosciences) beads. PCR-free genome library concentrations were quantified using a qPCR Library Quantification kit (KAPA, KK4824) prior to sequencing with paired-end 150 nucleotide reads on the Illumina HiSeqX platform using V4 chemistry according to manufacturer recommendations.

Variants from the genome data were identified using Strelka2[178], which we found to be more robust for identifying variants with low read support or low-level contamination of blood DNA with tumour cells (a feature of some MCLs). We removed variants with minor allele frequencies greater than 0.0001 in any gnomAD population[177].

2.5.6. Ribosomal RNA depletion RNA sequencing library construction and sequencing

To remove cytoplasmic and mitochondrial ribosomal RNA (rRNA) species from total RNA NEBNext rRNA Depletion Kit for Human/Mouse/Rat was used (NEB, E6310X). Enzymatic reactions were set-up in a 96-well plate (Thermo Fisher Scientific) on a Microlab NIMBUS liquid handler (Hamilton Robotics, USA). 100ng of DNase I treated total RNA in 6 µL was hybridized to rRNA probes in a 7.5 µL reaction. Heat-sealed plates were incubated at 95˚C for 2 minutes followed by incremental reduction in temperature by 0.1˚C per second to 22˚C (730 cycles). The rRNA in DNA hybrids were digested using RNase H in a 10 µL reaction incubated in a thermocycler at 37˚C for 30

minutes. To remove excess rRNA probes (DNA) and residual genomic DNA contamination, DNase I was added in a total reaction volume of 25 µL and incubated at 37˚C for 30 minutes. RNA was purified using RNA MagClean DX beads (Aline Biosciences, USA) with 15 minutes of binding time, 7 minutes clearing on a magnet followed by two 70% ethanol washes, 5 minutes to air dry the RNA pellet and elution in 36 µL DEPC water. The plate containing RNA was stored at -80˚C prior to cDNA synthesis.

First-strand cDNA was synthesized from the purified RNA (minus rRNA) using the Maxima H Minus First Strand cDNA Synthesis kit (Thermo-Fisher, USA) and random hexamer primers at a concentration of 8 ng/µL along with a final concentration of 0.4 µg/µL Actinomycin D, followed by PCR Clean DX bead purification on a Microlab NIMBUS robot (Hamilton Robotics, USA). The second strand cDNA was synthesized following the NEBNext Ultra Directional Second Strand cDNA Synthesis protocol (NEB) that incorporates dUTP in the dNTP mix, allowing the second strand to be digested using USERTM enzyme (NEB) in the post-adapter ligation reaction and thus achieving strand specificity.

cDNA was fragmented by Covaris LE220 sonication for 130 s (2x65 s) at a “Duty cycle” of 30%, 450 Peak Incident Power (W) and 200 Cycles per Burst in a 96-well microTUBE Plate (P/N: 520078) to achieve 200-250 bp average fragment lengths. The paired-end sequencing library was prepared following the BC Cancer Agency Genome Sciences Centre strand-specific, plate-based library construction protocol on a Microlab NIMBUS robot (Hamilton Robotics, USA). Briefly, the sheared cDNA was subject to end-repair and phosphorylation in a single reaction using an enzyme premix (NEB) containing T4 DNA polymerase, Klenow DNA Polymerase and T4 polynucleotide kinase, incubated at 20˚C for 30 minutes. Repaired cDNA was purified in 96-well format using PCR Clean DX beads (Aline Biosciences, USA), and 3’ A-tailed (adenylation) using Klenow fragment (3’ to 5’ exo minus) and incubation at 37˚C for 30 minutes prior to enzyme heat inactivation. Illumina PE adapters were ligated at 20˚C for 15 minutes. The adapter-ligated products were purified using PCR Clean DX beads, then digested with USERTM enzyme (1 U/µL, NEB) at 37˚C for 15 minutes followed immediately by 13 cycles of indexed PCR using Phusion DNA Polymerase (Thermo Fisher Scientific Inc. USA) and Illumina’s PE primer set. PCR parameters: 98˚C for 1 minute followed by 13 cycles of 98˚C 15 seconds, 65˚C 30 seconds and 72˚C 30 seconds, and then 72˚C 5

minutes. The PCR products were purified and size-selected using a 1:1 PCR Clean DX beads-to-sample ratio (twice), and the eluted DNA quality was assessed with Caliper LabChip GX for DNA samples using the High Sensitivity Assay (PerkinElmer, Inc. USA) and quantified using a Quant-iT dsDNA High Sensitivity Assay Kit on a Qubit fluorometer (Invitrogen) prior to library pooling and size-corrected final molar concentration calculation for Illumina HiSeq2500 sequencing with paired-end 75 base reads.

2.5.7. *HNRNPH1* variant calling in Rule et al.[179], Agarwal et al.[129] and Khodadoust et al.[130]

DNA was extracted from 145 tumour biopsies (including 88 matched germline samples) from MCL patients enrolled in RAY trial (NCT01646021)[94]. Additionally, DNA was extracted from B cell line, CEPH-1328. Libraries were prepared using a hybridization-based capture approach of a panel of 176 known recurrently mutated lymphoma genes. Pooled libraries were sequenced on HiSeq2500 followed by de-multiplexing and read mapping to human genome reference hg19 using bwa aln with the following parameters: -q 5, -l 32, -k 2, -o 1. Duplicate reads were marked using GATK MarkDuplicates and local indel realignment was performed using GATK IndelRealigner. Strelka2 was used for SNV and indel variant calling with matching germline DNA used as normal when applicable, and CEPH-1328 used in cases with no matching germline sample. Two filters were applied to remove potential germline variants. First, a panel of germline variants was found by running Strelka2 on the patient germline samples using CEPH-1328 cell line as a pseudo-normal. Variants with low allele frequencies (AF < 0.15) were removed from the panel-of-normal variants with the reasoning that these could be infiltrating somatic variants due to tumour-in-normal contamination, while the remaining variants were considered true germline variants. Tumour somatic variants were removed if they overlapped this panel-of-normal variant list. Second, variants with minor allele frequencies greater than 0.0001 in any gnomAD population were removed as common variants. Variant calls were then converted to GRCh38 coordinates using CrossMap.py using default parameters. Diagnostic tumour and matched germline exome FASTQs (N = 16) from Agarwal et al[129] were provided to us by the authors and underwent the same read alignment, post-processing, and variant calling as the BC exomes. The authors provided HNRNPH1 variant calls from 24 MCL exomes described

in Khodadoust et al[130] to us. Variant coordinates were converted to GRCh38 using CrossMap.py using default parameters.

2.5.8. RNA-seq analysis of MCL

Reads were aligned with STAR (version 2.5.3a) followed by Picard MarkDuplicates (version 2.14.1)[169,180]. We used FeatureCounts (version 1.6.0) to separately quantify the reads mapping to the relevant *HNRNPH1* splice junctions using the following parameters: juncCounts, countSplit, ignoreDup, requireBothEndsMapped, and minimum mapping quality >= 10. We defined the *HNRNPH1* exon skipping ratio as the ratio of reads spanning the exon 4-6 junction to the sum of reads spanning the exon 4-5 junction and 5-6 junction. RNA-seq samples were designated *HNRNPH1* exon 4 mutated or unmutated based on variants discovered in the matching DNA-sequencing sample. In cases where DNA-sequencing data was not available, we determined *HNRNPH1* exon 4 mutation status by examining the exon 4 region in the RNA-sequencing data in IGV.

2.5.9. Protein-RNA interactions of HNRNPH1

RNA-seq and iCLIP data from HeLa cells was provided by authors[128]. Using STAR, we aligned RNA-seq reads from untreated cells transfected with control siRNA (seq8 in Uren data) as described above. Here, we assigned counts using featureCounts with a flattened GTF file containing collapsed transcripts. For HNRNPH iCLIP reads, we pooled reads from two replicates and aligned each pool to GRCh38 with STAR. Peaks in iCLIP data were called by Piranha[181] in 50 bp bins, accounting for RNA abundance with log converted RNA-seq counts used as a covariate.

2.5.10. Cell-based experiments

We reconstituted cycloheximide (Sigma) in DMSO and added either cycloheximide or DMSO alone to cells for up to 6 hours. We then extracted RNA or protein, as appropriate, from pellets of 2 x 10^6 cells using the RNeasy mini kit (Qiagen) or RIPA buffer, respectively, and quantified protein using the Pierce BCA kit (ThermoFisher). Equal amounts of protein were used for SDS-PAGE, and we used Western blotting to observe HNRNPH (abcam ab10374, 1/10 000), histone H3 (Cell Signaling Technologies #9715, 1/10 000), and HA tag expression (abcam ab137838, 1/500).

2.5.11. Digital PCR

RNA from FFPE samples was reverse transcribed with random hexamers (iScript, BioRad) according to manufacturer's instructions. We pre-amplified cDNA using SsoAdvanced PreAmp SuperMix (BioRad) for 15 cycles with primers targeting *HNRNPH1* (all), *HNRNPH1* (canonical only), *HNRNPH1* (alternative only), *TBP*, *YWHAZ*, and *UBC* (Supplemental Data 3). Following a 1 in 5 dilution of pre-amplified cDNA, we performed ddPCR on the QX200 system (BioRad) using the primers and probes described. In each sample, we calculated the normalized expression relative to the geometric mean of the expression of three reference genes (*TBP*, *YWHAZ*, and *UBC*). Cell line RNA was similarly reverse transcribed prior to ddPCR. We normalized expression relative to the geometric mean of three reference genes (*ACTB*, *YWHAZ*, and *UBC*).

2.5.12. Plasmids

The HNRNPH1 insert (exons 2-6) was amplified from human genomic DNA (Promega, G3041) using Q5® High-Fidelity DNA Polymerase (NEB, M0491) and primers indicated in the Supplemental Data 4. Subsequent nested PCR was used to introduce a 3' HA tag and BamHI/XhoI restriction sites surrounding the selected exons. The insert was ligated into the pcDNA3.1+ vector using the BamHI/XhoI restriction sites and

functionally confirmed in HEK cells. Mutations were generated by site-directed mutagenesis using the indicated primers (Supplemental Data 5). All mutant clones (HNRNPH1_ex2-6 $^{\Delta A, \Delta B, \Delta C}$) were confirmed by Sanger sequencing (GeneWiz, New Jersey USA).

2.5.13. HEK transfections

Approximately $2x10^6$ cells were seeded in 60-mm dishes 24 h before transfection. 1 µg of each plasmid (combinations of HNRNPH1_ex2-6, pEGFP-C1, pEGFP-C1-HNRNPH1) was transfected into HEK cells using Lipofectamine 3000 (Invitrogen). Cells were collected after 24h and RNA or protein purified as described.

2.5.14. Tissue microarray and immunohistochemistry

Tissue microarrays (TMA) were constructed using duplicate 0.6mm cores from formalin-fixed paraffin-embedded MCL tissue. HNRNPH (polyclonal, Abcam, dilution 1:1000) immunohistochemistry was performed on 4µm TMA sections using routine protocols on a fully automated platform (Ventana Discovery XT). Protein expression intensity was assessed by one pathologist (GWS). Protein expression was uniformly expressed by all cells within each tumour but with variable intensity noted between cases, so tumours were scored in comparison to the on-slide control tonsil, which showed uniform moderate intensity expression in all lymphocytes. Tumours were assigned one of the following scores: 0 = absent/weak staining (less intense than tonsil); 1 = moderate staining (equally intense as tonsil); 2 = strong staining (more intense than tonsil).

2.5.15. Statistical analysis

Associations between gene mutation status and binary clinical characteristics were assessed using Fisher's exact test. Overall survival (OS) correlates were

separately tested in all patients regardless of treatment (N = 213) and in the subset of HSCT-untreated cases (N = 133) using univariate Cox proportional hazard modeling and P-values were corrected for multiple-hypothesis testing using Bonferroni-Hochberg. Corrected p-values less than 0.1 were considered significant. *HNRNPH1* exon 4 skipping ratio was derived from RNA-seq data and survival analysis using this ratio was limited to these cases (N = 102), using the median skipping ratio for all *HNRNPH1* mutated cases as a cutoff for “mutant-like” splicing. Multivariate survival associations were examined using Cox proportional hazard model on RNA-seq cases so that *HNRNPH1* exon 4 skipping ratio could be included in the model (N = 102). The final tested multivariate model included *TP53* and *NOTCH1* mutations, *HNRNPH1* mutant-like splicing and morphology.

Chapter 3.

Landscape of copy number and structural variations in MCL

The following chapter details my investigation into the copy number and structural variation landscape in MCL. Results from this chapter were derived from a collaborative effort involving many members of the Morin lab that sought to analyze the genomes from multiple B-cell lymphomas subtypes. My contribution to this effort included the following: development and application of recurrent structural variation modeling across the B-cell lymphomas and identification of MCL-specific SVs, identification of the recurrently amplified and deleted segments in MCL, and integrative analysis of copy number variations and structural variations with gene expression.

3.1. Introduction

Chromosomal aberrations, which can be copy-neutral or affect copy number, are common in many of the B-cell lymphomas. Some entities are defined, in part, based on the presence of inter-chromosomal rearrangements, such as *MYC*::*IGH* in BL, *IGH*::*BCL2* in FL. In the case of MCL, the *CCND1*::*IGH* translocation is observed in virtually all cases but is alone insufficient for oncogenic transformation. This is supported by the observation of t(11;14)(q13;q32) translocations in circulating blood of healthy individuals[182]. The acquisition of additional genetic aberrations within pre-malignant cells bearing this translocation is thought to be required for onset of MCL. These aberrations can roughly be divided into two types: copy-number variations (CNV) and structural variations (SV). CNVs change the total number of alleles of affected genes within the genome. In a diploid organism, deletions are a type of CNV that results in either the loss of one allele (heterozygous deletion) or loss of both alleles (homozygous deletion). Amplifications and gains increase gene copies within a genome and are distinguished by the degree of the increase. Gains refer to low-level increases in copy number while amplifications refer to large increases in copy number.

The most common SVs lead to increased expression by placing an oncogene in proximity to a potent enhancer without changing the dosage of either sequence. CNVs

are, by definition, changes in copy number but also generally involve underlying structural variations unless they involve entire chromosomal gains or losses, which are rare. By changing dosage of genes within the cell, CNVs may affect expression of genes in cis with amplifications and gains increasing and deletions causing reduced expression. CNVs and SNVs can also act synergistically through the loss-of-heterozygosity (LOH) such that the mutant allele becomes the only viable copy within the cell.

In Chapter 2, we demonstrated that *cis* regulatory mutations are an important under-appreciated component of the repertoire of drivers in MCL. This was facilitated by leveraging matched gene expression and genetic information. Several studies have examined the copy number landscape of MCL, relying on lower resolution methods including array-based technology or whole-exome sequencing. Since they alter gene dosage, any putative target of recurrent CNVs should manifest in altered gene expression of the genes within them. In general, previous studies lacked matching gene expression data, which could aid in the effect of CNV on gene expression[183,184]. In lieu of this information, the genes thought to be targeted by a set of CNVs have been largely based on proximity to or existence in the minimal common region.

Only WGS offers the potential to provide an unbiased view of the genomic landscape, allowing the identification CNVs and their underlying breakpoints along with copy-neutral SVs at base-pair resolution. Until recently, there has been limited whole genome data for MCL partly due to the relative rarity of the lymphoma and partly due to cost. SV breakpoints can form at any location in the genome and only through WGS could these positions be hoped to be found. SVs accruing in a specific region may hint towards the presence of a nearby oncogene or tumour suppressor and evidence of positive selection for disrupting the region. Identifying these recurrently altered SV regions has only recently been explored in solid tumours but have not been accomplished in MCL. Besides translocation of *CCND1* to IgH locus, no other recurrent structural variation has been described in MCL. If there exists uncharacterized recurrent SVs, these may represent MCL-associated drivers, some of which could facilitate the identification of new therapeutic strategies.

In this chapter, I explore the MCL mutational landscape from the structural variation and copy number perspective. By leveraging whole-genomes from different B-

cell lymphomas, I developed a background model of the SV rate across the entire genome. Applying this model to MCL, I identified a novel SV breakpoint cluster on 10p12.1. The region also harboured recurrent amplifications and gains. Expression analysis indicates an increase in expression of only a single gene in this region (*ABI1*). Finally, I clarify the putative targets of other commonly mutated regions.

3.2. Results

3.2.1. Clinical and molecular characteristics of MCL tumours

MCL cases in this chapter were part of a larger compendium representing some of the most common mature B-cell neoplasms. These are being analyzed within an internal project known as Genomic Analysis of Mature B-cell Lymphomas (GAMBL). The MCL cohort consisted of 160 tumours including 145 cMCL and 15 nnMCL. Cases were acquired from two sources: 103 cMCL genomes were sequenced locally from patients from British Columbia, Canada (the BC cohort) along with matching constitutive DNA for 83 patient-samples (Supplemental Data 6). An additional 42 cMCLs and 15 nnMCLs and matching constitutive DNA were acquired from the recent study by Nadeau *et al*[45] (the Barcelona cohort). A subset of the BC cohort (81 tumours) also underwent ribosomal RNA depletion RNA-seq. The original study containing the Barcelona cases did not use RNA-seq data and therefore none were available to be included. To facilitate comparisons between tumour and normal RNA data, I included six samples of naïve B cells, which are considered the hypothetical cell-of-origin of MCL[185].

Clinical and molecular characteristics are summarized in Table 3-1. Cases roughly follow the expected 3:1 male-to-female ratio ($p > 0.05$; one-sample proportions test). The median age within the BC cohort was 61 years old (range, 31 - 90). Similar proportions of cases were categorized as either low, intermediate, or high-risk according to MIPI index and MCL35 stratification. Both metrics indicate that majority of cases were considered low-risk - 63% and 56% using MIPI and MCL35, respectively. Most samples were derived from lymph node biopsies.

All MCL cases were used in subsequent analyses. The remaining (non-MCL) genomes in GAMBL (n=1054) were utilized for building the background model for finding

regions of recurrent SVs. MCL and non-MCL genomes were subjected to the same bioinformatic pipelines for pre-processing, sequence alignment and variant detection.

Table 3-1 Clinical and molecular characteristics of cMCL in the BC cohort

Characteristic	Level	BC (N=103)
Sex	Male	77 (75%)
	Female	26 (25%)
Age group	< 57 yr	25 (24%)
	57 - 65 yr	30 (29%)
	66 - 72 yr	23 (22%)
	> 73 yr	25 (24%)
Tissue biopsy	FF	102 (99%)
	FFPE	1 (1%)
Treatment	HSCT + RCHOP	39 (38%)
	Rituximab + Bendamustine	29 (28%)
	Rituximab + CHOP/CVP	11 (11%)
	CHOP/CVP	7 (7%)
	Observation	13 (13%)
	Other	4 (4%)
MIPI	Low	65 (63%)
	Intermediate	27 (26%)
	High	11 (11%)
Biopsy site	Lymph node	90 (87%)
	Spleen	4 (4%)
	Tonsil	3 (3%)
	Other	6 (6%)
MCL35 (N=91)	Low	51 (56%)
	Intermediate	28 (31%)
	High	12 (13%)

Abbreviations: cMCL, conventional MCL; FF, fresh frozen; FFPE, formalin-fixed paraffin embedded; HSCT, hematopoietic stem cell transplant; RCHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone; CHOP, cyclophosphamide, doxorubicin, vincristine, prednisone; CVP, cyclophosphamide, vincristine, prednisone; MIPI, mantle cell lymphoma international prognostic index

3.2.2. Recurrent structural variations in B-cell lymphomas

While some hematological malignancies are characterized by a single common chromosomal rearrangement such CCND1 in MCL or MYC in BL, it is appreciated that these are not sufficient for lymphomagenesis, instead requiring the acquisition of secondary mutations. Akin to other mutation types, SVs can contribute to carcinogenesis, but a minority of SVs represent drivers with the remainder being passenger mutations. Since drivers provide a selective advantage, these events are expected to be observed in a population of cancers more frequently than expected by chance. Identifying SV drivers therefore requires finding SVs clustered in a region across multiple tumours above a background rate of the occurrence of SV within that region. The background SV rate in this case would be a model of neutral selection in the tumour, that is, provide no selective advantage to the tumour but occur because of external factors such as chromosomal fragility.

To identify putative SV driver events in MCL, I began by creating a background model using SVs from MCL, DLBCL, FL, MM, CLL and BL. I split the genome into 100 Kbp regions (bins) and tabulated the number of SVs that occur within each bin across all tumours. A generalized linear model was used to model the SV counts along with potential covariates that may influence the propensity of a region to acquire breaks. These covariates included genomic characteristics (GC content, mappability, repetitive elements, DNAse sites, common fragile sites, gene density) and regulatory elements (chromatin states from Roadmap, and regulatory regions from ENCODE). This model was then fit onto SV counts per lymphoma type and the model fit was assessed using lambda values. Lambda values ranged from 0.99 and 1.07 indicating the background model provides a good fit between the expected and observed counts (Figure 3-1). Bins that accrued SVs more than would be expected by chance were deemed significant (adjusted $P < 0.1$). Significant bins within 1 Mbp were merged, and each bin was annotated with a probable target using a lymphoma driver gene list.

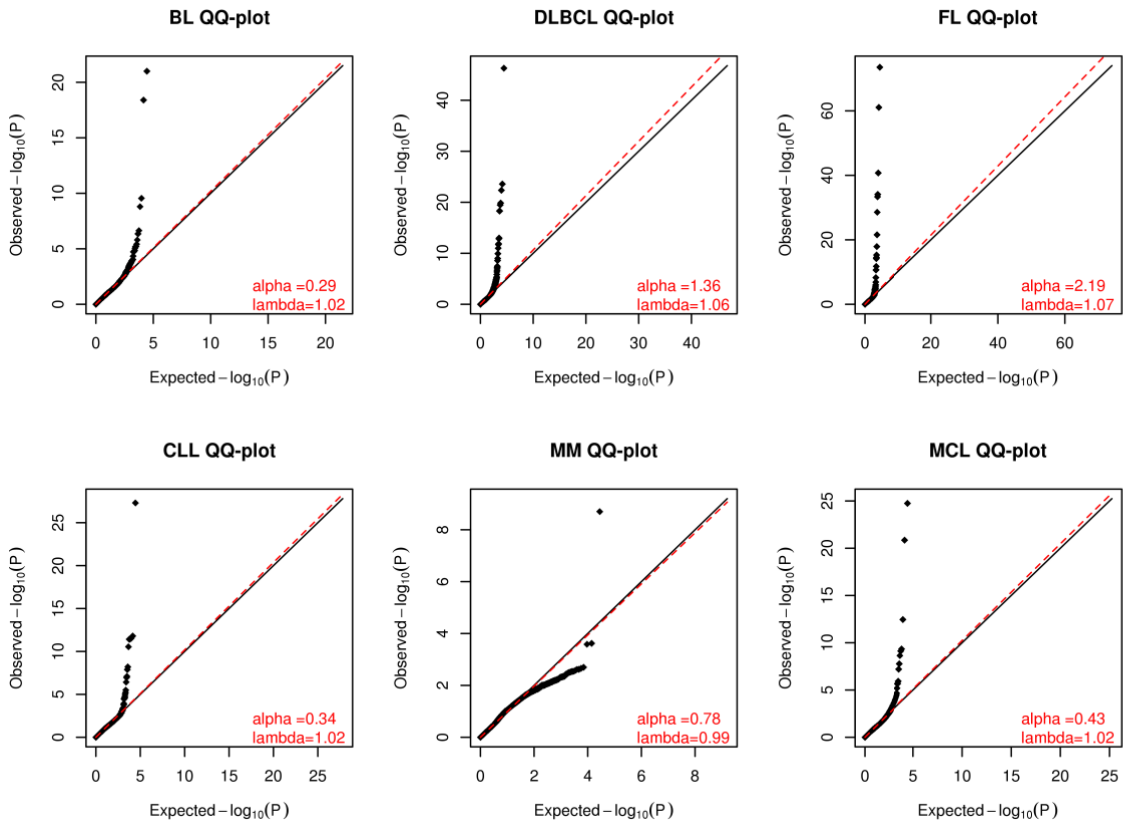

Figure 3-1 Q-Q plots of model fits for BL, DLBCL, FL, CLL, MM and MCL
Assessment of model fit for various B-cell lymphomas. Red dotted line represents extrapolated slope between expected and observed negative $\log_{10}$ P-values. Black solid line represents slope between expected and observed negative $\log_{10}$ P-values when model perfectly predicts observed counts. Genomic inflation factor (lambda) measures the ratio between red, dotted, and black, solid lines, where values closer to one is better. Alpha value is the estimate of the shape parameter for the Gamma distribution chosen for the model, where higher is better.

Among the regions identified, physiologic rearrangements within the three immunoglobulin loci were the most frequently affected by SV. Common SVs associated with the various lymphomas were also identified by this method, supporting its ability to detect SV drivers (Figure 3-2). For example, inter-chromosomal rearrangements affecting *BCL2* were the most common rearrangement in DLBCL and FL and SV involving *MYC* were the most common in BL and rearrangements upstream of *CCND1* were the primarily seen in MCL. This approach is not restricted to detection of translocations as deletions affecting *DLEU1* and *RNASEH2B* were the most frequent in CLL.

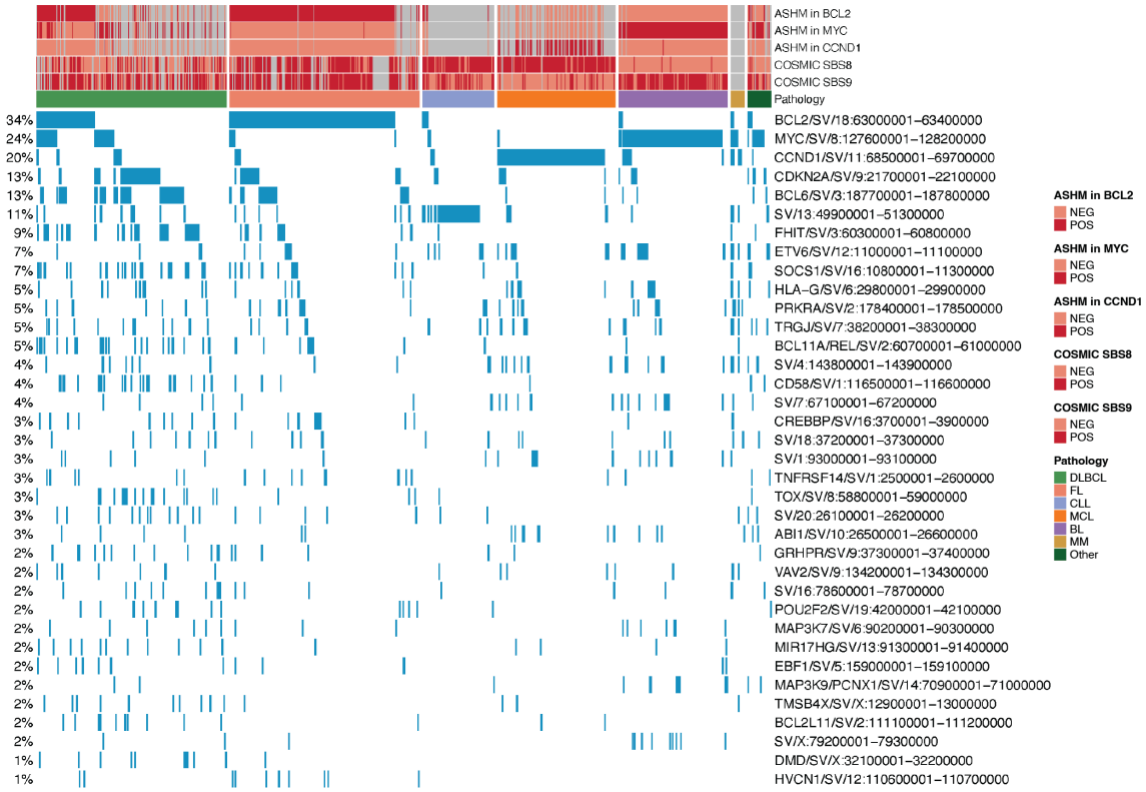

Figure 3-2 Oncoplot of recurrently altered bins across DLBCL, FL, CLL, MCL, BL and MM

3.2.3. *CCND1*::*IGH* rearrangements in MCL

Unsurprisingly, rearrangements within 200 Kbp of *CCND1* were the most common SV found within the MCL genomes. In the BC cohort, *CCND1*::*IGH* translocations were observed in 77 of 103 (75%) tumours (Figure 3-3; Supplemental Data 7). Three of these tumours also contained breakpoints between 33 kbp and 81 kbp telomeric of *CCND1* and mapping to IGH indicating an insertion of *CCND1* into the heavy-chain locus. Of the 77 cases with *CCND1*::*IGH* translocations, 68 (88%) tumours had breakpoints mapping to the IGH D-J gene cluster on chromosome 14. Tumours with D-J breakpoints were found to map to the 5' flanking region upstream of the affected D/J gene within putative RSS motifs, suggesting that these breakpoints were RAG-mediated translocations occurring during D-J joining. Two tumours were found with breakpoints mapping to the IGHV gene cluster (3%). In one tumour, the breakpoint mapped in the 5' of IGHV4-34 in an intergenic region, while in the second tumour, the breakpoint mapped to within the gene body of IGHV4-39. Four (6%) tumours had breakpoints mapping

centromeric of IGHJ gene cluster within the emu enhancer. The region lacked any RSS motifs and was telomeric of IGHM raising the possibility that these breakpoints occurred during class-switch recombination instead of V(D)J recombination. Lastly, additional two tumours (3%) mapped to within IGH constant gene cluster.

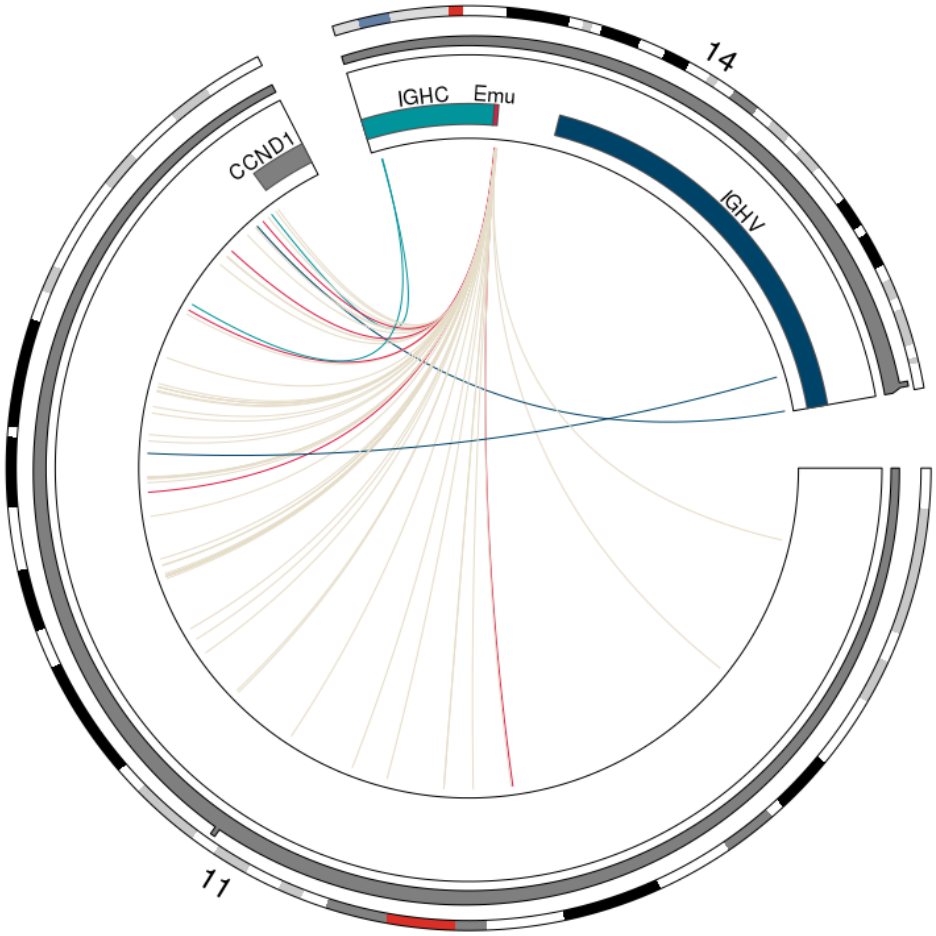

Figure 3-3 *CCND1*::*IGH* rearrangements in BC cMCL tumours
Detected rearrangements between *CCND1* (chr11) and IGH (chr14) locus are coloured by location of IGH breakpoints: blue, IGHV region; beige, IGHJ and IGHD region; red, Emu enhancer; teal, IGHC region.

I observed that some cMCLs acquired mutations within aberrant somatic hypermutation regions (aSHM) across the genome. Furthermore, mutations affecting CCND1 were localized to 5' flanking, promoter and first exon regions which was characteristic of aSHM (detailed in Chapter 4). Given that breakpoints on chromosome 11 varied in distance from *CCND1*, we investigated whether there existed a relationship

between *CCND1* breakpoint distance and total aSHM counts. I hypothesized that tumours with breakpoints closer to *CCND1* would have greater number of aSHM within the genome, a possible indication that proximal breakpoints were hypermutation-mediated[186]. Although there was not clear relationship between distance and aSHM counts, I observed that chromosome 11 breakpoints of high aSHM count tumours were largely found within 100kbp of *CCND1* start site (Figure 3-4).

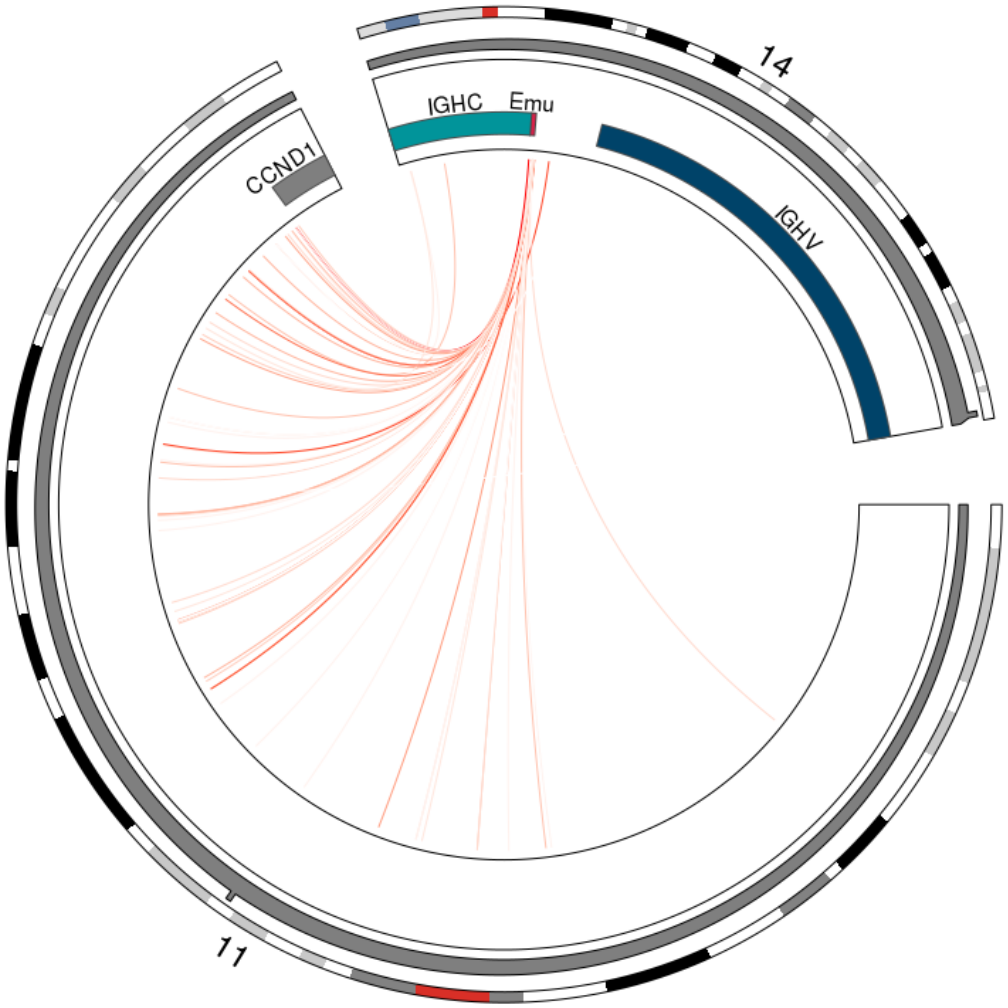

Figure 3-4 aSHM counts in relation to *CCND1*::*IGH* rearrangements in BC and Barcelona MCL tumours

Detected rearrangements between *CCND1* (chr11) and IGH (chr14) locus are coloured by total aSHM counts. Colour gradient from white to red correspond to 0 and 30 aSHM count, respectively.

Twenty-six tumours (25%) had undetectable *CCND1*::*IGH* translocations based on the WGS data. In some cases, MCL are known to harbour a translocation to one of the immunoglobulin light-chain loci rather than IGH[14,17,187]. Of these twenty-six tumours, only five had breakpoints linking the *CCND1* locus and one of the light-chains. In all cases however, the resulting derivative chromosome would be place *CCND1* several megabases away from the light-chain locus, leaving their relevance unclear. Besides translocations involving the light-chains, *CCND1* negative MCLs have been reported to contain rearrangements affecting either *CCND2* or *CCND3*[14]. After searching for these specifically, I was unable to find evidence of these rearrangements among the 26 cases. The reasons for lack of detection could be due to several factors including reduced sensitivity by the Manta algorithm or the region may have undergone complex rearrangements involving more than two chromosomes. While I only used Manta for this analysis, application of other structural variant callers may be able to identify the missing rearrangements. Additionally, an in-depth investigation of complex structural variations may be warranted. Despite not detecting *CCND1* rearrangements in these cases, these tumours appear to be MCLs by standard diagnostic criteria and other molecular evidence. MCL is associated with high CCND1 and, in the case of cMCL, high SOX11 expression. Both SOX11 and CCND1 gene expression were elevated in all MCL cases including those lacking a detectable t(11;14) relative to normal B cells and other B-NHLs (Figure 3-5). Although there were some outliers, the MCLs as a whole have markedly higher MCL marker expression than the other samples.

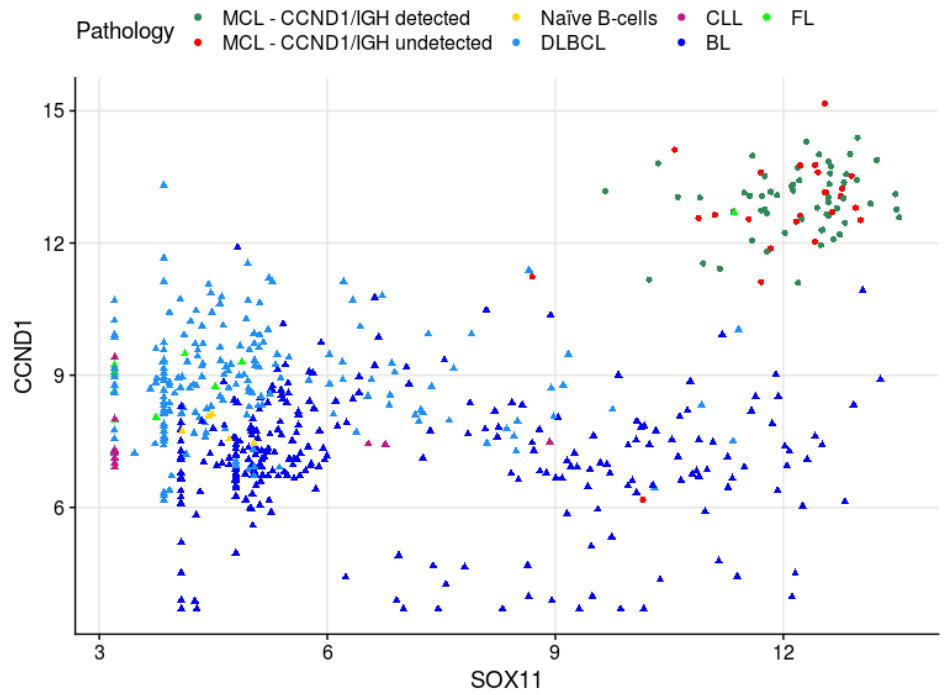

Figure 3-5 *CCND1* and *SOX11* expression of B-cell lymphomas and naïve B cells

MCL tumours were coloured red and green and are indicated by circles. Non-MCL B-cell lymphomas and naïve B cells are indicated by triangles.

3.2.4. *ABI1* frequently altered by structural variations in MCL

Besides *CCND1*, I noted two additional regions were enriched for SVs in MCL. In the first, SVs upstream of *MTF2* on chromosome 1p22.1 were part of intrachromosomal rearrangements spanning up to 1.2 Mbp on chromosome 1. These breakpoints were annotated as either deletions (5) or duplications (2) but a concomitant change in copy number was not supported by read-depth based copy-number analysis. Closer inspection of this region in IGV also failed to show characteristic changes in coverage at the boundaries of the events. While these may simply be rare events, due to the lack of corroborating read support, I excluded this region from subsequent analyses.

The second region that was uniquely enriched for SV in MCL genomes was 10p21.1. The region contained seven genes, with intra-chromosomal breakpoints overlapping *ABI1* and *PDSS1* observed in four unique tumours, and inter-chromosomal rearrangements upstream of *PDSS1* in two different tumours (Figure 3-6). The

translocations in 10p21.1 both affected an intergenic region on 1q42.2 within 75 kbp centromeric of *MAP10* in two different patients. Gains or amplifications affecting 10p21.1 were found in 27 tumours (26%) including 7 cases with copy-states of 4 or more. Taken together, the existence of recurrent SVs, and increases in copy number indicate a selective pressure to increase the expression of one or more genes on 10p21.1 (Figure 3-6A).

To identify the gene(s) most affected by the observed SVs and CNVs, I utilized matching RNA-seq expression data (where available). First, I stratified tumours according to their SV status and/or copy number status and performed two separate comparisons. First, I compared expression between mutated, unmutated MCLs. I separately compared the expression of these genes within mutated MCLs to the expression within available naïve B cell samples. Of the 20 genes, that reside in the recurrently altered region determined by GISTIC, only two (*EBLN1* and *ABI1*) exhibited differential expression between mutated (CNVs and SVs) and unmutated tumours ($P < 0.05$, Mann-Whitney U-test; Figure 3-6C). Upon further investigation, *EBLN1* expression was low across all three categories and the differences appear to be driven by outlier tumours. In contrast, *ABI1* expression appears affected by CNVs overlapping the gene. Interestingly, *ABI1* is normally expressed at lower levels in MCL relative to naïve B cells, where in the normal samples *ABI1* expression appears to be tightly regulated (Figure 3-6D). I speculate that *ABI1* may be downregulated within MCL upon transformation, but in a subset of tumours, structural variations that increase *ABI1* expression provide some survival benefit and are subjected to positive selection.

Amplifications and gains in 10p21.1 region had been previously reported in MCL, however, the target gene nominated previously was *BMI1*. When performing the same comparison, *BMI1* expression could not be differentiated between normal, unmutated and mutated tumours suggesting that this gene may not be the target of the structural variations (Figure 3-6E).

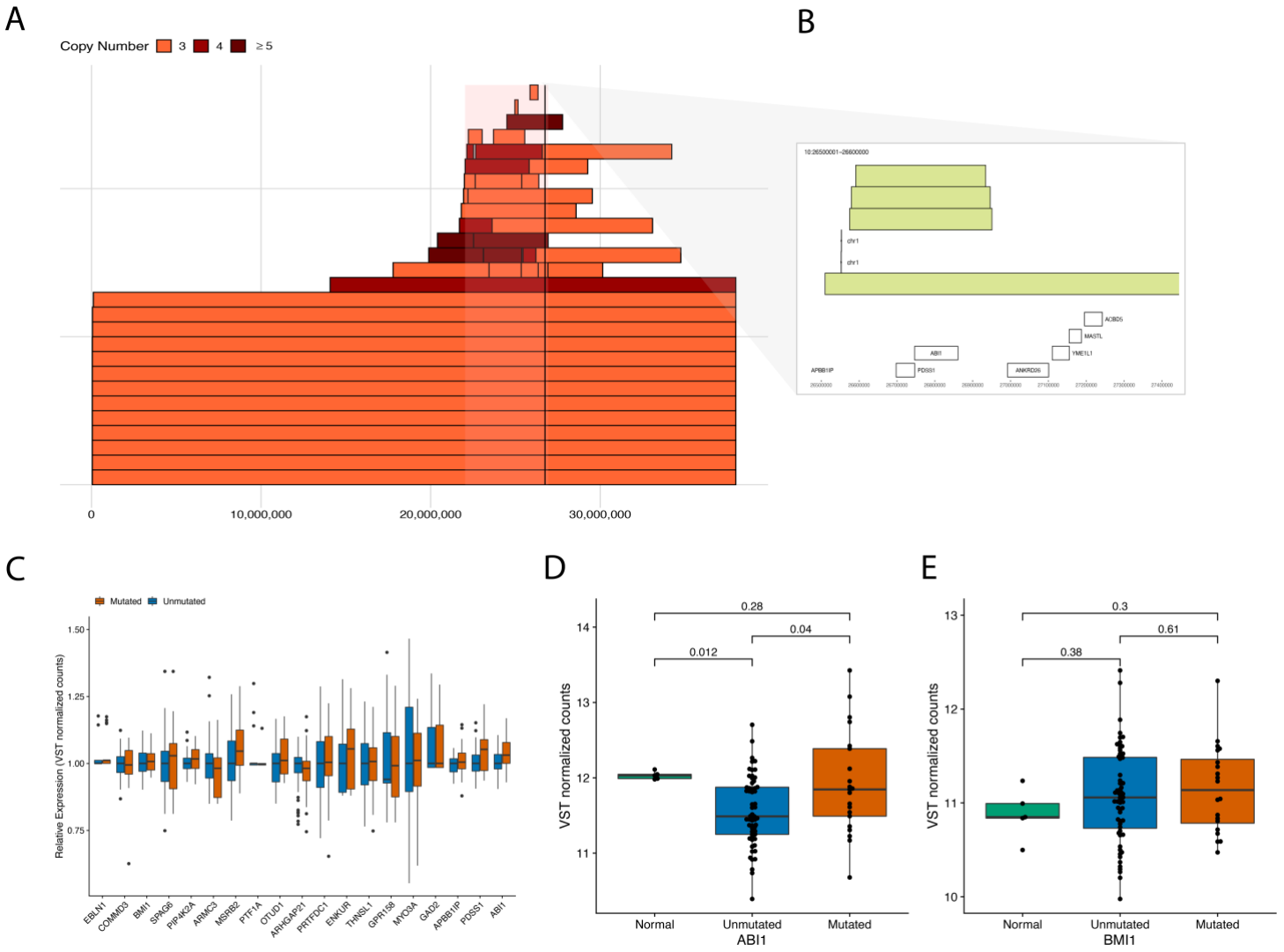
A
Copy Number 3 4 ≥5
0
10,000,000
20,000,000
30,000,000
B
10:26500001-26600000
chr1
chr1
AC3D5
MASTL
ABI1
YME1L1
APBB1IP
POSS1
ANKRD26
C
Mutated
Unmutated
Relative Expression (VST normalized counts)
1.50
1.25
1.00
0.75
EBLN1
COMMD3
BMI1
SPAG6
PIP4K2A
ARMC3
MSRB2
PTF1A
OTUD1
ARHGAP21
PRTFDC1
ENKUR
TMNSL1
GPR158
MYO3A
GAD2
APBB1IP
POSS1
ABI1
D
0.28
0.04
0.012
VST normalized counts
14
13
12
11
Normal
Unmutated
Mutated
ABI1
E
0.3
0.61
0.38
VST normalized counts
13
12
11
10
Normal
Unmutated
Mutated
BMI1

Figure 3-6 Chromosomal alterations affecting *ABI1* on chromosome 10
(A) Overlapping CNV gains and amplifications in 10p21.1. Black solid line shows location of *ABI1* relative to CNVs. Pink overlay shows boundaries of GISTIC2.0 peak. (B) Structural variations found overlapping and upstream of *ABI1*. (C) Gene expression values of genes within GISTIC peak stratified by CNV and SV status. (D) *ABI1* and (E) *BMI1* gene expression between mutated and unmutated tumours along with expression in naïve B cell normal samples.

ABI1 (ableson interactor-1) encodes a major regulator of actin polymerization acting through the large macromolecular WAVE complex to influence cytoskeletal remodeling, lamellopedia protrusions, and micropinocytosis. It was originally identified as an interactor of the ABL1 kinase. ABI1 dysregulation has been observed in multiple cancer types and has an influence of prognosis, albeit with conflicting results. Some data supports its role as tumour suppressor, such as recurrent deletions and reduced expression of ABI1 in prostate cancer [188]. In contrast, ABI1 was shown to positively regulate breast cancer cell proliferation, migration and invasion[189,190], and ABI1 was found to sustain epithelial-to-mesenchymal transition in ovarian cancer cells by acting through the SOS1/EPS8/ABI1 complex to transmit signal from Ras to Rac[191,192].

Interestingly, the expression of ABL1 itself appears to be uniformly upregulated in MCL compared to naïve B cells, consistent with a potential role of ABI1 in this neoplasm. As a downstream target of ABL1, ABI1 has been suggested as having possible oncogenic function in ABL-driven malignancies such as chronic myeloid leukemia and acute myeloid leukemia. Oncogenic fusion protein *ABL1*::*BCR* was found to phosphorylate ABI1 within WAVE complex causing localization to the plasma membrane to induce actin polymerization[193]. Although ABL1 appears upregulated in MCL, the substrates of *ABL1*::*BCR* may not be entirely shared with wild-type ABL such that the regulation of ABI1 outside malignancies harbouring *ABL1*::*BCR* remains speculative. Given the success of targeted interventions in other lymphoid cancers, the confirmation of ABI1 deregulation and functional exploration of its role and possible interplay with ABL1 in MCL are warranted.

3.2.5. Integrative analysis of copy number variations

By applying the GISTIC2.0 algorithm to all cases profiled by WGS, I identified 20 significant regions affected by CNVs or "peaks" (Q-value < 0.1). All CNVs used in GISTIC2.0 are detailed in Supplemental Data 8. These were evenly split between regions affected by amplifications and those affected by deletions (Figure 3-7; Table 3-2). Most of these intersect with recurrent CNVs that have previously been documented as features of MCL such as gains of 3q25, 7p22-22, 8q21, 10p11-12, 18q11-q23 and losses of 1p13-p31, 6q23-27, 8p21, 9p21-22, 9q21, 11q22-q23, 13q11-q13, 13q14-q34[45,111,184,194–200]. Even among these established CNVs, many still lack a definitive gene that represents the functional target of the event, leaving their role in lymphomagenesis unclear. To address this, I performed a similar integrative analysis (using matched RNA-seq) to determine which genes were most affected by CNV. For each gene within each peak peaks, tumours were stratified into mutated and unmutated groups based on the copy number status of the gene and gene expression values were compared (Supplemental Data 9). Additionally, expression from naïve B cells were included to assess the physiological expression of genes.

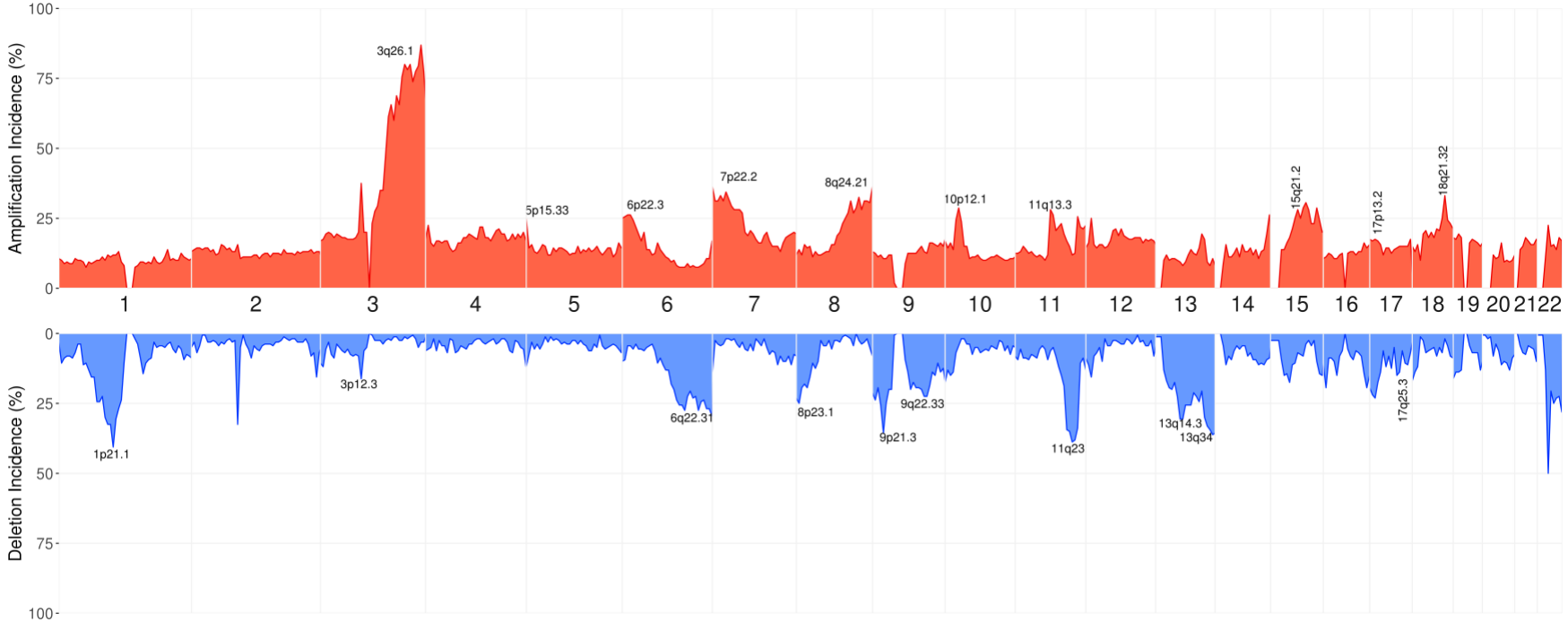

Figure 3-7 CNV incidence rates in MCL

Amplification (red) and deletion (blue) incidences per 100kbp regions across 160 MCL tumour genomes. Recurrently altered regions identified by GISTIC2.0 are annotated with the putative target gene based on gene expression, or with GISTIC2.0 cytochrome band when putative target is absent.

Table 3-2 Recurrently amplified and deleted regions in MCL

Type	Bands	Q-value	Frequency (N = 160)
Amp	3q22.3-3q29	2.81×10^{-3}	70 (44%)
Amp	5p15.33	5.22×10^{-3}	11 (7%)
Amp	6p25.3-6p22.3	2.97×10^{-3}	18 (11%)
Amp	7p22.3-7p12.2	1.38×10^{-2}	26 (16%)
Amp	8q24.13-8q24.23	8.64×10^{-8}	26 (16%)
Amp	10p12.31-10p12.1	1.30×10^{-2}	13 (8%)
Amp	11q13.3-11q13.4	1.38×10^{-2}	18 (11%)
Amp	15q21.2	1.76×10^{-4}	15 (9%)
Amp	17p13.3-17p11.2	9.49×10^{-4}	10 (6%)
Amp	18q21.32-18q22.1	2.97×10^{-3}	16 (10%)
Del	1p22.1-1p21.1	1.60×10^{-28}	53 (33%)
Del	3p12.3	1.86×10^{-4}	15 (9%)
Del	6q16.3-6q25.3	6.86×10^{-15}	34 (21%)
Del	8p23.3-8p21.2	1.96×10^{-2}	15 (9%)
Del	9p21.3	2.82×10^{-15}	34 (21%)
Del	9q21.31-9q31.1	1.58×10^{-5}	30 (19%)
Del	11q22.3-11q23.1	1.66×10^{-39}	57 (36%)
Del	13q14.2-13q14.3	2.05×10^{-7}	50 (31%)
Del	13q32.3-13q34	2.05×10^{-7}	51 (32%)
Del	17q25.3	4.13×10^{-3}	18 (11%)

Deregulation of RB1 via the CCND1-CDK4/6 axis

The 11q13.3 amplification peak overlaps *CCND1* and there was a total of 8 genomes with focal amplifications of this gene. *CCND1* expression was not significantly different between cases having and lacking this CNV. This is likely confounded by the influence of other cis regulatory mutations such as point mutations and structural variations (mostly deletions) affecting the 3'UTR. These can introduce a pre-mature polyadenylation signal via upstream truncating and frameshift mutations, or by physical loss of the 3'UTR due to deletions and structural variations thereby possibly removing RNA-destabilizing element[115,201,202]. Accordingly, structural variations and deletions were observed in 14 tumours that overlapped *CCND1* 3'UTR (Figure 3-8) with six tumours harbouring focal deletions within or spanning the UTR. Four tumours showed inversions overlapping the 3'UTR including one tumour with an inversion of the coding exons of

CCND1 that separates them from the 3' UTR. Similarly, one tumour showed tandem duplication event overlapping the majority of *CCND1* except for the UTR. Lastly, three tumours displayed inter-chromosomal breakpoints within the 3'UTR and mapping to chromosome 14. These effectively represent an insertion of *CCND1* (lacking its 3' UTR) into the IGH locus. When the 3'UTR mutated tumours are included, *CCND1* expression was higher among the five tumours with any cis regulatory mutation than among the remaining tumours (P-value = 0.029; Mann-Whitney U-test).

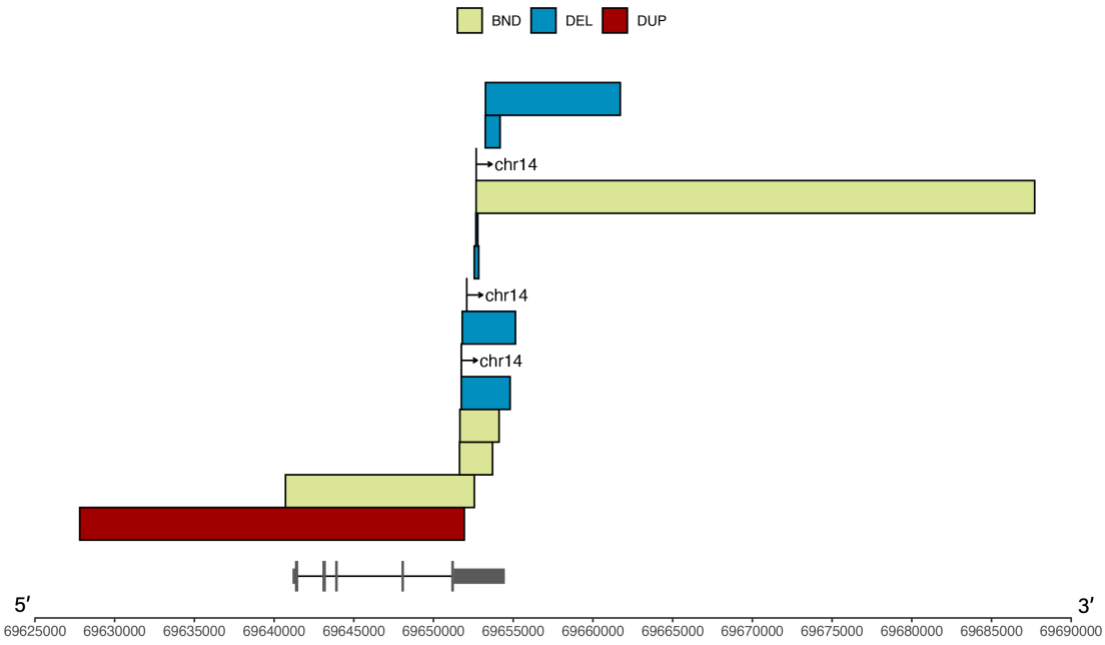

Figure 3-8 Structural variations overlapping *CCND1* 3'UTR in BC and Barcelona cohorts

Manta derived SVs overlapping CCND1 3'UTR are displayed coloured by SV type. BND, inversion; DEL, deletion; DUP, tandem duplication. Inter-chromosomal SVs are denoted by an arrow and chromosome location of the matching breakpoint. *CCND1* gene shown at the bottom.

Common focal deletions on 9p21.3 were found in 34 tumours (21%) that affect both *MTAP* and *CDKN2A*. *CDKN2A* encodes for two structurally distinct tumour suppressor proteins: $p16^{INK4A}$ and $p14^{ARF}$ which inhibit CDK4 and TP53, respectively. It is reasonable to assume that loss of CDK4 could be advantageous to MCL given the central importance of CCND1. Surprisingly, *CDKN2A* expression was similar between deleted and non-deleted tumours. In contrast, *MTAP* expression was significantly reduced in cases with this deletion (Q-value = 9.3×10^{-4}, Mann-Whitney U-test). *MTAP*

encodes a metabolic protein involved in catalyzing adenine and methionine salvage. *MTAP* was often co-deleted with *CDKN2A* leading to aggressive tumours with poor prognosis. Recently, *MTAP* loss was found to be synthetic lethal to *MAT2A* loss[203–205] and inhibiting MAT2A in *MTAP*-deleted tumours reduced proliferation of tumour cells[206]. Alternatively, *PRMT5* encodes an arginine methyltransferase that interacts downstream of *MTAP*. Inhibition of PRMT5 in MCL led to restoration of cell-cycle regulation, apoptosis, and negative regulators of BCR signaling[207]. Importantly, wild-type *TP53* and *CDKN2A/MTAP* deletions were positive prognostic indicators of response to PRMT5 inhibition indicating synthetic lethal relationship between PRMT5 and CDKN2A/MTAP[207].

Lastly, deletion of 13q14.3 which encompass *RB1* are observed in 50 tumours (31%) and *RB1* expression was downregulated in deleted tumours (Q-value = 5.2×10^{-4}, Mann-Whitney U-test). The variety of structural variations affecting components of the CCND1-CDK4/6-RB1 axis appears to be a common feature in MCL. Additionally, evaluation of chromosome 17 revealed several amplifications and gains encompassing the entirety of 17p. This deviated from the expectation that 17p would acquire deletions due to the presence of *TP53* in this region. I examined whether these 17p amplifications affected *TP53* expression and found that expression was not different between mutated and unmutated tumours. However, tumours with deletions were associated with reduced *TP53* (P = 0.0016; Mann-Whitney U-test).

Integrative analysis of CNVs and expression

The most frequent aberration was amp(3q26.1) spanning a 64 Mbp region between 3q22.2 to 3q29 and was gained in 44% (70/160) of tumours. In total, 161 genes were significantly over-expressed in amp(3q26.1) tumours compared to the remaining cases with no CNV in this region (Q-value < 0.01; Mann-Whitney U-test). 3q amplifications had been reported in other B-cell lymphomas where *BCL6* is considered the target[208–210]. In MCL however *BCL6* was not found upregulated due to CNVs (Q-value = 0.11). Of the twenty within the region having the most extreme gene expression difference relative to normals (based on fold-change), *PLCH1*, *MB21D2*, and *GP5* were most affected by gains and amplifications (Figure 3-9). Further supporting their relevance, each of these genes was highly expressed in MCL compared to naïve B cells and appear to be further up-regulated by the presence of gains or amplifications.

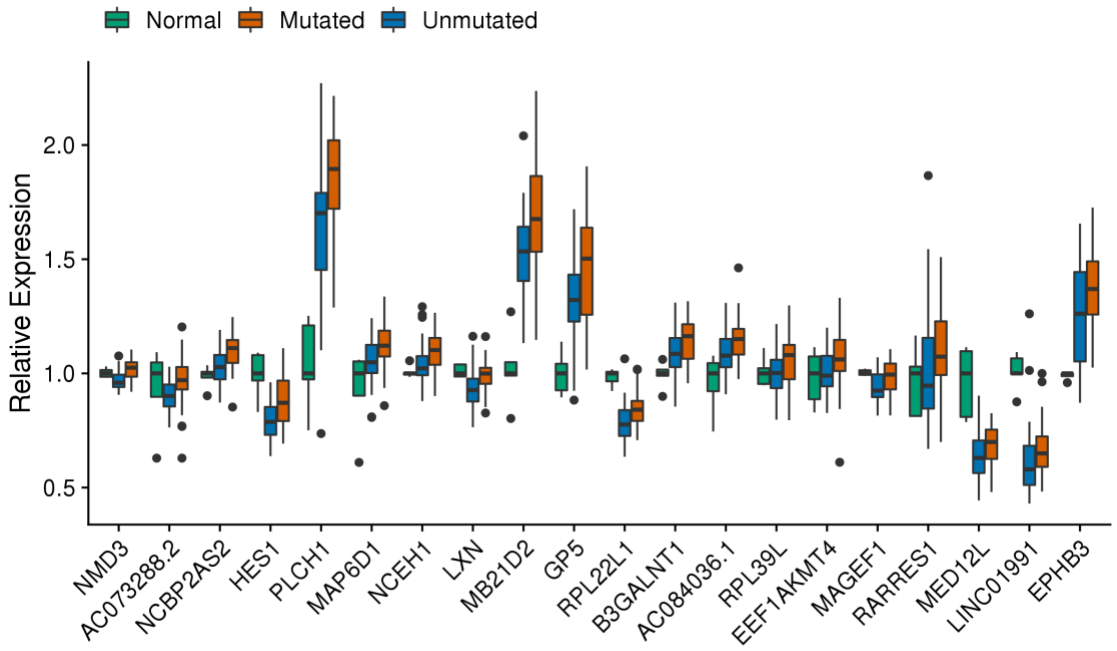

Figure 3-9 Variance-stabilized expression differences between amp(3q26.1) mutated, unmutated and normal samples.

Displaying the top 20 of 161 significant genes based on fold-change between MCL count median and naïve B cell count median. Gene expression of each was normalized to the median expression of the gene within naïve B cell normal samples.

Another common CNV was gains and amplifications affecting 6p22.3, which were observed in 11% (18/160). In this region, only two genes were significantly up-regulated at the mRNA level: *MYLIP* and *PRPF4B*. *PRPF4B* encodes for PRPF4, a pre-mRNA splicing factor containing Arg/Ser domain on its N-terminus, while the C-terminus domain shares homology with cyclin-dependent kinases and mitogen-activated protein kinases[211]. PRPF4 is an integral part of the U4/U6 small nuclear ribonucleoprotein[212,213]. Park et al[214] showed that PRPF4 influenced cell growth, migration, invasion and apoptosis in breast cancer cells and Gao et al[215] showed knockdown of *PRPF4* in pancreatic and colorectal cancer cell lines reduced cancer cell viability and demonstrated that PRPF4 is amenable for pharmacological inhibition. Targeted overexpression of *PRFP4* adds further support to the role of RNA-metabolism dysregulation in MCL.

Twelve genes were significantly upregulated in 7p22.2 amplified region (Figure 3-10). This region represents an arm-level event spanning 50Mbp and was observed in

16% (26) tumours. Among the 12 overexpressed genes, cytochrome oxidase C assembly factor 19 (*COX19*) showed increased mutant expression above both unmutated and normal cells (Q = 0.036; Mann-Whitney U-test). COX19 is an assembly component of cytochrome oxidase (COX), a terminal enzyme of the electron transport chain during oxidative phosphorylation. Due to its involvement in respiration, selective overexpression of COX19 may associated with increased respiratory need within the tumour cell. Interestingly, inhibition of COX19 by mir-21 in non-small cell lung cancer reduced cell proliferation and increased apoptosis[216]. Increased expression of COX19 may play in attenuating apoptosis and increasing proliferation.

Like COX19, Kelch-like Family Member 7 (*KLHL7*) displayed increased expression due to 7p22.2 gains and expression was overall higher in MCL tumours than in normal naïve B cells. *KLHL7* encodes for a protein that complexes with Cullin3 (CUL3) and facilitates proteosomal degradation via ubiquitination of target proteins[217]. Increased KLHL7 protein expression has been observed in various cancers and may be a marker of poor prognosis[218,219].

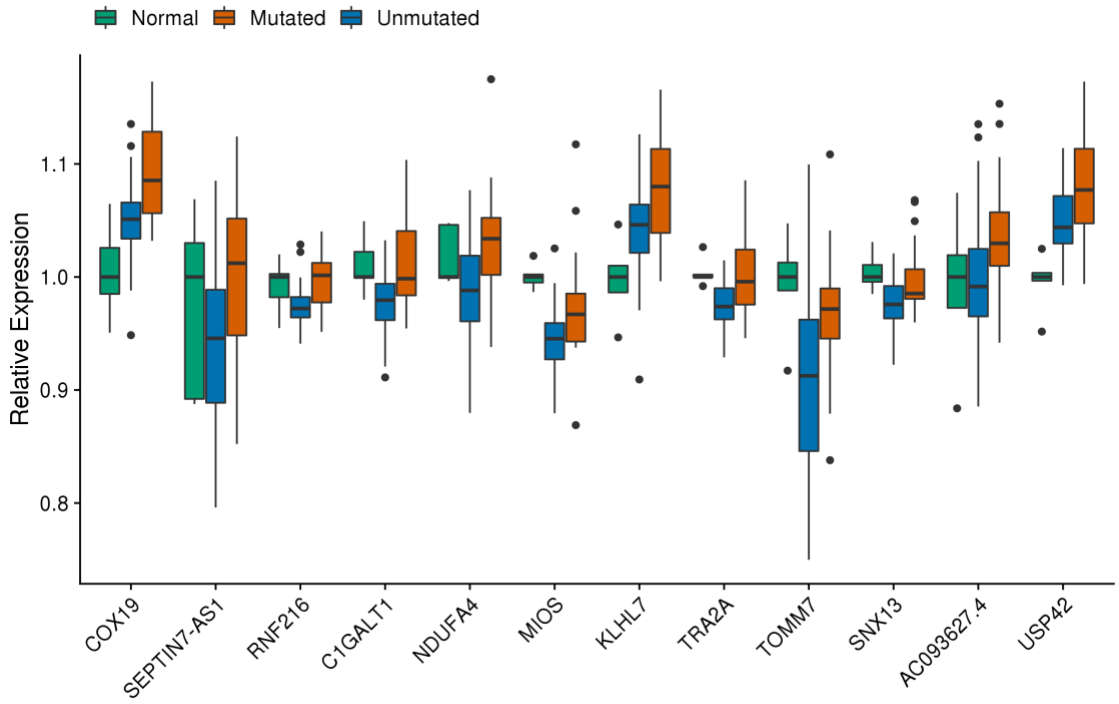

Figure 3-10 Variance-stabilizied expression differences between amp(7p22.2) mutated, unmutated and normal samples.

Gene expression of each was normalized to the median expression of the gene within naïve B cell normal samples.

Gains and amplifications around 8q24.21 were observed in 16% (26/160) of tumours. The amp(8q24.21) peak region spanned between 8q24.13 and 8q24.23 and was approximately 17 Mbp. This peak overlapped *MYC* and two genomes with focal amplifications surrounding this gene were noted. However, gene expression revealed that MYC is downregulated overall in MCL compared to naïve B cells (Figure 3-11). Only two genes within amp(8q24.21) were up-regulated: *FBXO32* and *MIR7874*. *FBXO32* encoded for F-Box containing protein that functions as the substrate recognition component of SCF E3 ubiquitin-protein ligase and therefore directs target proteins for degradation via the proteosome. Interestingly, FBXO32 was shown to specifically target MYC for proteasomal degradation and inhibited MYC activity within HEK293T cell lines[220].

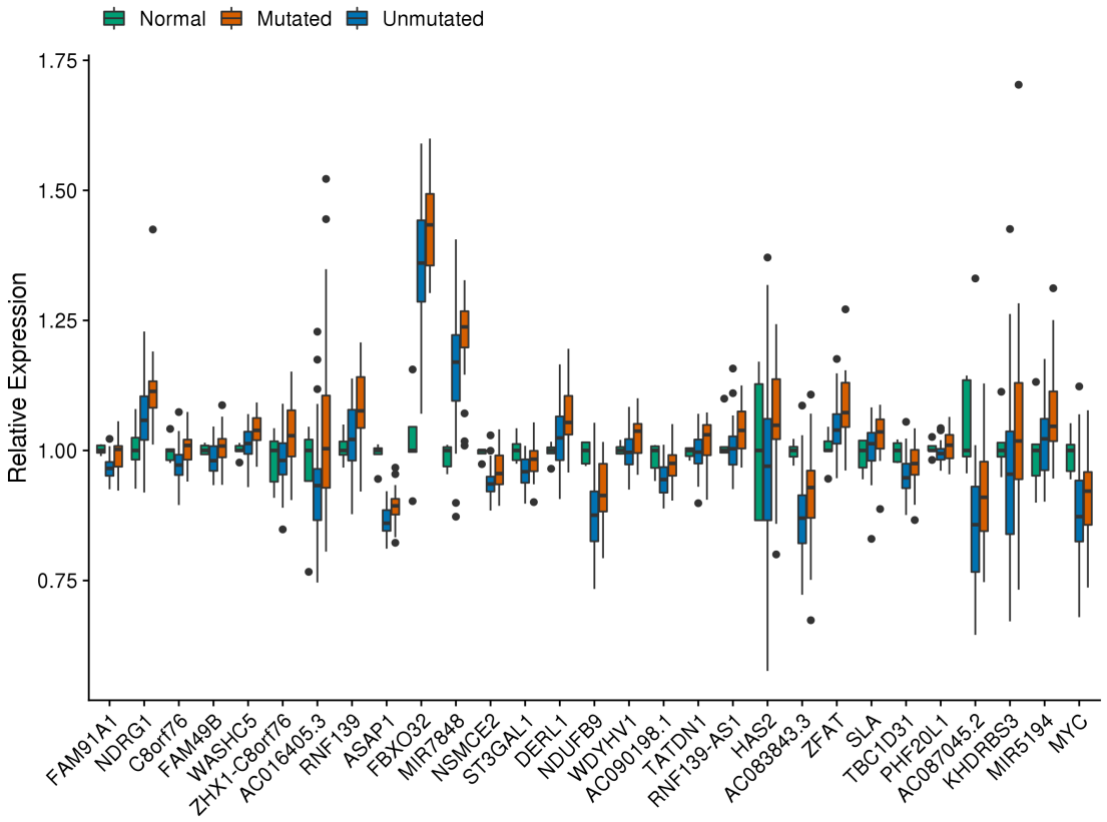

Figure 3-11 Variance-stabilizied expression differences between amp(8q24.21) mutated, unmutated and normal samples.

Gene expression of each was normalized to the median expression of the gene within naïve B cell normal samples.

Amp(15q21.2) peak region was mutated in 9% (15/160) of tumours with 8 tumours harbouring focal gains and amplifications affecting the *BCL2L10* gene (Figure 3-12). Expression of genes within the peak revealed *BCL2L10* overexpression in copy-number altered tumours compared to other MCL and to naïve B cells. Interestingly, an additional copy-gain peak was found overlapping 18q21.32-18q22.1 (10%; 16/160) with focal amplifications in four tumours overlapping *BCL2*. Both *BCL2* and *BCL2L10* encodes an anti-apoptotic factor involved in regulating the intrinsic apoptosis pathway[221,222] and BCL2 was known to be over-expressed in MCL[223]. Although 18q21 amplifications have been reported in MCL, the relevant gene affected by this event as not, to my knowledge, been nominated. Up-regulation of *BCL2L10* and *BCL2*, along with mutations and deletions of *TP53*, implicate avoidance of the intrinsic apoptosis pathway in MCL.

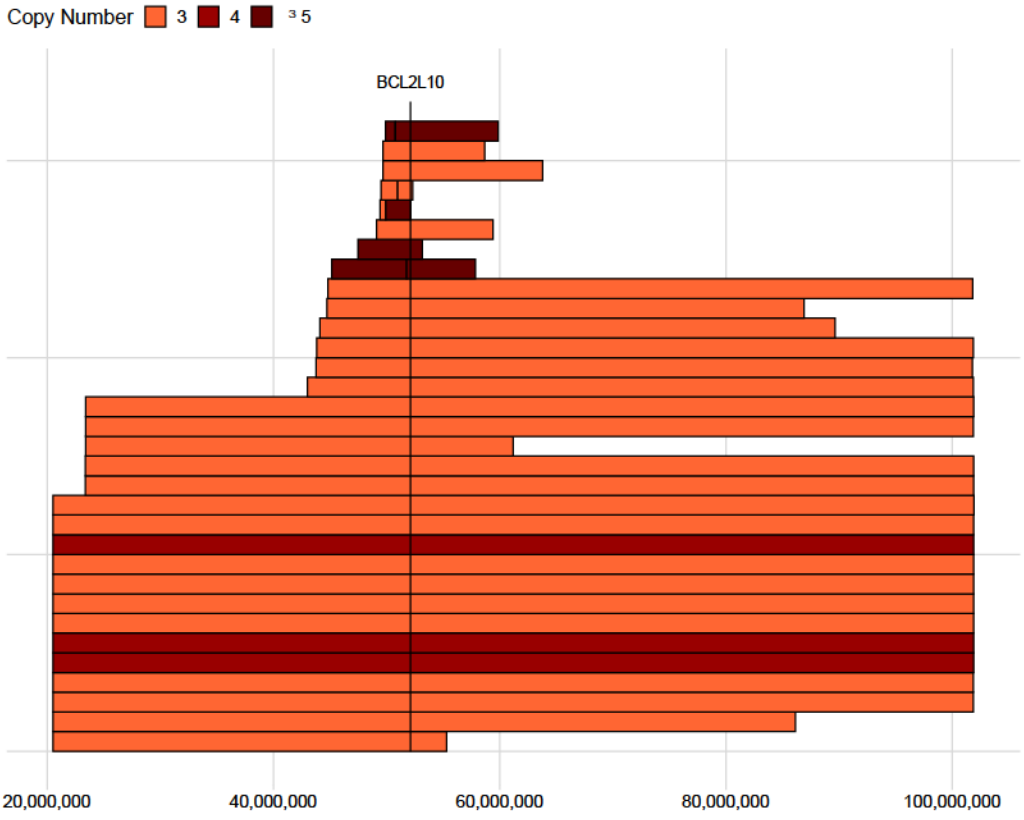

Figure 3-12 Copy number gains and amplifications overlapping 15q21.26 GISTIC peak

Gains (2 < CN < 4) and amplifications (CN ≥ 4) CNVs overlapping 15q21.26 GISTIC (15:49880428-52116453; GRCh38). Location of *BCL2L10* denoted by black solid line.

There were multiple regions where gene expression analysis did not reveal any significant gene expression differences affected by CNV. These regions included amplifications affecting 11q13.3 (11%; 18/160), 5p15.33 (7%; 11/160), 17p13.3-17p11.2 (6%; 10/160), and deletions affecting 3p12.3 (9%; 15/160) and 17q25.3 (11%; 18/160). Deletions on 3p12.3 overlap *ZNF717* and the first two exons of *ROBO1*, however, neither gene exhibited reduced expression in these cases. The 17q25.3 deletion peak was 1.2Mbp in size and overlapped 55 genes.

3.3. Methods

Whole-genome sequencing data used in the validation cohort in Pararajalingam et al[224] were incorporated with additional cases from BC Cancer patients, and whole-genomes from Nadeu et al[45] for this analysis. Cases which underwent whole-genome sequencing were selected for RNA sequencing with an additional 27 tumours undergoing only RNA-sequencing. In total, this analysis utilized 160 MCL cases whole-genomes and 129 cases RNA-seq.

3.3.1. Read mapping

Whole genomes sequenced at the BC Genomes Sciences Centre were aligned to the human reference genome (GRCh38) lacking alternate contigs. Alignment was executed using BWA mem (version 0.7.6; parameters: -M)[167]. Aligned genomes from Nadeu *et al*[45] in BAM format were downloaded from EGA (EGAD00001006025).

3.3.2. Copy number variation analysis

Regions containing copy number variations (CNV) were identified using Battenberg for cases with matched normal DNA, and Control-FREEC for cases without a matched normal[225,226]. Control-FREEC was run using a hard-masked reference file obtained from UCSC. The Battenberg R package was forked from the development version and modified to allow the usage of chromosome prefixed reference files. CNVs found by Battenberg were converted to SEG format using a custom Python script. To reduce noise, each seg file was filtered using CNVfilteR (version 1.6.1) to keep segments that were concordant with VAFs of overlapping SSMs[227]. Tumour SEG files were processed to fill in missing genomic regions with copy-neutral segments using a custom Python script and merged. The merged SEG file was as input for GISTIC2.0 (parameters -genegistic 1 -broad 1 -conf 0.99 -gcm extreme -js 8 -qvt 0.5 -td 0.5 -armpeel 1 -maxspace 5000 -v 30 -cap 2)[228]. GISTIC2.0 was run using an hg38 reference file supplemented with miRNA genes. Gains and amplifications were defined by

GISTIC2.0 where segments with absolute copy numbers between 3 and 4 were considered gains, and segments with copy numbers greater than 4 were considered amplifications. Heterozygous and homozygous deletions were not separated for analyses in this thesis. Resulting wide peaks were filtered to retain peaks greater than 1Mbp and had Q-value less than 0.1. Any overlapping peaks were merged to prevent double counting in subsequent analyses. Peaks less than 20 Mbp in length were considered focal and were annotated using a curated lymphoma driver gene list.

3.3.3. Structural variation analysis

Large scale structural variations (SV) supported by paired and split-read evidence were discovered using Manta on default settings in tumour/normal mode with an hg38 reference file[229] (Full list of MCL SVs detailed in Supplemental Data 10). Manta VCF outputs containing somatic SVs were converted to BEDPE format using svtools vcftobedpe (https://github.com/hall-lab/svtools). In addition to MCL tumours, Manta was run on tumour-normal pairs from diffuse large cell B-cell lymphoma (DLBCL), Burkitt lymphoma (BL), follicular lymphoma (FL), chronic lymphocytic leukemia (CLL), multiple myeloma (MM), and other B-cell lymphomas. Sequencing data for non-MCL cases were a combination of samples sequenced at BC Genomes Science Centre and data downloaded from public repositories. For non-MCL tumours that were aligned to a reference besides hg38, Manta was run using the appropriate reference file and the resulting SV BEDPE files were converted to hg38 coordinates with liftOver (version 3.6.6) and “hg19tohg38” chain file provided by UCSC Genome Browser.

Genomic regions containing SVs in multiple tumours may be evidence of positive selection, but such clusters are confounded by the presence of mapping and genomic artifacts. To distinguish between regions containing potential oncogenic drivers and passenger events, the region-specific background SV rate was estimated by fitting a negative binomial generalized linear model (GLM) using the fishHook R package (version 0.1) (https://github.com/mskilab/fishHook). The genome was split into 100kbp regions (i.e., bins) and the number of SVs per region per tumour was tabulated using SVs from all lymphoma types, thus providing the cohort-wide SV rate per bin. To prevent inflation of bin counts by highly rearranged tumours, each tumour could contribute a

maximum of one count per bin. Heterochromatin and genomic gap coordinates were downloaded from UCSC genome browser, and bins that overlapped these regions were excluded.

The background SV rate of any given bin may be influenced by a variety of factors. To build an accurate model of the background SV rate, 28 covariates were initially selected for GLM fitting. From UCSC Genome Browser, SINE, LINE, LTR positions from the repeatmasker track, ENCODE CD20 DNAase positions, genomic GC content, genomic mappability, and candidate cis-regulatory elements were downloaded in BED format. Genomic fragile sites were downloaded from HumCFS database[230]. Predicted chromatin states from B-cell epigenomes were downloaded from RoadMap database[231]. Lastly, gene density within a 3 Mbp sliding window (1 Mbp offset) was included as a model covariate. Stepwise regression of the linear model using all covariates was performed and the model that achieved the highest AIC value was selected. The optimal model included 26 covariates. This model was then fit to each lymphoma type SV count separately and significantly enriched bins after multiple-testing correction were identified (Q-value < 0.1; Bonferroni). Bins that were within 1Mbp of another bin were merged and were annotated using a curated lymphoma driver gene.

3.3.4. Quantification of gene expression

Gene expression was quantified at the transcript level using salmon (version 1.10)[232]. RNA-sequencing reads were pseudo-aligned to the human transcriptome (Gencode release 33). Transcript level counts were imported into R (version 4.1.3) using tximport (version 1.22.0) and summarised to gene counts[233]. Top 1000 most variably expressed were used to investigate batch effects by principal component analysis (PCA) in R. PCA components were visualized in R using ggplot2 (version 3.3.5). Batch effects associated sample storage were removed by including sample storage variable in the DESeq2 (version 1.34.0) model[234]. Corrected counts were then normalized for library size and heteroskedasticity was lessened using variance-stabilization DESeq2. The resulting transformed counts were used for all RNA-seq analyses and visualizations. For the integrative analysis, all genes with available RNA-seq expression data within GISTIC2.0 peaks were analyzed.

3.3.5. Statistical analyses

R (version 4.1.3) was used to perform Mann-Whitney U-test and multiple testing correction using Bonferroni method where applicable. P-values below 0.05 and Q-values below 0.10 were considered significant.

Chapter 4.

Consensus clustering reveals MCL genetic subgroups

The following chapter builds on my prior findings by integrating the genetic features from chapters 2 and 3 to resolve the genetic heterogeneity in our MCL genome cohort. I performed and designed all analyses described in this section, including feature selection, model training, and data analysis.

4.1. Introduction

There is no international consensus for the standard treatment of MCL with decisions for specific treatments dependent on region, expertise, access to therapies and on clinical features such as patient age/frailty. The exploration of new treatment options or combinations that may extend survival of MCL patients is ongoing. The recent SHINE trial[235] evaluated the efficacy of combining standard treatment with either ibrutinib or placebo in previously untreated, older MCL patients. While patients treated with ibrutinib showed modest increase in progression-free survival relative to the control arm, complete response rates between both arms remained the same. Furthermore, an exploration of the efficacy of ibrutinib within MIPI stratified patient groups showed that the ibrutinib treatment was not better than that of placebo within the high-risk patient population. These results highlight that generalized application of treatments are, at best, leading to modest improvements in patient outcome and highlight the need for superior approaches. One avenue of improvement may be the application of precision medicine whereby patients are treated based on the specific genetic characteristics of their respective tumours. Such an approach would enable future clinical trials to consider the interplay between a treatment's mechanism of action and tumour genetics. Theoretically, this could enable the identification of patients with the greatest chance of benefiting from a particular treatment, thereby improving outcomes beyond the modest results achieved thus far in randomized controlled trials.

The genetic landscape of MCL delineated in the preceding chapters represents a substantial expansion of our understanding of the genes and biological pathways that

are commonly perturbed in this neoplasm. While these mutations have advanced our knowledge into MCL aetiology, these results also affirm the notion of widespread genetic heterogeneity among patients. Although these results add further complexity to the study of MCL, the genetic diversity may provide an opportunity to form the basis for understanding the known clinical heterogeneity of the disease. Prior efforts to delineate clinical differences among MCL patients led to the current classification, which is based on putative cell-of-origin (i.e., cMCL and nnMCL). While this has cemented differential application of treatments between cMCL and nnMCL patients, due to the residual clinical heterogeneity within cMCL, clinical management is an ongoing challenge. As cMCL cases represent the majority (approximately 90% of new diagnoses)[33,34,236] and cMCL have shorter survival in general, this subgroup requires further attention.

Until recently, DLBCL has been conventionally divided into two subgroups. However, for nearly 20 years, clinical trials attempting to apply precision medicine to DLBCL have been unsuccessful. The two original molecular subgroups were termed germinal centre B-cell like (GCB) and, post-germinal centre B-cell like (ABC). In a research setting, these subgroups enabled greater understanding of oncogenic processes occurring in DLBCL including finding treatable genetic pathways unique to each subgroup. However, clinical trials of therapies exploiting these genetic features largely failed to improve patient outcome[237]. Using mutation, recurrent copy number and SV information, Schmitz et al[143] and Chapuy et al[238] showed that DLBCL comprised a more diverse collection of genetic subgroups. The eventual establishment of the LymphGen classifier by Wright et al[239], extended nomenclature and subgroups by Schmitz et al[143]. LymphGen classified tumours into one of seven groups: MCD, BN2, EZB-MYC+, EZB-MYC-, ST2, A53, and N1.

Each DLBCL subgroup has mutations in pathways that might indicate therapeutic vulnerabilities. For example, it has long been thought that ABC-DLBCL would be susceptible to inhibitors like ibrutinib which perturb BCR signaling[240]. Similarly, MCD tumours commonly feature *MYD88*^{L265P} and *CD79B* mutations, which indicate reliance on pathways normally activated by B-cell receptor signalling. Based on the hypothesis that BCR signaling inhibition would benefit this subset of DLBCL patients, the PHOENIX trial was conducted to evaluate the efficacy of adding ibrutinib to R-CHOP in previously untreated DLBCL[241]. Evaluation of the PHOENIX through the lens of LymphGen classification groups showed improved performance of ibrutinib administered to MCD

tumour patients versus placebo[242]. This study demonstrated the feasibility and effectiveness of applying rational therapeutic approach for a non-Hodgkin lymphoma. This approach could be applied to other lymphomas such as MCL, but an understanding of the genetic subgroups is currently incomplete.

Motivated by the potential benefits afforded by the recent transition of DLBCL classification from a cell-of-origin based scheme to genetic subgroups, I sought to explore whether a similar approach may be appropriate for MCL. Specifically, I hypothesize that the underlying complexity of MCL may be more adequately described through its genetic features, thereby informing on clinical and biological heterogeneity. Similarly, a recent study attempted to apply genetic subgrouping to MCL, demonstrating the potential utility of this approach. Yi et al[243] used clustering to dissect the genetic heterogeneity of 134 MCLs. Their approach resembled that applied to DLBCL by Chapuy et al[238] including the use of non-negative matrix factorization (NMF) clustering. Their features included SNVs within their candidate MCL driver genes, recurrent CNVs, and IGHV mutation status where IGHV was designated as being mutated in a tumour if the variable region shared less than 97% identity with the germline sequence. Their analysis resulted in four clusters (C1, C2, C3, C4). Interestingly, these clusters displayed differential clinical outcomes (overall survival and progression-free survival) with the largest difference between C1 and C4. Tumours in C1 had the most favourable prognosis and consisted primarily of indolent nnMCL cases, thereby recapitulating a known group. Unsurprisingly, one of the primary features associated C1 was IGHV mutation status, which is a feature of nnMCL. They also noted an enrichment of cases with mutations in *CCND1* and amplifications affecting this locus (11q13). Patients assigned to clusters C2 and C3 had similar survival, with average OS intermediate to C1 and C4 tumours. The majority of C2 cases harboured deletions 11q overlapping *ATM* and concomitant *ATM* SNVs. Furthermore, this group also harboured deletions of 1p21.1, and while not stated in the study, overlapped *S1PR1* which we had previously identified as an MCL driver gene in our exome analysis. Tumours in C3 harboured changes in multiple features without an obvious unifying set of driver mutations. The primary alterations in this group included amplifications in 3q21.1-29 which is commonly observed in MCL, and harboured deletions on 6q25.3-26. Importantly, C4 was enriched for patients with poor prognosis. Accordingly, MCLs in this group contained genetic features already known to be associated with high-risk disease such as deletions of

17p13.3 (*TP53*), *TP53* mutations, deletions of 13q14.2 (*RB1*) and deletions of 9p21.3 (*CDKN2A/MTAP*).

The clusters identified by Yi and colleagues represent the first evidence of genetic subtypes in MCL. However, the genetic clustering approach may be further improved. For example, the number of features could be enhanced by incorporating SV information, or other known genetic changes associated with outcome[201]. Additionally, features that are predicted to affect the same gene could be consolidated where applicable. This is particularly relevant for C2 and C4 which are primarily associated with *ATM* mutations/deletions and *TP53* mutations/deletions, respectively. Failing to combine similar features may be inflating the importance of a select few features in cluster formation.

Notably, in the DLBCL literature the genetic subgroups are in flux, whereby different analytical and/or sequencing methods have led to genetic subgroups with only partial overlap[237] and ongoing work in our lab is exploring how WGS data may lead to improvements and extension of existing classification systems. Notably, the study by Yi et al[243] relied mainly on coding mutations from exome data and did not include other mutations that could be detected by WGS. Lastly, usage of IGHV mutation status as a feature may separate nnMCL from cMCL, but this may also prevent clustering of cMCLs with nnMCLs that would otherwise be genetically similar. To address these problems, I sought to determine whether a more comprehensive set of driver mutations and other mutational features could lead to a more robust set of clusters and possibly novel or more clinically meaningful subgroups in MCL.

4.2. Results

4.2.1. Somatic hypermutation in MCL

Successful application of any machine learning approach, including clustering, depends on the features selected for inclusion. For this analysis, I included the full extent of protein-coding mutations in significantly mutated genes and recurrent copy number aberrations that were described in Chapter 3. Mutations in WGS were identified using a consensus variant calling pipeline developed for the GAMBL project (Figure 4-1;

Supplemental Data 11). I was also interested in improving our ability to distinguish cMCL from nnMCL using genome-wide features. Given the nature of nnMCL and that these tumours likely undergo germinal centre reaction during its development, I hypothesized that the inclusion of suspected passenger mutations resulting from somatic hypermutation (SHM) and aberrant somatic hypermutation (aSHM) may be informative for this separation. We applied the same methods used to find DLBCL and BL aSHM regions to MCL, however the only result was the promoter region of *CCND1*. We hypothesized that MCLs undergoing germinal centre reaction may share aSHM mutational pattern to that of other post-germinal centre lymphomas and so we supplemented this analysis with DLBCL and BL derived aSHM regions. The specific set of regions were comprehensively determined in previous analyses of DLBCL genomes[132] and were subsequently shown to be affected by aSHM in other B-cell lymphomas such as BL[131]. These regions of high mutation burden were hypothesized to be the result of off-target activity of cytidine deaminase (AID), which is active during physiological B-cell development to increase antigen variability within immunoglobulin variable regions.

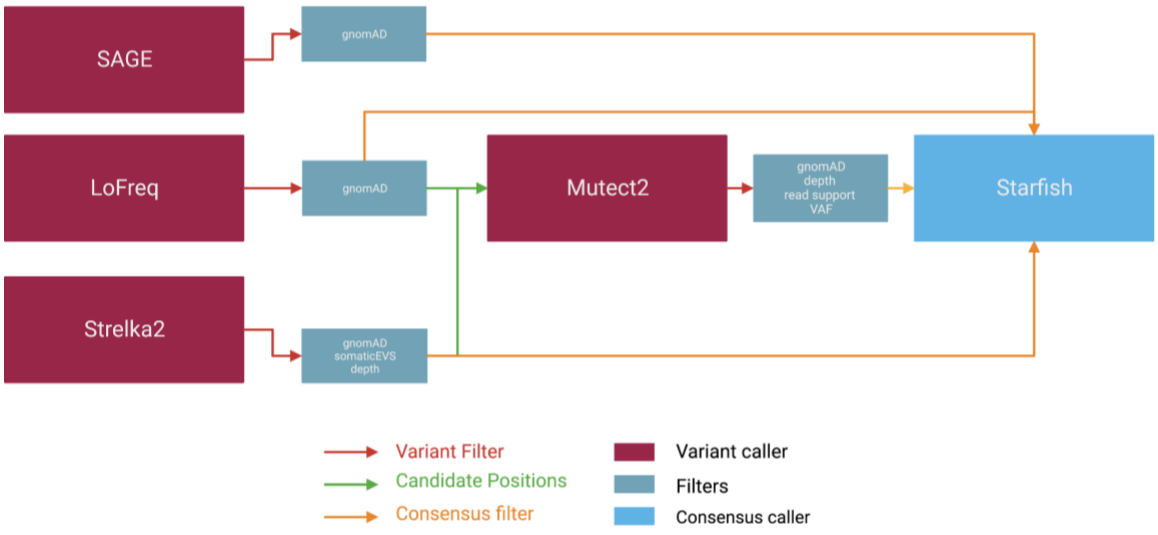

Figure 4-1 Algorithmic view of SSM variant calling pipeline for MCL genomes
Maroon boxes indicate variant calling tools; teal boxes are variant filters used; light-blue box indicates consensus calling software used to finalize variant calls.

In addition to the regions documented as hypermutated in DLBCL and BL, I also sought to find any non-coding regions uniquely mutated in MCL. Inspection of the SNV distribution affecting *CCND1* including 500 bp flank regions, I found that *CCND1* exhibited a mutational pattern akin that of other aSHM regions (Figure 4-2A). Interestingly, there were examples of each of nnMCL and some cMCL with mutations in the 5' UTR and first exon regions of *CCND1*, which could suggest that at least some cMCL cases may undergo germinal centre reactions (Figure 4-2A). Examining the degree to which aSHM regions were mutated across the MCLs revealed that majority of DLBCL/BL aSHM targets had fewer than two mutations in each MCL tumour (Figure 4-2B; Supplemental Data 12). One reason for this difference could be that the off-target potential of an aSHM site may be dependent on the transcriptional activity and chromatin state of the site when AID is active[244]. Since the cell of origin for MCL and DLBCL are distinct, they would be expected to have different sets of active regulatory regions. In contrast, *CCND1* was highly mutated in MCL which coincides with the fact that *CCND1* is highly expressed in MCL and is translocated within the IGH locus, which provides opportunity for mutation by AID.

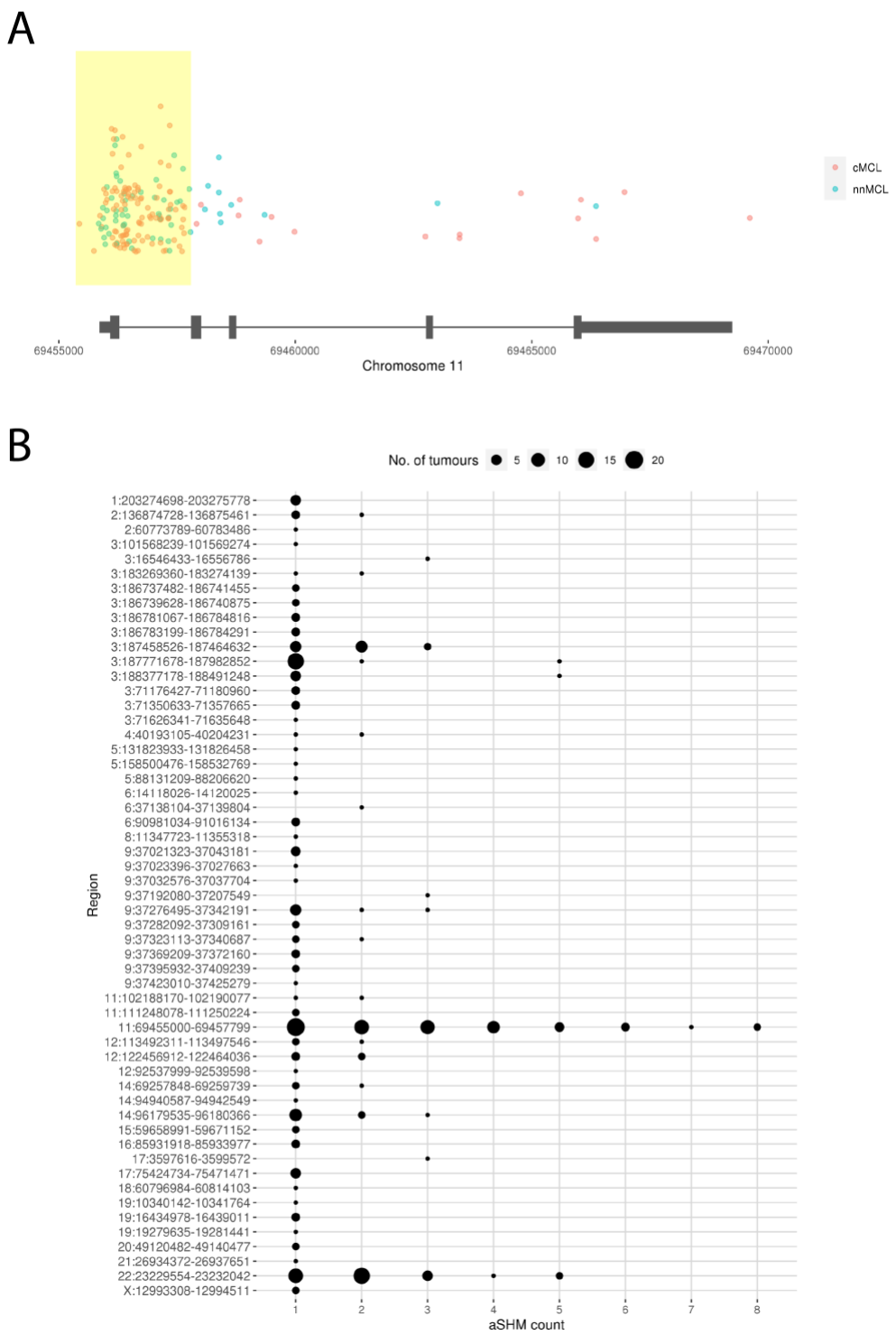

Figure 4-2 Putative aSHM in MCL

(A) Mutation burden of *CCND1* in cMCL (red) and nnMCL (blue). Yellow highlight indicates the region of CCND1 that contained the bulk of mutations derived from aSHM and were used in NMF clustering. (B) Mutation count across GC lymphoma aSHM sites and *CCND1* aSHM regions (hg19 coordinates). Point size indicates number of MCL tumours with a given aSHM count per region. Tumours with zero aSHM mutations were omitted.

4.2.2. Genetic clustering in MCL

To identify genetic subgroups in MCL based on shared biological features, I applied a non-negative matrix factorization (NMF) clustering to 39 genetic alterations identified in chapters 2 and 3 along with mutated aSHM targets (Figure 4-3A). Genetic features which could potentially be affecting the same driver gene were consolidated into single features to reduce the chance of overfitting clusters and avoid double-counting of redundant events. After consolidation and removal of tumours that had fewer than three features, a total of 158 tumours were viable for clustering. We used a standard methodology for performing NMF which relies on testing multiple possible solutions and selecting the rank which maximizes cophenetic and silhouette values (Figure 4-3B). The optimal solution converged on five genetic subgroups with robust genetic features associated with each cluster.

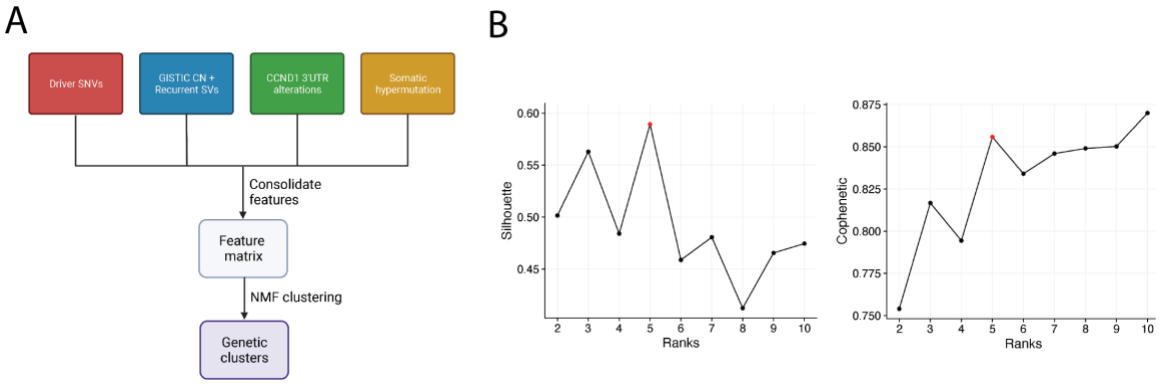

Figure 4-3 NMF clustering schematic and rank selection metrics
(A) SSMs of MCL driver genes, recurrent CNVs and SVs, *CCND1* 3' UTR and somatic hypermutation regions were consolidated into tumour-feature matrix. Consensus clustering on feature matrix was performed using NMF. (B) Ranks between 2 and 10 were tested using nsNMF algorithm and silhouette and cophenetic constants were used to select optimal NMF rank.

Evaluation of survival characteristics of these five genetic subgroups reveals distinct survival outcomes for each cluster. When considering cMCL and nnMCL cases together (i.e., the entire cohort), the clusters differed significantly in overall survival (OS) ($P < 0.0001$; Figure 4-4A). Patients in cluster 1 had the best outcomes compared to other clusters (median OS not reached) and were significantly longer than all other clusters except cluster 2. Patients in both clusters 2 and 3 had intermediate OS with median OS of 12.8 years and 13.3 years, respectively, and were not significantly

different (P = 0.6; log-rank test). Patients in clusters 4 and 5 did not differ in OS (P = 0.15; log-rank test) and had median OS of 5.89 years and 3.07, respectively. Although cluster 1 patients had the best OS, 41% of tumours were nnMCL (13/31) which could explain the apparent increase in overall survival for this group. Considering this confounder, I omitted all nnMCL tumours from cluster one and re-evaluated overall survival. After this, overall survival remained significantly different from other clusters (P < 0.0001; median OS not reached; Figure 4-4B). In other words, patients with cMCL tumours that group within cluster 1 tended to perform better in terms of survival.

Besides OS, clusters were also had significant differences in progression-free survival (PFS) and disease-specific survival (DSS) in only cMCL tumours. Median PFS of clusters four and five were lower compared to remaining clusters (Figure 4-4C). Disease-specific survival followed a similar pattern in cMCL tumours (Figure 4-4D).

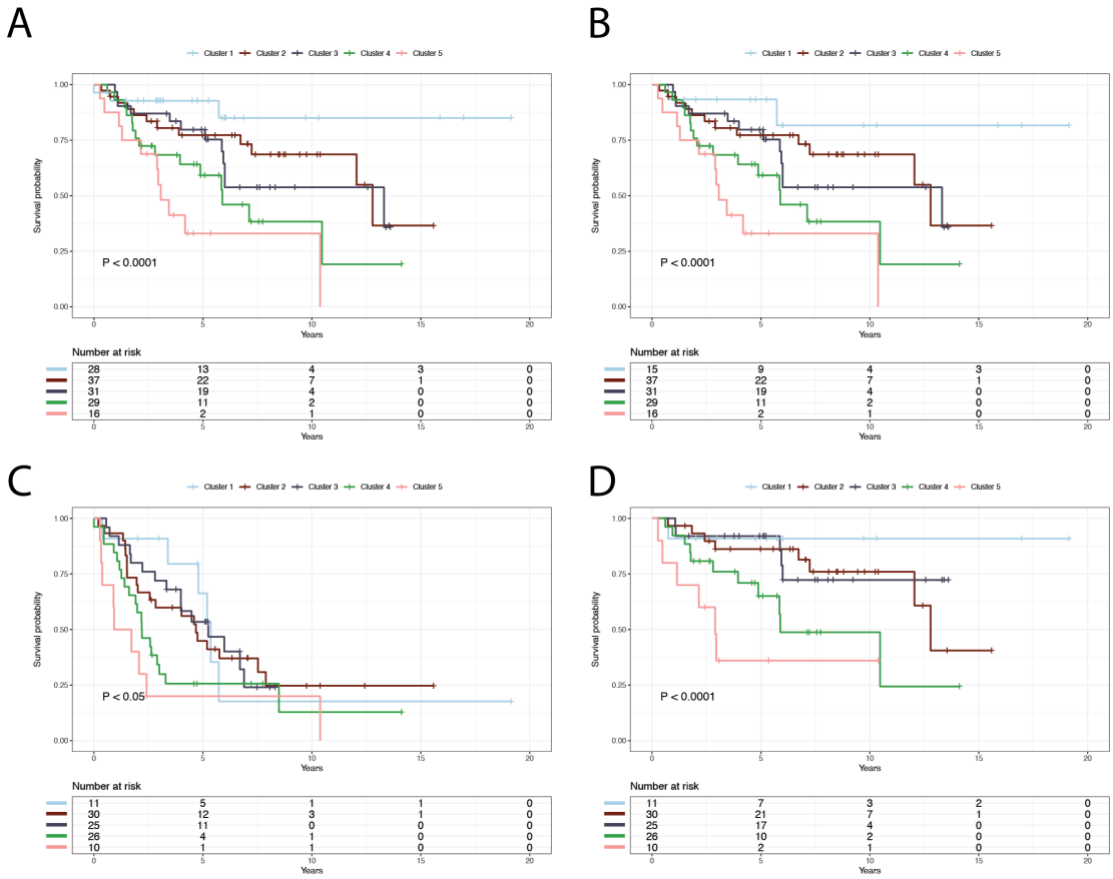

Figure 4-4 Survival impact of NMF clusters
(A) Overall survival (OS) of cMCL and nnMCL tumours stratified by cluster membership. (B) OS of cMCL tumours stratified by clusters. (C) Progression-free survival (PFS) of cMCL tumours stratified by clusters. (D) Disease specific survival (DSS) of cMCL tumours stratified by clusters.

Characteristics of MCL Cluster 1

MCL Cluster 1 (MC1) had the highest representation of SHM features, with the most common mutations affecting the *CCND1* and *IGLL5* loci (Figure 4-5). This cluster contained nearly all nnMCL tumours except for one case in cluster 4. In addition to nnMCL tumours, cluster 1 also contained 18 cMCL tumours. These cMCLs assigned to Cluster 1 also had genetic aberrations affecting other genes that were shared with the other clusters. In other words, the distinguishing feature of Cluster 1 cMCLs is the presence of mutations associated with aSHMs, a feature they share with the nnMCLs.

Owing to the unexpected presence of SHM-associated mutations, the cMCL tumours in this cluster may represent MCLs that have experienced the germinal centre at some point prior to, or during, lymphomagenesis. Evidence of germinal centre reactions may be inferred through the usage of signature analysis where it is expected that mutations generated by hypermutation processes can occur both within and outside SHM regions. To investigate the mutational processes occurring in MCL and to determine whether mutational processes differed within SHM regions, I performed signature analysis using all mutations and separately using mutations only within SHM regions including 1 kbp region containing *CCND1* promoter. After decomposing *de novo* signatures into known COSMIC signatures, signatures SBS9 and SBS84 were prevalent within cluster 1, but were enriched separately in cMCL versus nnMCL tumours. As expected, there was strong evidence of SBS9 in cluster 1 nnMCL cases, which suggests mutations arising from repair by the error-prone polymerase eta (POLH). In contrast, the SBS84 was more common in cluster 1 cMCL, which implies canonical AID activity without repair by POLH. It appeared that while mutational processes associated with germinal centre development occurred in both MCL subtypes in cluster 1, the two tumour types undergo distinct development.

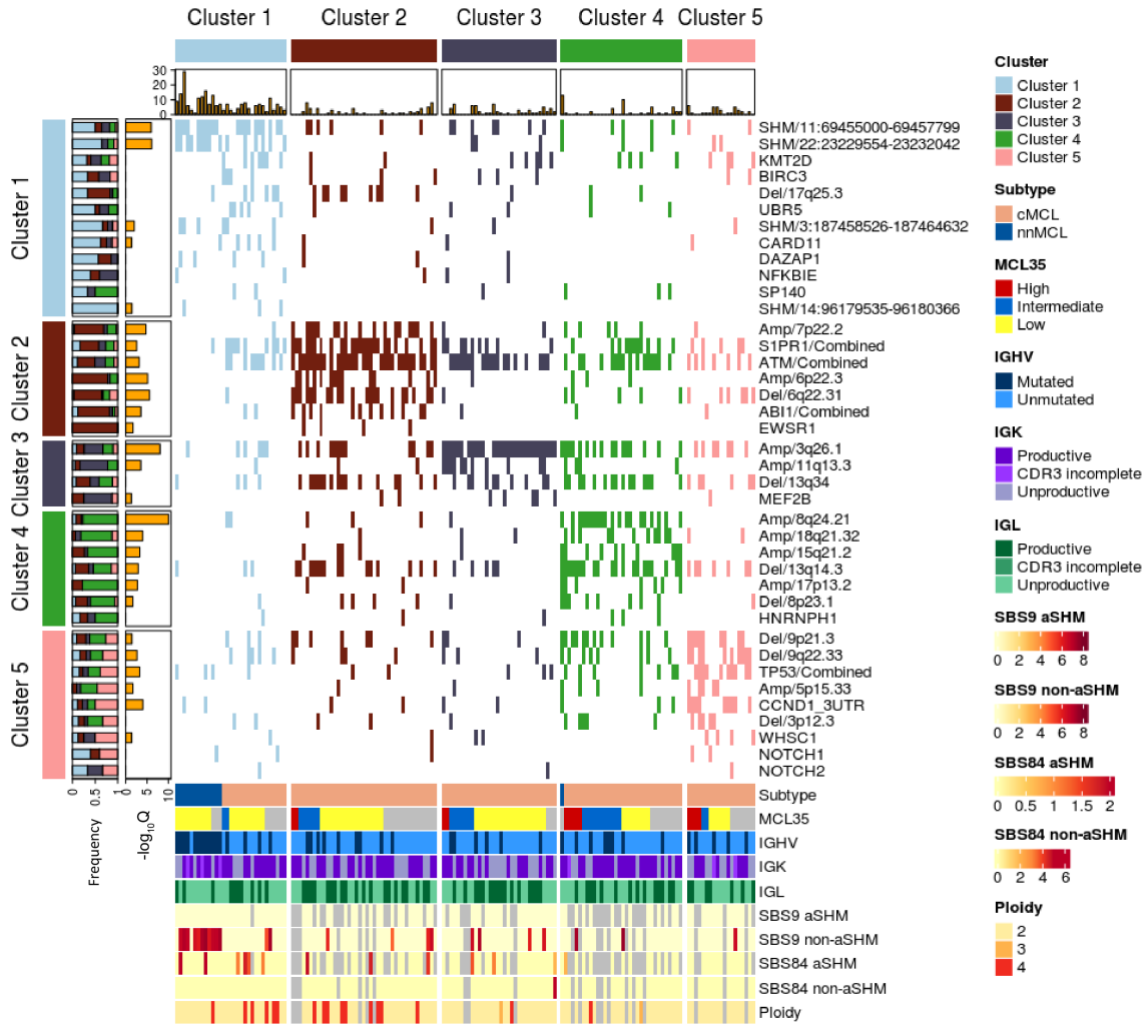

Figure 4-5 Heatmap of tumour and mutational features of NMF genetic clusters
Top barplot: total count per tumour of mutations within aSHM regions. Left stacked barplot: Frequency of mutations for a given feature within each cluster. Left barplot: Fisher's test -$\log_{10}$Q values for a feature's assigned cluster. Abbreviations: IGHV, immunoglobulin heavy-chain variable; IGK, immunoglobulin light-chain lambda; IGL, immunoglobulin light-chain kappa; SBS9, COSMIC signature 9; SBS84, COSMIC signature 84.

Characteristics of MCL Cluster 2

The second cluster of patients (MC2) contained several predominant genetic features. The most abundant were large gains around spanning between 7p22.3 to 7p12.1 with peak at 7p22.2, present in 41% of cluster 2 tumours (17/41). Apart from these amplifications, this cluster also featured deletions and mutations that encompass *ATM* or *S1PR1*. Although not significant (P = 0.051, Fisher's exact test), *ATM* and *S1PR1* mutations and deletions were nearly two-times more likely to co-occur in the

same tumour than not (OR = 1.92). Both aberrations were observed in other cluster tumours especially *ATM*. Another common feature was deletions spanning 6q16.3 to 6q25.3 and a peak at 6q22.31 which was observed deleted in 21 cluster 2 genomes (51%). This cluster also had the highest representation of amplifications and structural variations of *ABI1,* collectively found in 29% of cluster 2 genomes (12/41).

Characteristics of MCL Cluster 3

MCL Cluster 3 (MC3) had the lowest number of unique genetic features, with an amplification peak at 3q26.1 being the most predominant, followed by large amplifications spanning 3q22.2 and 3q29. Copy number gains across this region have been extensively reported in MCL in prior studies. However, as previously discussed, the target of this event is unclear even after my search for cis regulatory effects on gene expression.

Characteristics of MCL Cluster 4

MCL Cluster 4 (MC4) represents the first group enriched for patients with inferior prognosis. The most prominent features of this cluster are amplifications 8q24.1 and to a lesser degree 13q14.3 deletions. Interestingly, many tumours in this cluster also contain aberrations in features associated with cluster 5 such as 9p21.3 deletions, 9q23.33 deletions and *TP53* deletions and mutations.

Characteristics of MCL Cluster 5

The cluster of patients with the worst clinical outcomes is MCL Cluster 5 (MC5). The most prominent features in this cluster are many of the known aberrations for poor prognosis including deletions of 9p21.3 (*CDKN2A/MTAP*), *TP53* mutations and deletions, 5p15.33 amplifications (*TERT*) and *CCND1* 3'UTR mutations.

4.3. Methods

4.3.1. Detecting Simple Somatic Mutations

Simple somatic mutations (SSM), consisting of SNVs and indels, were identified from WGS data for all tumour/normal pairs using a consensus calling approach. This pipeline relies on votes from four somatic variant calling algorithms: Strelka2, LoFreq, SAGE and Mutect2[245–247]. Candidate variants were first individually identified using Strelka2, LoFreq and Sage. Strelka2 (version 2.9.10) variants were post-filtered to remove variants with GnomAD population allele frequencies (AF) greater than 0.0001, SomaticEVS score and depth less than 10. LoFreq (version 2.1.5) and SAGE (version 2.6) variants post-filtered to remove variants with GnomAD AF greater than 0.0001[248]. Candidate variants from Strelka2 and LoFreq were combined and subsequently used as candidate variant positions for running Mutect2 (version 4.1.8.1). Variants called by Mutect2 were filtered to remove variants with GnomAD AF greater than 0.0001, read depths less than 10, less than four reads supporting the alternate allele, and variant allele frequency less than 0.1. Variants from all four variant callers were intersected by position using Starfish (version 0.2.2) and variants that were found by at least three callers were retained making up the final call set. Variants were annotated and converted to MAF format using vcf2maf (version 1.6.18) and Variant Effect Predictor (release 86)[171]. Canonical variants were selected during variant annotation except in cases where more non-synonymous mutations outnumbered in a non-canonical gene isoform.

Regions affected by aberrant somatic hypermutation (aSHM) were examined in MCL using the presence of SSMs. Specifically, variants within a curated set of DLBCL aSHM regions along with variants within the first exon, 500 bp upstream the first exon, and the first intron of CCND1 were counted per tumour. Tumours that had at least two variants within a given aSHM region were considered hypermutated for that region.

4.3.2. SSM mutational signature analysis

De novo and COSMIC single base substitution (SBS) mutational signatures were extracted from the SSM data using the SigProfiler software suite (version 1.1.0)[249]. First, a matrix of tri-nucleotide mutational contexts was tabulated using SigProfilerMatrixGenerator and hg38 human reference. De novo mutation signatures were extracted using a two-phase approach. In the first phase, SigProfilerExtractor was run for each rank between 2 and 25 with iterations set to 30. Next, the optimal number of signatures from the previous step was obtained and SigProfilerExtractor was run again with the optimal signature rank and number of iterations set to 200. De novo signatures were then "decomposed" to COSMIC signatures (version 3.1) with SigProfilerExtractor. COSMIC signature activity values per sample were then processed in R and used for all analyses.

4.3.3. Immunoglobulin gene usage

MiXCR (version 4.3.2) and IgCaller (version 1.2) were used for immunoglobulin clonotyping[250–252]. Since MiXCR considers expressed immunoglobulin transcripts for assessing dominant clonotypes, MiXCR results were selected over IgCaller in cases where both were available. For both MiXCR and IgCaller results, clonotypes were considered mutated if the variable region sequence shared less than 98% similarity to the aligned germline sequence corresponding to the CLL germline sequence threshold.

MCL RNA-seq was run using MiXCR 'analyze shotgun' command. Resulting FASTQ files containing assembled clonotypes were queried against a local copy of the igblastn databse (version 1.17.1) and the highest matching queries were retained for each clonotype. Dominant clonotypes were selected for each sample by considering only productive clonotypes that had the highest clone fraction and were supported by at least 50 reads. IgCaller results were used for cases where RNA-seq was unavailable. IgCaller was run with default settings and references in paired or unpaired mode depending on availability of matching normal samples. Results were further filtered to remove low scoring clonotypes (score < 10) and to remove clonotypes predicted to

produce unproductive transcripts. In cases where more than one viable clonotype was observed in a sample, the clonotype with the highest score was retained.

4.3.4. Consensus clustering using non-negative matrix factorisation

Genetic subgroups were discovered using non-negative matrix factorization (NMF) using the NMF package (version 0.31.1) in R (version 3.6.3). First, a feature matrix of mutations and tumours was created. Non-synonymous somatic mutations in MCL driver genes identified in Chapter 2 were encoded as 1 if a mutation was present in a sample and 0 otherwise. Mutations affecting *HNRNPH1* also included synonymous mutations if they overlapped splice regions flanking exon 4. GISTIC2.0 peaks identified in Chapter 3 were used for copy number variations. High copy amplifications (CN > 3.7) and homozygous deletions (CN < 0.8) were encoded as 2, low-copy gains (2.1 < CN < 3.7) and heterozygous deletions (1.4 > CN > 0.8) were encoded as 1, absence of CNVs were encoded as 0. Structural variations affecting *ABI1* and 3'UTR of *CCND1* identified in Chapter 3 were encoded as 1 if present and 0 if absent. Somatic hypermutation regions were encoded as 1 if present and 0 if absent. Deletion of *TP53* (17p) and non-synonymous mutations affecting *TP53* were encoded as 1 if present and 0 if absent. Mutation features that were shown to likely affect the same gene were collapsed into a single feature which included: 10p12.1 copy gains/amplifications and *ABI1* SVs, 11q23.1 deletions and *ATM* mutations, and 1p21.1 deletions and *S1PR1* mutations. Features were included for clustering if they were observed in at least 3 tumours overall.

The optimal NMF rank and NMF algorithm was selected by testing NMF solutions using ranks between 2 and 10 for each algorithm and selecting the rank and algorithm that achieved the highest silhouette and cophenetic metrics after 30 iterations. An optimum of 5 ranks using the nsNMF algorithm were used for the final model. NMF was subsequently run again with algorithm set to nsNMF, rank set to 5 and iterations set to 100. Heatmaps of clustered mutations and tumours were made using pheatmap (version 1.0.12) in R.

4.3.5. Survival Analyses

Kaplan-Meier curves were used to assess survival differences between NMF-derived patient-genetic clusters. The survival R package (version 3.1-8) was used to estimate progression-free survival (PFS), overall survival (OS) and disease-specific survival (DSS) depending on survival variable availability. Kaplan-Meier survival curves visualized using survminer R package (version 0.4.8). Log-rank test was used to quantify the difference between groups where a P-value below 0.05 was considered significant. OS was defined as the time between diagnosis and time of any death, PFS was the time interval from diagnosis to the first instance of disease progression or death, and DSS was the time between diagnosis and time of death due to disease or acute toxicity of treatment. All Kaplan-Meier curves produced using right-censored data. OS data was available 144 samples including samples belonging to the Barcelona cohort. PFS and DSS values were only available for patients belonging to the British-Columbia (BC) cohort.

Chapter 5.

General Discussion

5.1. Summary of Research Findings

The first two data chapters (Chapters 2 and 3) aimed at comprehensively delineating the pattern and frequency of known and novel genetic alterations in MCL with the goal to expand the current knowledge of MCL oncogenesis. In Chapter 4, I identified genetic clusters of MCLs based on the patterns of these mutations. The genetics of each cluster might imply distinct natural histories and biological features. The patients in some of these genetic clusters had distinct clinical outcomes, indicating the potential prognostic utility of this result beyond current approaches.

In Chapter 2, I describe a meta-analysis of previously published MCL tumour exomes along with extended analysis using newly sequenced exomes and matching RNA-seq data. I employed a highly accurate in-house SNV variant calling approach and multiple driver gene discovery tools to enhance our power to detect significantly mutated genes. This afforded the identification of novel recurrent mutations, such as those affecting RNA-binding proteins, in a subset of MCL tumours. This revealed a previously undescribed mechanism of RNA-metabolism as a target for mutations.

Among the novel genes identified herein, we highlight each of *DAZAP1*, *EWSR1* and *HNRNPH1* as genes encoding RNA-binding proteins with putative functions in MCL. The frequency of deleterious mutations along the length of *EWSR1* gene suggests a role as tumour suppressor in MCL. EWSR1 plays a multifunctional role in non-malignant cells where it regulates transcription and RNA splicing[253]. Recently, the role of EWSR1 was explored in mature B cells using conditional knockout mice[254]. From this study, researchers learned that attenuating EWSR1 in B cells caused an expansion of pre-germinal centre B cells. It is conceivable that reduced EWSR1 protein achieved through loss-of-function mutations as seen in MCL may similarly cause proliferation of pre-germinal centre B cells. In *DAZAP1*, we identified a highly recurrent pattern of missense or truncating mutations near the C terminus, with the latter predicted to remove only the terminal region of the protein including signal peptides. DAZAP1 C-terminus was shown to contain a putative nuclear localization domain that may be disrupted by these

mutations[255]. As a member of the large group of hnRNP proteins, DAZAP1 regulates mRNA splicing[256], transport into the cytoplasm[257], and translation[258]. Sequestration of DAZAP1 into the cytoplasm via loss of the nuclear localization domain will likely hinder DAZAP1 ability to carry out these essential functions. HNRNPH1, another hnRNP protein, accrued non-coding mutations primarily within intronic poly-G motifs upstream of exon 4. Given the existence of alternative isoforms whereby exon 4 was excised out, this led to the hypothesis that the intronic mutations may alter splicing of *HNRNPH1*. Mutated *HNRNPH1* tumours were observed with increase abundance of full-length isoform relative to the alternative isoform lacking exon 4.

Chapter 3 further examines the MCL mutational landscape by focussing on large chromosomal changes such as CNVs and SVs. Here, I relied on WGS data generated from an expanded cohort, which was also used for all subsequent analyses. The use of WGS was necessary for comprehensive identification of breakpoints and accurate resolution of the boundaries associated with CNVs. Although tools already exist for finding recurrent CNVs, at the time of this analysis, tools for inferring recurrent SV targets were not available. To address this deficiency, I developed a model for estimating background SV rate across genomic space using unrelated genomes as a background. This model was applied to a variety of B-cell lymphomas and, supporting its utility, it identified many of the regions that accumulate SVs in DLBCL, BL, CLL, and FL. While I only focused on MCL for this thesis, the additional recurrent SVs identified in other the lymphomas can be the subject of further exploration. In MCL, besides the SVs located centromeric to *CCND1*, the only other region of significant recurrence was the on chromosome 10. This same region harboured amplifications and gains centred around the gene *ABI1*. Using matching RNA-seq expression, I found that only *ABI1* mRNA levels were significantly elevated in the presence of either CNV or SV. Together, these results appear to suggest that *ABI1* was the relevant target gene for these chromosomal variations. While the region had previously identified wherein *BMI1* was attributed as the target gene, in this analysis the expression level of *BMI1* mRNA was not associated with the presence of SVs or CNVs.

Our dataset of both genomic and expression data enabled greater exploration of genes targeted by common copy number events. In total 10 regions affected by recurrent amplifications/gains and 10 regions affected by recurrent deletions were identified here. While many of these regions have already been described in earlier

studies in MCL, these results were the first to leverage expression data on a large MCL tumour cohort. Consistent with prior studies, the most observed amplification affected 3q26.1. Owing to the extent of the minimal common region, the relevant target gene(s) of this event have not yet been definitively ascertained. Here, I attempted to determine the genes with deregulated expression, but this led to a list of 161 genes. Based solely on the genes that displayed the highest differential expression between mutation statuses, three genes: *PLCH1*, *MB21D2* and *GP5*, were the best candidates. While the issue of gene target for these amplifications remains unresolved, these results could allow for more targeted studies on a smaller list of possible genes. In other cases, my analysis highlighted potential cis regulatory targets of other recurrently amplified and gained regions.

Using our new compendium of potential drivers in MCL, I combined these and applied machine learning to search for robust genetic subdivisions. This analysis also included other possibly informative features such as aSHM/SHM regions and *CCND1* 3'UTR. A comparison of clinical outcomes (OS) across the five resulting genetic clusters displayed distinct trajectories (MC1-MC5). The genetics of each cluster were unique and align with known biomarkers of clinical outcome.

MC1 tumours were mostly devoid of common MCL alterations but were enriched for SHM mutations and consisted of nearly all nnMCLs. Importantly, a subset of cMCL tumours were also assigned to MC1 and these cMCLs had the longest OS of all tumours. Although grouped with nnMCLs, these cMCLs differed in many important ways such as lack of IGHV hypermutation and presence common MCL alterations (i.e., *ATM* mutations/deletions). However, unlike other cMCLs, cluster one cMCLs acquired mutations in SHM regions particularly in the *CCND1* promoter region. These differences may be useful in identifying patients with good clinical outcomes and can allow for the application of a wait-and-observe strategy. MC2 patients had intermediate OS and this cluster was enriched for tumours with CNVs and SVs affecting *ABI1*, aberrations involving *ATM* and *S1PR1*, and amplification of 7p22.2. As previously described, multiple genes were overexpressed because of gains and amplifications on 7p22.2 and *COX19* and *KLHL7* displayed higher expression in MCL overall compared to normal counterparts. Altogether it appears that multiple pathways may be affected by these changes which include respiration, apoptosis, DNA repair, cell-cycle control, and cytoskeleton regulation. MC3 also displayed intermediate OS but unlike cluster two, the

genetic features of this cluster are not well characterized. Specifically, the main feature of this cluster was amplifications of 3q26.1. In addition to this large region not having a clear target gene, the amplifications/gains observed were common amongst tumours in other clusters. This cluster appears to be defined by what features are absent rather than the features that occur. MC4 and MC5 consisted of the worst performing MCL patients, and the features associated with these reflect that observation. *TP53* deletions/mutations, *CDKN2A/MTAP* deletions, *RB1* deletions, *TERT* amplifications and *CCND1* 3'UTR mutations were observed in both clusters, but especially in cluster five tumours. In contrast to cluster five, cluster four tumours acquired additional CNVs. Integrative analysis of these regions has nominated some possible gene targets within these CNVs such as *FBXO32*, *MIR7874* and *BCL2L10*.

5.2. RNA-splicing dysregulation

A variety of cancers have been observed with splicing abnormalities which include intron retention, abnormal expression of isoforms, and splicing errors of tumour suppressors or oncogenes[259,260]. Perturbation of splicing can occur through a variety of mechanisms. These include alteration of key *trans*-acting RNA splicing proteins or RNAs[162,261–265], altered expression of splicing factors[266–271], and mutations of *cis*-splicing elements[224,272]. In MCL, three genes that encode RNA-binding proteins (*DAZAP1*, *EWSR1* and *HNRNPH1*) acquired recurrent SNVs, while *PRPF4* was found overexpressed due to 6q22.3 amplifications and gains. These results contribute to an emerging theme of splicing dysregulation in B-cell malignancies. As splicing factors, alteration of concentrations or functions of these genes are predicted to have widespread changes to either splicing landscape and/or expression of a variety of genes including ones that are apart of cellular processes considered hallmarks of cancer[273]. In glioblastoma, for example, high concentrations of HNRNPH/F was found to maintain DNA-damage response genes in a linear form facilitating translation of these genes, and resistant to chemo- and radiotherapy[274].

5.3. Towards a new classification scheme for MCL

Genetics-based classification schemes have been of substantial interest to the lymphoma community recently. Prior to this work only one study explored the existence of genetic subgroups in MCL. In their study, Yi *et al*[243] described a clustering solution with four candidate genetic subgroups numbered between C1 and C4 (Figure 5-1). Although aspects of their subgroups are shared with my own results, there are also some key differences. Of the similarities, MC1 and C1 were both able to capture long-lived MCL patients consisting of nnMCLs and cMCLs. In contrast with C1, our MC1 included a greater proportion of cMCL tumours and those cMCLs harboured a similar mutation repertoire to other cMCLs (i.e., *ATM* and *S1PR1* mutations/deletions). Unlike their approach, my analysis did not rely directly on IGHV mutation status for cluster assignment, instead leveraging information from additional sites affected by SHM. This difference may explain why MC1 included a mixture of cMCLs and nnMCLs. These nnMCL shared genetic features with cMCLs including SHM mutations in *CCND1* and *IGLL5*. An interesting similarity was that MC2 and MC3 in my results paralleled the C2 and C3 reported by Yi *et al*[243], respectively. Both C2 and MC2 featured ATM and S1PR1 mutations and deletions, while C3 and cluster 3 contained 3q26.1 amplifications.

Despite differences in sequencing methodology, feature selection and clinical samples, the genetic clusters from the Yi study share similarities with the five-cluster solution described herein. This supports the existence of *bona fide* distinct genetic substructure amongst MCL patients but leaves the exact identity and number of clusters unresolved. The greatest difference between their result and the clusters described here was the single C4 roughly corresponded to two clusters (MC4 and MC5) in my results. Unsurprisingly, each of C4, MC4 and MC5 represented poorly performing patients and contained high-risk variants such as *TP53* mutations/deletions, *CDKN2A/MTAP* deletions and *RB1* deletions. However, MC4 tumours were also characterized by several CNVs that were not reported in Yi *et al*[243], possibly due to a more restricted sets of features utilized by their analysis or the decision to leave similar genetic aberrations as separate features. These CNVs separate high-risk MCLs into two genetically distinct tumour populations that may rely on unique oncogenic mechanisms. The differences observed between my results and those reported by Yi et al[243] could be further explored

since both studies used tumour genetic data from Nadeu et al[45], however, cluster membership information from the Yi study was not published.

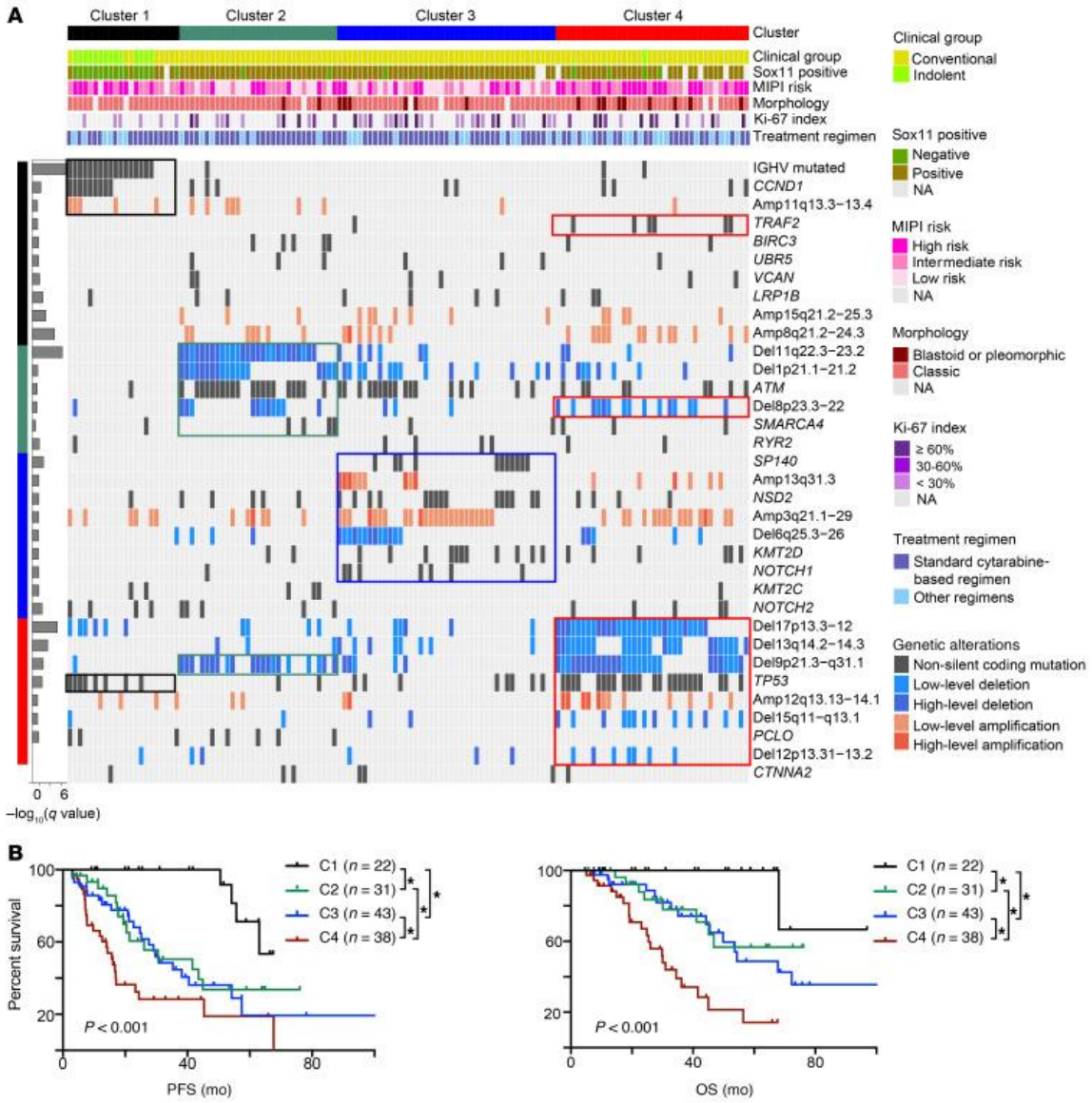

Figure 5-1 MCL NMF clustering described by Yi et al.[243]

(A) Heatmap of tumours and features of four genetic subgroups (C1 to C4). (B) PFS and (C) OS of genetic subgroups in months. Repurposed here from Yi S, Yan Y, Jin M, et al. Genomic and transcriptomic profiling reveals distinct molecular subsets associated with outcomes in mantle cell lymphoma. *J Clin Invest*. 132(3):e153283, which is distributed under the terms of Creative Commons Attribution 4.0 International License.

5.4. Future Directions

5.4.1. From Clusters to a Classification System

The genetic subgroups detailed in this thesis and similar results reported recently by Yi *et al*[243] highlight the existence of subgroupings within MCL that can be reproducibly identified based on the presence of recurrent genetic features. Some of these clusters are enriched for patients with significantly longer (MC1) or shorter (MC4, MC5) survival. Interestingly, the group of tumours with the best prognosis was marked by evidence of germinal centre involvement (IGHV mutations in Yi *et al*[243], and SHM activity in my results). Whereas in Yi *et al*[243], this group largely comprised nnMCLs, my results showed a greater representation of cMCLs within this indolent group, which may indicate the existence of a subset of cMCL with germinal centre experience and a more indolent clinical course. Mutations in *IGLL5* and 5' region of *CCND1* may be useful for detecting cMCL patients with better prognosis, which could affect clinical management decisions such as treatment timing and intensity. Whether membership in this (and the other clusters) can be adequately assigned using data from a targeted sequencing panel remains to be determined.

Translation of genetic clusters into a clinical setting will require additional work. Using an unsupervised clustering approach such as NMF has revealed organic groupings of MCL tumours based on similar genetic aberrations, however, classification of new tumours is not easily accomplished using NMF. The next step, using the genetic cluster assignments as labels, would be to train a classifier using a supervised machine learning algorithm such as random forest or neural network[239]. Development of a classifier would enable quick identification of new tumours that could be used in future clinical trials.

Risk-adapted clinical trial design may be a useful tool in the ongoing attempt to improve outcomes for MCL patients. For example, in the recent WINDOW-2 trial[275], patients were given an induction regimen of ibrutinib/rituximab/venotclax followed by either observation or consolidation therapy based on patient and tumour characteristics. Patients that displayed low Ki-67 index, low MIPI score, small tumour sizes, and lacked high-risk features including *TP53*, *NSD2*, *NOTCH1/2* mutations, complex karyotype, 17p deletions, MYC positivity, large tumour size, blastoid/pleomorphic histology or partial

response to induction therapy were assigned to the observation arm. All other patients were given high intensity immunochemotherapy. Good clinical outcomes were reached with high ORR (96%) and median OS and PFS not reached after median follow-up of 24 months. Incorporation of additional biomarkers into such trial designs, such as the genetic clusters presented herein, could provide more meaningful risk stratification within this type of clinical trial.

While targeted therapies in MCL has largely focused on inhibitors of BTK (ibrutinib[179], acalabrutinib[97], zanubrutinib[95]), some new therapeutics are currently being investigated. For example, Venetoclax, a BH3 mimetic targeting the BCL2 and the intrinsic apoptotic pathway has shown some efficacy in R/R MCL in combination with ibrutinib in phase II trials[276]. In Chapter 3, I described the recurrence of copy number gains and focal amplifications of *BCL2* (18q21.32) and *BCL2L10* (15q21.2) with BCL2L10 showing higher expression due to CNVs. *BCL2* and *BCL2L10* amplifications were features designated specifically to MC4 and were largely absent in MC5. The presence of these CNVs could indicate increased dependence of apoptotic inhibition in MC4 tumours and therefore may be amenable to inhibition by venetoclax or other BH3 mimetics.

Inhibition of the cell-cycle is hypothesized to improve clinical outcome in MCL patients because of MCL tumour reliance on CCND1. Phase I clinical trials of CDK4/6 inhibitors, either as single-agents or in combination therapy, in relapse-refractory MCL has so far yielded moderate responses with overall response rates between 18 and 67% and complete response rates between 6 and 37%[277–281]. MC4 and MC5 clusters include several aberrations related to cell-cycle deregulation, namely loss of *CDKN2A*/*MTAP*, *RB1* and *TP53*. Clinical trial designs incorporating genetic subgroups may reveal insight into the efficacy of targeted therapeutics such as CDK4/6 inhibitors including which patients do and do not benefit.

5.4.2. CRISPR screens for essential genes

With the additional recurrent driver mutations defined in this thesis, including multiple genes involved in RNA maturation, there is additional opportunity for the

identification of completely new therapeutic strategies. Chapter 3 highlights several recurrent CNVs and nominated some putative cis regulatory targets associated with these. The number of candidate genes with elevated expression associated with gains and amplification could be as low as two or up to 161, in one case (3q26.1). Additionally, there were five regions where no gene was found over- or under-expressed due to CNVs. In both situations, additional data and functional experiments will be needed to determine the biologically meaningful targets affected by these CNVs (i.e., the oncogene). One approach currently being investigated in our lab is to use proteomics from MCL tumours to identify genes in amplified regions that exhibit elevated protein levels. We should expect that genes within elevated copy regions would display greater concentration of the associated protein. Using mass-spectrometry, isoform concentrations in a large cohort of MCL tumours could be determined and associated with any relevant genetic changes including copy number events.

Another approach for identifying genes of relevance to MCL biology would utilize a targeted CRISPR screen to determine which of the CNV-affected genes are essential in MCL cell lines[282]. Using CRISPR dropout screens, this would involve introduction of a pooled library of guide RNAs (gRNAs) targeting candidate genes in these regions, which could be informed by the list of putative targets found in Chapter 3. After introduction of the CRISPR-cas9, individual cells with guides that cause mutation (loss) of essential genes would be lost, causing drop-out of those guides from the library. Such functional genomics experiments, guided by the results described herein, may allow the delineation of new oncogenes with relevance to MCL, thereby leading to new and potentially "druggable" vulnerabilities.

5.4.3. RNA-binding proteins in MCL and B-cell lymphomas

Several ongoing projects in the Morin lab have been initiated because of the findings described in Chapter 2. Although the splicing landscape was not explored in this thesis, our current data set which includes ribosomal depleted RNA-seq would enable a comprehensive survey of changes in splicing in MCL due to mutations affecting RNA-binding proteins and has been performed. Initial results showed altered splicing in specific genes when comparing tumours with and without mutations affecting RNA-

binding proteins although a pattern of splicing changes has not emerged. This may be attributed to difficulties in detecting some aberrantly spliced products that do not reside within the cell for long, or that aberrantly spliced products may exist at low concentrations compared to more common isoforms and therefore are less likely to sequenced.

Beyond splicing, RNA-binding proteins can regulate mRNA expression, transport, and translation. Multiple ongoing projects are investigating these functions in various lymphoma subtypes including MCL. For example, *HNRNPU* was frequently truncated in BL and high-grade B-cell lymphomas with *BCL2* translocations; both lymphomas rely on *MYC* driven oncogenesis. Functional characterization of *HNRNPHU* within BL, liver cancer and myelogenous leukemia cell lines showed that *HNRNPU* expression positively correlated with MYC expression suggesting a role as modulator of *MYC* (unpublished). Loss of *HNRNPU* within these lymphomas may be to reduce MYC-induced apoptosis and proliferative stress. In the Morin laboratory, there are ongoing experiments involving RNA-binding proteins to find and characterize binding targets of these proteins using targeted pull-down protocols and investigate changes in binding ability and specificity of these RNA-binding proteins due to mutations.

5.4.4. Indolent MCL

In this thesis, I had focused primarily on cMCLs and the differences within this MCL subtype. Although beyond the scope of the current analysis, it was possible to compare cMCLs to nnMCLs. While I did not perform a comparative analysis between MCL subtypes, many previous studies have focused on this topic[23,33,45,283].

From the clustering results I showed that a subset of cMCLs group together with nnMCLs in cluster MC1 and exhibit prolonged overall survival. While these cMCLs shared putative aSHM features with nnMCLs in the same cluster, these tumours also resembled non-MC1 cMCLs. A better understanding of the reason for these differences would improve our knowledge about MCL development and possibly reveal ways to intervene. Given the genetic similarity between MC1 and non-MC1 cMCLs, the cause for the differential patient outcomes observed may be due to non-genetic causes, i.e.,

epigenetics. To investigate this hypothesis, I propose that tumours of both MCL subtypes undergo ATAC-seq (assess chromatin accessibility), CHIP-seq (determine histone regulatory markers), and bisulfite-sequencing (assess chromatin methylation patterns). With this dataset, a comparative analysis can be performed examining the differential methylation and regulatory markers between MCL cell-of-origin subtypes, and between tumours assigned to different genetic clusters. Prior investigations into epigenetic landscape of MCL have focused mostly on the differences between nnMCL and cMCL subtypes[185], but with evidence of genetic differences between cMCLs described in this thesis, an expanded epigenetic survey is warranted.

References

1. Banks PM, Chan J, Cleary ML, et al. Mantle Cell Lymphoma A Proposal for Unification of Morphologic, Immunologic, and Molecular Data. *Am. J. Surg. Pathol.* 1992;16(7):637–640.
2. Lennert K, Stein H, Kaiserling E. Cytological and functional criteria for the classification of malignant lymphomata. *Br. J. Cancer*. 1975;31(SupII):29–43.
3. Berard CW, Dorfman RF. Histopathology of malignant lymphomas. *Clinics in Haematology*. 1974;3(1):39–76.
4. Weisenburger DD, Kim H, Rappaport H. Mantle-zone lymphoma: a follicular variant of intermediate lymphocytic lymphoma. *Cancer*. 1982;49(7):1429–1438.
5. Swerdlow SH, Habeshaw JA, Murray LJ, et al. Centrocytic lymphoma: a distinct clinicopathologic and immunologic entity. A multiparameter study of 18 cases at diagnosis and relapse. *Am J Pathol*. 1983;113(2):181–197.
6. Weisenburger DD, Sanger WG, Armitage JO, Purtilo DT. Intermediate Lymphocytic Lymphoma: Immunophenotypic and Cytogenetic Findings. *Blood*. 1987;69(6):1617–1621.
7. Lardelli P, Bookman MA, Sundeen J, Longo DL, Jaffe ES. Lymphocytic lymphoma of intermediate differentiation. Morphologic and immunophenotypic spectrum and clinical correlations. *Am J Surg Pathol*. 1990;14(8):752–763.
8. Cossman J, Neckers LM, Hsu S, Longo D, Jaffe ES. Low-grade lymphomas. Expression of developmentally regulated B-cell antigens. *Am J Pathol*. 1984;115(1):117–124.
9. Medeiros LJ, Van Krieken JH, Jaffe ES, Raffeld M. Association of bcl-1 rearrangements with lymphocytic lymphoma of intermediate differentiation. *Blood*. 1990;76(10):2086–2090.
10. Tsujimoto Y, Yunis J, Onorato-Showe L, et al. Molecular Cloning of the Chromosomal Breakpoint of B-Cell Lymphomas and Leukemias with the t(11;14) Chromosome Translocation. *Science*. 1984;224(4656):1403–1406.
11. Bosch F, Jares P, Campo E, et al. PRAD-1/cyclin D1 gene overexpression in chronic lymphoproliferative disorders: a highly specific marker of mantle cell lymphoma. *Blood*. 1994;84(8):2726–2732.
12. Motokura T, Bloom T, Kim HG, et al. A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature*. 1991;350(6318):512–515.
13. Zucca E, Stein H, Coiffier B, et al. European Lymphoma Task Force (ELTF): Report of the workshop on Mantle Cell Lymphoma (MCL). *Ann. Oncol.* 1994;5(6):507–511.
14. Salaverria I, Royo C, Carvajal-Cuenca A, et al. CCND2 rearrangements are the most frequent genetic events in cyclin D1− mantle cell lymphoma. *Blood*. 2013;121(8):1394–1402.
15. Fu K, Weisenburger DD, Greiner TC, et al. Cyclin D1-negative mantle cell lymphoma: a clinicopathologic study based on gene expression profiling. *Blood*. 2005;106(13):4315–4321.
16. Rosenwald A, Wright G, Wiestner A, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell*. 2003;3(2):185–197.
17. Martín-Garcia D, Navarro A, Valdés-Mas R, et al. CCND2 and CCND3 hijack immunoglobulin light-chain enhancers in cyclin D1- mantle cell lymphoma. *Blood*. 2019;133(9):940–951.

18. Gesk S, Klapper W, Martín-Subero JI, et al. A chromosomal translocation in cyclin D1-negative/cyclin D2-positive mantle cell lymphoma fuses the CCND2 gene to the IGK locus. *Blood*. 2006;108(3):1109–1110.
19. Shiller SM, Zieske A, Holmes H, et al. CD5-positive, cyclinD1-negative mantle cell lymphoma with a translocation involving the CCND2 gene and the IGL locus. *Cancer Genet*. 2011;204(3):162–164.
20. Herens C, Lambert F, Quintanilla-Martinez L, et al. Cyclin D1-negative mantle cell lymphoma with cryptic t(12;14)(p13;q32) and cyclin D2 overexpression. *Blood*. 2008;111(3):1745–1746.
21. Mozos A, Royo C, Hartmann E, et al. SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. *Haematologica*. 2009;94(11):1555–1562.
22. Meggendorfer M, Kern W, Haferlach C, Haferlach T, Schnittger S. SOX11 overexpression is a specific marker for mantle cell lymphoma and correlates with t(11;14) translocation, CCND1 expression and an adverse prognosis. *Leukemia*. 2013;27(12):2388–2391.
23. Navarro A, Clot G, Royo C, et al. Molecular subsets of mantle cell lymphoma defined by the IGHV mutational status and SOX11 expression have distinct biological and clinical features. *Cancer Res*. 2012;72(20):5307–5316.
24. Palomero J, Vegliante MC, Eguileor A, et al. SOX11 defines two different subtypes of mantle cell lymphoma through transcriptional regulation of BCL6. *Leukemia*. 2016;30(7):1596–1599.
25. Swerdlow SH, Campo E, Harris NL, et al. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: International Agency for Research on Cancer; 2017.
26. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375–2390.
27. Bernard M, Gressin R, Lefrère F, et al. Blastic variant of mantle cell lymphoma: a rare but highly aggressive subtype. *Leukemia*. 2001;15(11):1785–1791.
28. Dreyling M, Klapper W, Rule S. Blastoid and pleomorphic mantle cell lymphoma: still a diagnostic and therapeutic challenge! *Blood*. 2018;132(26):2722–2729.
29. Tiemann M, Schrader C, Klapper W, et al. Histopathology, cell proliferation indices and clinical outcome in 304 patients with mantle cell lymphoma (MCL): a clinicopathological study from the European MCL Network. *Br. J. Haematol.* 2005;131(1):29–38.
30. Chen D, Viswanatha DS, Zent CS, et al. Indolent Mantle Cell Lymphoma: A Distinct Subgroup Characterized by Leukemic Phase Disease without Lymphadenopathy. *Blood*. 2009;114(22):3937–3937.
31. Espinet B, Salaverria I, Beà S, et al. Incidence and prognostic impact of secondary cytogenetic aberrations in a series of 145 patients with mantle cell lymphoma. *Genes Chromosomes Cancer*. 2010;49(5):439–451.
32. Isaac KM, Portell CA, Williams ME. Leukemic Variant of Mantle Cell Lymphoma: Clinical Presentation and Management. *Curr Oncol Rep*. 2021;23(9):102.
33. Fernàndez V, Salamero O, Espinet B, et al. Genomic and Gene Expression Profiling Defines Indolent Forms of Mantle Cell Lymphoma. *Cancer Res*. 2010;70(4):1408–1418.
34. Orchard J, Garand R, Davis Z, et al. A subset of t(11;14) lymphoma with mantle cell features displays mutated IgVH genes and includes patients with good prognosis, nonnodal disease. *Blood*. 2003;101(12):4975–4981.
35. Royo C, Salaverria I, Hartmann EM, et al. The complex landscape of genetic alterations in mantle cell lymphoma. *Semin Cancer Biol*. 2011;21(5):322–334.

36. Chapman-Fredricks J, Sandoval-Sus J, Vega F, Lossos IS. Progressive leukemic non-nodal mantle cell lymphoma associated with deletions of TP53, ATM, and/or 13q14. *Ann Diagn Pathol*. 2014;18(4):214–219.
37. Ye H, Desai A, Zeng D, et al. Smoldering mantle cell lymphoma. *J Exp Clin Cancer Res*. 2017;36(1):185.
38. Royo C, Navarro A, Clot G, et al. Non-nodal type of mantle cell lymphoma is a specific biological and clinical subgroup of the disease. *Leukemia*. 2012;26(8):1895–1898.
39. Teras LR, DeSantis CE, Cerhan JR, et al. 2016 US lymphoid malignancy statistics by World Health Organization subtypes: 2016 US Lymphoid Malignancy Statistics by World Health Organization Subtypes. *CA Cancer J. Clin.* 2016;66(6):443–459.
40. Wang Y, Ma S. Risk factors for etiology and prognosis of mantle cell lymphoma. *Expert Rev Hematol*. 2014;7(2):233–243.
41. Wang SS, Slager SL, Brennan P, et al. Family history of hematopoietic malignancies and risk of non-Hodgkin lymphoma (NHL): a pooled analysis of 10 211 cases and 11 905 controls from the International Lymphoma Epidemiology Consortium (InterLymph). *Blood*. 2007;109(8):3479–3488.
42. Küppers R, Dalla-Favera R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. 2001;20(40):5580–5594.
43. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood*. 2008;112(5):1570–1580.
44. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*. 2020;160(3):233–247.
45. Nadeu F, Martin-Garcia D, Clot G, et al. Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood*. 2020;136(12):1419–1432.
46. Rosenberg CL, Wong E, Petty EM, et al. PRAD1, a candidate BCL1 oncogene: mapping and expression in centrocytic lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* 1991;88(21):9638–9642.
47. Arnold A, Motokura T, Bloom T, et al. PRAD1 (cyclin D1): a parathyroid neoplasia gene on 11q13. *Henry Ford Hosp Med J*. 1992;40(3–4):177–180.
48. Inaba T, Matsushime H, Valentine M, et al. Genomic organization, chromosomal localization, and independent expression of human cyclin D genes. *Genomics*. 1992;13(3):565–574.
49. Seto M, Yamamoto K, Iida S, et al. Gene rearrangement and overexpression of PRAD1 in lymphoid malignancy with t(11;14)(q13;q32) translocation. *Oncogene*. 1992;7(7):1401–1406.
50. de Boer CJ, van Krieken JH, Kluin-Nelemans HC, Kluin PM, Schuuring E. Cyclin D1 messenger RNA overexpression as a marker for mantle cell lymphoma. *Oncogene*. 1995;10(9):1833–1840.
51. Kato J, Matsushime H, Hiebert SW, Ewen ME, Sherr CJ. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev*. 1993;7(3):331–342.
52. Gladden AB, Woolery R, Aggarwal P, Wasik MA, Diehl JA. Expression of constitutively nuclear cyclin D1 in murine lymphocytes induces B-cell lymphoma. *Oncogene*. 2006;25(7):998–1007.
53. Campo E, Cymbalista F, Ghia P, et al. TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. *Haematologica*. 2018;103(12):1956–1968.

54. O'Shea D, O'Riain C, Taylor C, et al. The presence of TP53 mutation at diagnosis of follicular lymphoma identifies a high-risk group of patients with shortened time to disease progression and poorer overall survival. *Blood*. 2008;112(8):3126–3129.
55. Eskelund CW, Dahl C, Hansen JW, et al. TP53 mutations identify younger mantle cell lymphoma patients who do not benefit from intensive chemoimmunotherapy. *Blood*. 2017;130(17):1903–1910.
56. Pinyol M, Hernandez L, Cazorla M, et al. Deletions and loss of expression of p16INK4a and p21Waf1 genes are associated with aggressive variants of mantle cell lymphomas. *Blood*. 1997;89(1):272–280.
57. Hernández L, Beà S, Pinyol M, et al. CDK4 and MDM2 gene alterations mainly occur in highly proliferative and aggressive mantle cell lymphomas with wild-type INK4a/ARF locus. *Cancer Res.* 2005;65(6):2199–2206.
58. Salaverria I, Zettl A, Beà S, et al. Specific secondary genetic alterations in mantle cell lymphoma provide prognostic information independent of the gene expression-based proliferation signature. *J Clin Oncol*. 2007;25(10):1216–1222.
59. Stilgenbauer S, Schaffner C, Winkler D, et al. The ATM gene in the pathogenesis of mantle-cell lymphoma. *Ann Oncol*. 2000;11 Suppl 1:127–130.
60. Bullrich F, Rasio D, Kitada S, et al. ATM mutations in B-cell chronic lymphocytic leukemia. *Cancer Res*. 1999;59(1):24–27.
61. Staber PB, Herling M, Bellido M, et al. Consensus criteria for diagnosis, staging, and treatment response assessment of T-cell prolymphocytic leukemia. *Blood*. 2019;134(14):1132–1143.
62. Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N Engl J Med*. 1991;325(26):1831–1836.
63. Choi M, Kipps T, Kurzrock R. ATM Mutations in Cancer: Therapeutic Implications. *Mol. Cancer Ther.* 2016;15(8):1781–1791.
64. Wu C, de Miranda NF, Chen L, et al. Genetic heterogeneity in primary and relapsed mantle cell lymphomas: Impact of recurrent CARD11 mutations. *Oncotarget*. 2016;7(25):38180–38190.
65. Zhang J, Jima D, Moffitt AB, et al. The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. *Blood*. 2014;123(19):2988–2996.
66. Bea S, Valdes-Mas R, Navarro A, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* 2013;110(45):18250–18255.
67. Frazzi R. BIRC3 and BIRC5: multi-faceted inhibitors in cancer. *Cell Biosci*. 2021;11(1):8.
68. Hershkovitz-Rokah O, Pulver D, Lenz G, Shpilberg O. Ibrutinib resistance in mantle cell lymphoma: clinical, molecular and treatment aspects. *Br J Haematol*. 2018;181(3):306–319.
69. Diop F, Moia R, Favini C, et al. Biological and clinical implications of BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica*. 2020;105(2):448–456.
70. Sharpless NE. INK4a/ARF: A multifunctional tumor suppressor locus. *MUTAT RES-FUND MOL M*. 2005;576(1–2):22–38.
71. Martin P, Chadburn A, Christos P, et al. Outcome of Deferred Initial Therapy in Mantle-Cell Lymphoma. *J. Clin. Oncol.* 2009;27(8):1209–1213.
72. Abrisqueta P, Slack GW, Scott DW, et al. Outcome of Observation As Initial Strategy in Patients with Mantle Cell Lymphoma. *Blood*. 2015;126(23):2699–2699.
73. Sachanas S, Pangalis GA, Vassilakopoulos TP, et al. Combination of rituximab with chlorambucil as first line treatment in patients with mantle cell lymphoma: a highly effective regimen. *Leuk. Lymphoma*. 2011;52(3):387–393.

74. Jain AG, Chang C-C, Ahmad S, Mori S. Leukemic Non-nodal Mantle Cell Lymphoma: Diagnosis and Treatment. *Curr. Treat. Options in Oncol.* 2019;20(12):85.
75. Abrisqueta P, Scott DW, Slack GW, et al. Observation as the initial management strategy in patients with mantle cell lymphoma. *Ann. Oncol.* 2017;28(10):2489–2495.
76. Kumar A. Who is a Candidate for Watch and Wait Therapy in Mantle Cell Lymphoma? *Clin. Lymphoma Myeloma Leuk.* 2019;19:S109–S111.
77. Dreyling M, Lenz G, Hoster E, et al. Early consolidation by myeloablative radiochemotherapy followed by autologous stem cell transplantation in first remission significantly prolongs progression-free survival in mantle-cell lymphoma: results of a prospective randomized trial of the European MCL Network. *Blood*. 2005;105(7):2677–2684.
78. Romaguera JE, Fayad LE, Feng L, et al. Ten-year follow-up after intense chemoimmunotherapy with Rituximab-HyperCVAD alternating with Rituximab-high dose methotrexate/cytarabine (R-MA) and without stem cell transplantation in patients with untreated aggressive mantle cell lymphoma. *Br J Haematol*. 2010;150(2):200–208.
79. Hermine O, Hoster E, Walewski J, et al. Addition of high-dose cytarabine to immunochemotherapy before autologous stem-cell transplantation in patients aged 65 years or younger with mantle cell lymphoma (MCL Younger): a randomised, open-label, phase 3 trial of the European Mantle Cell Lymphoma Network. *Lancet*. 2016;388(10044):565–575.
80. Delarue R, Haioun C, Ribrag V, et al. CHOP and DHAP plus rituximab followed by autologous stem cell transplantation in mantle cell lymphoma: a phase 2 study from the Groupe d'Etude des Lymphomes de l'Adulte. *Blood*. 2013;121(1):48–53.
81. Geisler CH, Kolstad A, Laurell A, et al. Long-term progression-free survival of mantle cell lymphoma after intensive front-line immunochemotherapy with in vivo-purged stem cell rescue: a nonrandomized phase 2 multicenter study by the Nordic Lymphoma Group. *Blood*. 2008;112(7):2687–2693.
82. Maddocks K. Update on Mantle Cell Lymphoma. *Blood*. 2018;blood-2018-03-791392.
83. Kluin-Nelemans HC, Hoster E, Hermine O, et al. Treatment of older patients with mantle-cell lymphoma. *N Engl J Med*. 2012;367(6):520–531.
84. Rummel MJ, Niederle N, Maschmeyer G, et al. Bendamustine plus rituximab versus CHOP plus rituximab as first-line treatment for patients with indolent and mantle-cell lymphomas: an open-label, multicentre, randomised, phase 3 non-inferiority trial. *Lancet*. 2013;381(9873):1203–1210.
85. Flinn IW, van der Jagt R, Kahl BS, et al. Randomized trial of bendamustine-rituximab or R-CHOP/R-CVP in first-line treatment of indolent NHL or MCL: the BRIGHT study. *Blood*. 2014;123(19):2944–2952.
86. Villa D, Sehn LH, Savage KJ, et al. Bendamustine and rituximab as induction therapy in both transplant-eligible and -ineligible patients with mantle cell lymphoma. *Blood Adv*. 2020;4(15):3486–3494.
87. Robak T, Huang H, Jin J, et al. Bortezomib-based therapy for newly diagnosed mantle-cell lymphoma. *N Engl J Med*. 2015;372(10):944–953.
88. Castellino A, Wang Y, Larson MC, et al. Evolving frontline immunochemotherapy for mantle cell lymphoma and the impact on survival outcomes. *Blood Adv*. 2022;6(4):1350–1360.
89. Goy A, Bernstein SH, Kahl BS, et al. Bortezomib in patients with relapsed or refractory mantle cell lymphoma: updated time-to-event analyses of the multicenter phase 2 PINNACLE study. *Ann Oncol*. 2009;20(3):520–525.

90. Ansell SM, Tang H, Kurtin PJ, et al. Temsirolimus and rituximab in patients with relapsed or refractory mantle cell lymphoma: a phase 2 study. *Lancet Oncol.* 2011;12(4):361–368.
91. Kahl BS, Spurgeon SE, Furman RR, et al. A phase 1 study of the PI3Kδ inhibitor idelalisib in patients with relapsed/refractory mantle cell lymphoma (MCL). *Blood*. 2014;123(22):3398–3405.
92. Wang ML, Rule S, Martin P, et al. Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. *N Engl J Med*. 2013;369(6):507–516.
93. Dreyling M, Jurczak W, Jerkeman M, et al. Ibrutinib versus temsirolimus in patients with relapsed or refractory mantle-cell lymphoma: an international, randomised, open-label, phase 3 study. *Lancet*. 2016;387(10020):770–778.
94. Rule S, Jurczak W, Jerkeman M, et al. Ibrutinib versus temsirolimus: 3-year follow-up of patients with previously treated mantle cell lymphoma from the phase 3, international, randomized, open-label RAY study. *Leukemia*. 2018;32(8):1799–1803.
95. Tam CS, Opat S, Simpson D, et al. Zanubrutinib for the treatment of relapsed or refractory mantle cell lymphoma. *Blood Adv.* 2021;5(12):2577–2585.
96. Song Y, Zhou K, Zou D, et al. Zanubrutinib in relapsed/refractory mantle cell lymphoma: long-term efficacy and safety results from a phase 2 study. *Blood*. 2022;139(21):3148–3158.
97. Wang M, Rule S, Zinzani PL, et al. Acalabrutinib in relapsed or refractory mantle cell lymphoma (ACE-LY-004): a single-arm, multicentre, phase 2 trial. *Lancet*. 2018;391(10121):659–667.
98. Eyre TA, Cheah CY, Wang ML. Therapeutic options for relapsed/refractory mantle cell lymphoma. *Blood*. 2022;139(5):666–677.
99. Andersen NS, Jensen MK, De Nully Brown P, Geisler CH. A Danish population-based analysis of 105 mantle cell lymphoma patients. *Eur. J. Cancer*. 2002;38(3):401–408.
100. Velders GA, Kluin-Nelemans JC, De Boer CJ, et al. Mantle-cell lymphoma: a population-based clinical study. *J. Clin. Oncol.* 1996;14(4):1269–1274.
101. Argatoff LH, Connors JM, Klasa RJ, Horsman DE, Gascoyne RD. Mantle Cell Lymphoma: A Clinicopathologic Study of 80 Cases. *Blood*. 1997;89(6):2067–2078.
102. Räty R, Franssila K, Joensuu H, Teerenhovi L, Elonen E. Ki-67 expression level, histological subtype, and the International Prognostic Index as outcome predictors in mantle cell lymphoma: Mantle cell lymphoma. *Eur. J. Haematol.* 2002;69(1):11–20.
103. Hoster E, Dreyling M, Klapper W, et al. A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood*. 2008;111(2):558–565.
104. Geisler CH, Kolstad A, Laurell A, et al. The Mantle Cell Lymphoma International Prognostic Index (MIPI) is superior to the International Prognostic Index (IPI) in predicting survival following intensive first-line immunochemotherapy and autologous stem cell transplantation (ASCT). *Blood*. 2010;115(8):1530–1533.
105. Schrader C, Meusers P, Brittinger G, et al. Topoisomerase IIα expression in mantle cell lymphoma: a marker of cell proliferation and a prognostic factor for clinical outcome. *Leukemia*. 2004;18(7):1200–1206.
106. Determann O, Hoster E, Ott G, et al. Ki-67 predicts outcome in advanced-stage mantle cell lymphoma patients treated with anti-CD20 immunochemotherapy: results from randomized trials of the European MCL Network and the German Low Grade Lymphoma Study Group. *Blood*. 2008;111(4):2385–2387.
107. Halldórsdóttir AM, Lundin A, Murray F, et al. Impact of *TP53* mutation and 17p deletion in mantle cell lymphoma. *Leukemia*. 2011;25(12):1904–1908.
108. Greiner TC, Dasgupta C, Ho VV, et al. Mutation and genomic deletion status of *ataxia telangiectasia mutated* (*ATM*) and *p53* confer specific gene expression

profiles in mantle cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* 2006;103(7):2352–2357.
109. Nordström L, Sernbo S, Eden P, et al. 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*. 2014;166(1):98–108.
110. Aukema SM, Hoster E, Rosenwald A, et al. Expression of TP53 is associated with the outcome of MCL independent of MIPI and Ki-67 in trials of the European MCL Network. *Blood*. 2018;131(4):417–420.
111. Rubio-Moscardo F, Climent J, Siebert R, et al. Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood*. 2005;105(11):4445–4454.
112. Streich L, Sukhanova M, Lu X, et al. Aggressive morphologic variants of mantle cell lymphoma characterized with high genomic instability showing frequent chromothripsis, CDKN2A /B loss, and TP53 mutations: A multi-institutional study. *Genes Chromosomes Cancer*. 2020;59(8):484–494.
113. Delfau-Larue M-H, Klapper W, Berger F, et al. High-dose cytarabine does not overcome the adverse prognostic value of CDKN2A and TP53 deletions in mantle cell lymphoma. *Blood*. 2015;126(5):604–611.
114. Lebwohl DE, Muise-Helmericks R, Sepp-Lorenzino L, et al. A truncated cyclin D1 gene encodes a stable mRNA in a human breast cancer cell line. *Oncogene*. 1994;9(7):1925–1929.
115. Rimokh R, Berger F, Bastard C, et al. Rearrangement of CCND1 (BCL1/PRAD1) 3' untranslated region in mantle-cell lymphomas and t(11q13)-associated leukemias. *Blood*. 1994;83(12):3689–3696.
116. Scott DW, Abrisqueta P, Wright GW, et al. New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies. *J. Clin. Oncol.* 2017;35(15):1668–1677.
117. Meissner B, Kridel R, Lim RS, et al. The E3 ubiquitin ligase UBR5 is recurrently mutated in mantle cell lymphoma. *Blood*. 2013;121(16):3161–3164.
118. Ahmed M, Zhang L, Nomie K, Lam L, Wang M. Gene mutations and actionable genetic lesions in mantle cell lymphoma. *Oncotarget*. 2016;7(36):58638–58648.
119. Smith RWP, Anderson RC, Smith JWS, et al. DAZAP1, an RNA-binding protein required for development and spermatogenesis, can regulate mRNA translation. *RNA*. 2011;17(7):1282–1295.
120. Paronetto MP, Miñana B, Valcárcel J. The Ewing Sarcoma Protein Regulates DNA Damage-Induced Alternative Splicing. *Molecular Cell*. 2011;43(3):353–368.
121. Mohanty A, Sandoval N, Das M, et al. CCND1 mutations increase protein stability and promote ibrutinib resistance in mantle cell lymphoma. *Oncotarget*. 2016;7(45):73558–73572.
122. Kridel R, Meissner B, Rogic S, et al. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood*. 2012;119(9):1963–1971.
123. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* 2011;208(7):1389–1401.
124. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298–303.
125. Lin Y-T, Yen PH. A novel nucleocytoplasmic shuttling sequence of DAZAP1, a testis-abundant RNA-binding protein. *RNA*. 2006;12(8):1486–1493.

126. Dai T, Vera Y, Salido EC, Yen PH. Characterization of the mouse Dazap1 gene encoding an RNA-binding protein that interacts with infertility factors DAZ and DAZL. *BMC Genomics*. 2001;2:6.
127. Choudhury R, Roy SG, Tsai YS, et al. The splicing activator DAZAP1 integrates splicing control into MEK/Erk-regulated cell proliferation and migration. *Nat. Commun.* 2014;5:3078.
128. Uren PJ, Bahrami-Samani E, de Araujo PR, et al. High-throughput analyses of hnRNP H1 dissects its multi-functional aspect. *RNA Biol*. 2016;13(4):400–411.
129. Agarwal R, Chan Y-C, Tam CS, et al. Dynamic molecular monitoring reveals that SWI–SNF mutations mediate resistance to ibrutinib plus venetoclax in mantle cell lymphoma. *Nat. Med.* 2019;25(1):119.
130. Khodadoust MS, Olsson N, Wagar LE, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature*. 2017;543(7647):723–727.
131. Grande BM, Gerhard DS, Jiang A, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood*. 2019;133(12):1313–1324.
132. Arthur SE, Jiang A, Grande BM, et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. *Nat Commun*. 2018;9(1):4001.
133. Martinez FJ, Pratt GA, Van Nostrand EL, et al. Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron*. 2016;92(4):780–795.
134. Rossbach O, Hung L-H, Schreiner S, et al. Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol Cell Biol*. 2009;29(6):1442–1451.
135. Ferrero S, Rossi D, Rinaldi A, et al. KMT2D mutations and TP53 disruptions are poor prognostic biomarkers in mantle cell lymphoma receiving high-dose therapy: a FIL study. *Haematologica*. 2019;
136. Änkö M-L, Müller-McNicoll M, Brandl H, et al. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* 2012;13(3):R17.
137. Saltzman AL, Pan Q, Blencowe BJ. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* 2011;25(4):373–384.
138. Pereverzev AP, Gurskaya NG, Ermakova GV, et al. Method for quantitative analysis of nonsense-mediated mRNA decay at the single cell level. *Sci Rep*. 2015;5:7729.
139. Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat. Biotechnol.* 2001;19(5):434–439.
140. Yang P, Zhang W, Wang J, et al. Genomic landscape and prognostic analysis of mantle cell lymphoma. *Cancer Gene Ther.* 2018;25(5–6):129–140.
141. Grünewald TGP, Cidre-Aranaz F, Surdez D, et al. Ewing sarcoma. *Nat Rev Dis Primers*. 2018;4(1):5.
142. Sanchez G, Delattre O, Auboeuf D, Dutertre M. Coupled alteration of transcription and splicing by a single oncogene: Boosting the effect on cyclin D1 activity. *Cell Cycle*. 2008;7(15):2299–2305.
143. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med*. 2018;378(15):1396–1407.
144. Chen H-Y, Yu Y-H, Yen PH. DAZAP1 regulates the splicing of Crem, Crisp2 and Pot1a transcripts. *Nucleic Acids Res*. 2013;41(21):9858–9869.
145. Honoré B, Baandrup U, Vorum H. Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. *Exp. Cell Res.* 2004;294(1):199–209.

146. Rauch J, O'Neill E, Mack B, et al. Heterogeneous nuclear ribonucleoprotein H blocks MST2-mediated apoptosis in cancer cells by regulating A-Raf transcription. *Cancer Res.* 2010;70(4):1679–1688.
147. Sun Y-L, Liu F, Liu F, Zhao X-H. Protein and gene expression characteristics of heterogeneous nuclear ribonucleoprotein H1 in esophageal squamous cell carcinoma. *World J Gastroenterol*. 2016;22(32):7322–7331.
148. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* 2016;135(8):851–867.
149. Wang E, Aslanzadeh V, Papa F, et al. Global profiling of alternative splicing events and gene expression regulated by hnRNPH/F. *PLoS ONE*. 2012;7(12):e51266.
150. Ankö M-L, Morales L, Henry I, Beyer A, Neugebauer KM. Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.* 2010;17(8):962–970.
151. Pervouchine D, Popov Y, Berry A, et al. Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic Acids Res.* 2019;47(10):5293–5306.
152. Jumaa H, Nielsen PJ. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J*. 1997;16(16):5077–5085.
153. Cannell IG, Merrick KA, Morandell S, et al. A Pleiotropic RNA-Binding Protein Controls Distinct Cell Cycle Checkpoints to Drive Resistance of p53-defective Tumors to Chemotherapy. *Cancer Cell*. 2015;28(5):623–637.
154. Díaz-Muñoz MD, Turner M. Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. *Front. Immunol.* 2018;9:.
155. Saha S, Murmu KC, Biswas M, et al. Transcriptomic Analysis Identifies RNA Binding Proteins as Putative Regulators of Myelopoiesis and Leukemia. *Front. Oncol.* 2019;9:.
156. Golan-Gerstl R, Cohen M, Shilo A, et al. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* 2011;71(13):4464–4472.
157. Xu C, Xie N, Su Y, et al. HnRNP F/H associate with hTERC and telomerase holoenzyme to modulate telomerase function and promote cell proliferation. *Cell Death Differ*. 2020;27(6):1998–2013.
158. LeFave CV, Squatrito M, Vorlova S, et al. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J*. 2011;30(19):4084–4097.
159. Hong S. RNA Binding Protein as an Emerging Therapeutic Target for Cancer Prevention and Treatment. *J Cancer Prev*. 2017;22(4):203–210.
160. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* 2011;44(1):47–52.
161. Rossi D, Bruscaggin A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 2011;118(26):6904–6908.
162. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2011;365(26):2497–2506.
163. Wang L, Brooks AN, Fan J, et al. Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell*. 2016;30(5):750–763.
164. Hacken E ten, Valentin R, Regis FFD, et al. Splicing modulation sensitizes chronic lymphocytic leukemia cells to venetoclax by remodeling mitochondrial apoptotic dependencies. *JCI Insight*. 2018;3(19):.

165. Te Raa GD, Derks I a. M, Navrkalova V, et al. The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia*. 2015;29(5):1133–1142.
166. Óskarsdóttir A, Másson G, Melsted P. BamHash: a checksum program for verifying the integrity of sequence data. *Bioinformatics*. 2016;32(1):140–141.
167. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;arXiv:1303.3997:
168. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 2015;25(6):918–925.
169. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1-33.
170. Saunders CT, Wong WSW, Swamy S, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811–1817.
171. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
172. Gonzalez-Perez A, Lopez-Bigas N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* 2012;40(21):e169.
173. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016;17(1):128.
174. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013;29(18):2238–2244.
175. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–218.
176. Mansouri L, Noerenberg D, Young E, et al. Frequent NFKBIE deletions are associated with poor outcome in primary mediastinal B-cell lymphoma. *Blood*. 2016;128(23):2666–2670.
177. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019;531210.
178. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods*. 2018;15(8):591–594.
179. Rule S, Dreyling M, Goy A, et al. Ibrutinib for the treatment of relapsed/refractory mantle cell lymphoma: extended 3.5-year follow up from a pooled analysis. *Haematologica*. 2019;104(5):e211–e214.
180. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
181. Uren PJ, Bahrami-Samani E, Burns SC, et al. Site identification in high-throughput RNA-protein interaction data. *Bioinformatics*. 2012;28(23):3013–3020.
182. Lecluse Y, Lebailly P, Roulland S, et al. t(11;14)-positive clones can persist over a long period of time in the peripheral blood of healthy individuals. *Leukemia*. 2009;23(6):1190–1193.
183. Jardin F, Picquenot J-M, Parmentier F, et al. Detection of gene copy number aberrations in mantle cell lymphoma by a single quantitative multiplex PCR assay: clinicopathological relevance and prognosis value. *Br J Haematol*. 2009;146(6):607–618.

184. Flordal Thelander E, Ichimura K, Collins VP, et al. Detailed assessment of copy number alterations revealing homozygous deletions in 1p and 13q in mantle cell lymphoma. *Leuk. Res.* 2007;31(9):1219–1230.
185. Queirós AC, Beekman R, Vilarrasa-Blasi R, et al. Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. *Cancer Cell*. 2016;30(5):806–821.
186. Thomas N, Dreval K, Gerhard DS, et al. Genetic subgroups inform on pathobiology in adult and pediatric Burkitt lymphoma. *Blood*. 2023;141(8):904–916.
187. Peterson JF, Baughn LB, Ketterling RP, et al. Characterization of a cryptic IGH/CCND1 rearrangement in a case of mantle cell lymphoma with negative CCND1 FISH studies. *Blood Adv*. 2019;3(8):1298–1302.
188. Macoska JA, Xu J, Ziemnicka D, et al. Loss of expression of human spectrin src homology domain binding protein 1 is associated with 10p loss in human prostatic adenocarcinoma. *Neoplasia*. 2001;3(2):99–104.
189. Wang C, Navab R, Iakovlev V, et al. Abelson interactor protein-1 positively regulates breast cancer cell proliferation, migration, and invasion. *Mol Cancer Res*. 2007;5(10):1031–1039.
190. Qi Y, Liu J, Chao J, et al. PTEN suppresses epithelial–mesenchymal transition and cancer stem cell activity by downregulating Abi1. *Sci Rep*. 2020;10(1):12685.
191. Chen H, Wu X, Pan ZK, Huang S. Integrity of SOS1/EPS8/ABI1 Tri-Complex Determines Ovarian Cancer Metastasis. *Cancer Res.* 2010;70(23):9979–9990.
192. Fan PD, Goff SP. Abl interactor 1 binds to sos and inhibits epidermal growth factor- and v-Abl-induced activation of extracellular signal-regulated kinases. *Mol Cell Biol*. 2000;20(20):7591–7601.
193. Li Y, Clough N, Sun X, et al. Bcr-Abl induces abnormal cytoskeleton remodeling, beta1 integrin clustering and increased cell adhesion to fibronectin through the Abl interactor 1 pathway. *J Cell Sci*. 2007;120(Pt 8):1436–1446.
194. Tagawa H, Karnan S, Suzuki R, et al. Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene*. 2005;24(8):1348–1358.
195. Schraders M, Pfundt R, Straatman HMP, et al. Novel chromosomal imbalances in mantle cell lymphoma detected by genome-wide array-based comparative genomic hybridization. *Blood*. 2005;105(4):1686–1693.
196. Rinaldi A, Kwee I, Taborelli M, et al. Genomic and expression profiling identifies the B-cell associated tyrosine kinase Syk as a possible therapeutic target in mantle cell lymphoma. *Br J Haematol*. 2006;132(3):303–316.
197. Kohlhammer H, Schwaenen C, Wessendorf S, et al. Genomic DNA-chip hybridization in t(11;14)-positive mantle cell lymphomas shows a high frequency of aberrations and allows a refined characterization of consensus regions. *Blood*. 2004;104(3):795–801.
198. Bentz M, Plesch A, Bullinger L, et al. t(11;14)-positive mantle cell lymphomas exhibit complex karyotypes and share similarities with B-cell chronic lymphocytic leukemia. *Genes Chromosom. Cancer*. 2000;27(3):285–294.
199. Beà S, Ribas M, Hernández JM, et al. Increased Number of Chromosomal Imbalances and High-Level DNA Amplifications in Mantle Cell Lymphoma Are Associated With Blastoid Variants. *Blood*. 1999;93(12):4365–4374.
200. Allen JE, Hough RE, Goepel JR, et al. Identification of novel regions of amplification and deletion within mantle cell lymphoma DNA by comparative genomic hybridization. *Br J Haematol*. 2002;116(2):291–298.

201. Wiestner A, Tehrani M, Chiorazzi M, et al. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood*. 2007;109(11):4599–4606.
202. Chen RW, Bemis L, Amato C, et al. Truncations in CCND1 mRNA 3′ UTR Alters microRNA Regulation in Mantle Cell Lymphoma. *Blood*. 2007;110(11):3184.
203. Marjon K, Cameron MJ, Quang P, et al. MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. *Cell Rep*. 2016;15(3):574–587.
204. Kryukov GV, Wilson FH, Ruth JR, et al. *MTAP* deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science*. 2016;351(6278):1214–1218.
205. Mavrakis KJ, McDonald ER, Schlabach MR, et al. Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. *Science*. 2016;351(6278):1208–1213.
206. Kalev P, Hyer ML, Gross S, et al. MAT2A Inhibition Blocks the Growth of MTAP-Deleted Cancer Cells by Reducing PRMT5-Dependent mRNA Splicing and Inducing DNA Damage. *Cancer Cell*. 2021;39(2):209-224.e11.
207. Sloan SL, Brown F, Long ME, et al. PRMT5 supports multiple oncogenic pathways in mantle cell lymphoma. *Blood*. 2023;blood.2022019419.
208. Skinnider B, Horsman D, Dupuis B, Gascoyne R. Bcl-6 and Bcl-2 protein expression in diffuse large B-cell lymphoma and follicular lymphoma: correlation with 3q27 and 18q21 chromosomal abnormalities. *Hum Pathol*. 30(7):803–808.
209. Huang S, Nong L, Wang W, et al. Prognostic impact of diffuse large B-cell lymphoma with extra copies of MYC, BCL2 and/or BCL6: comparison with double/triple hit lymphoma and double expressor lymphoma. *Diagnostic Pathol.* 2019;14(1):81.
210. Karube K, Ying G, Tagawa H, et al. BCL6 gene amplification/3q27 gain is associated with unique clinicopathological characteristics among follicular lymphoma without BCL2 gene translocation. *Mod Pathol*. 2008;21(8):973–978.
211. Ayadi L, Callebaut I, Saguez C, et al. Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. *J. Mol. Biol.* 1998;284(3):673–687.
212. Lauber J, Plessel G, Prehn S, et al. The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA*. 1997;3(8):926–941.
213. Dellaire G, Makarov EM, Cowger JeffJM, et al. Mammalian PRP4 Kinase Copurifies and Interacts with Components of Both the U5 snRNP and the N-CoR Deacetylase Complexes. *Mol Cell Biol*. 2002;22(14):5141–5156.
214. Park S, Han S-H, Kim H-G, et al. PRPF4 is a novel therapeutic target for the treatment of breast cancer by influencing growth, migration, invasion, and apoptosis of breast cancer cells via p38 MAPK signaling pathway. *Mol Cell Probes*. 2019;47:101440.
215. Gao Q, Mechin I, Kothari N, et al. Evaluation of Cancer Dependence and Druggability of PRP4 Kinase Using Cellular, Biochemical, and Structural Approaches. *J. Biol. Chem.* 2013;288(42):30125–30138.
216. Guo Q, Zhang H, Zhang L, et al. MicroRNA-21 regulates non-small cell lung cancer cell proliferation by affecting cell apoptosis via COX-19. *Int J Clin Exp Med*. 2015;8(6):8835–8841.
217. Kigoshi Y, Tsuruta F, Chiba T. Ubiquitin ligase activity of Cul3-KLHL7 protein is attenuated by autosomal dominant retinitis pigmentosa causative mutation. *J. Biol. Chem.* 2011;286(38):33613–33621.

218. Bredholt G, Storstein A, Haugen M, et al. Detection of autoantibodies to the BTB-kelch protein KLHL7 in cancer sera. *Scand J Immunol*. 2006;64(3):325–335.
219. Kurozumi S, Joseph C, Sonbul S, et al. Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis. *Breast Cancer Res Treat*. 2018;170(3):525–533.
220. Mei Z, Zhang D, Hu B, et al. FBXO32 Targets c-Myc for Proteasomal Degradation and Inhibits c-Myc Activity. *J. Biol. Chem.* 2015;290(26):16202–16214.
221. Ke N, Godzik A, Reed JC. Bcl-B, a novel Bcl-2 family member that differentially binds and regulates Bax and Bak. *J. Biol. Chem.* 2001;276(16):12481–12484.
222. Klanova M, Klener P. BCL-2 Proteins in Pathogenesis and Therapy of B-Cell Non-Hodgkin Lymphomas. *Cancers (Basel)*. 2020;12(4):938.
223. Hofmann WK, de Vos S, Tsukasaki K, et al. Altered apoptosis pathways in mantle cell lymphoma detected by oligonucleotide microarray. *Blood*. 2001;98(3):787–794.
224. Pararajalingam P, Coyle KM, Arthur SE, et al. Coding and noncoding drivers of mantle cell lymphoma identified through exome and genome sequencing. *Blood*. 2020;136(5):572–584.
225. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994–1007.
226. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. 2012;28(3):423–425.
227. Moreno-Cabrera JM, Del Valle J, Castellanos E, et al. CNVfilteR: an R/Bioconductor package to identify false positives produced by germline NGS CNV detection tools. *Bioinformatics*. 2021;37(22):4227–4229.
228. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12:R41.
229. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.
230. Kumar R, Nagpal G, Kumar V, et al. HumCFS: a database of fragile sites in human chromosomes. *BMC Genom.* 2019;19(Suppl 9):985.
231. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–330.
232. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Meth*. 2017;14(4):417–419.
233. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000research*. 2015;4:1521.
234. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):.
235. Wang ML, Jurczak W, Jerkeman M, et al. Ibrutinib plus Bendamustine and Rituximab in Untreated Mantle-Cell Lymphoma. *N. Engl. J. Med.* 2022;386(26):2482–2494.
236. Ondrejka SL, Lai R, Smith SD, Hsi ED. Indolent mantle cell leukemia: a clinicopathological variant characterized by isolated lymphocytosis, interstitial bone marrow involvement, kappa light chain restriction, and good prognosis. *Haematologica*. 2011;96(8):1121–1127.
237. Morin RD, Arthur SE, Hodson DJ. Molecular profiling in diffuse large B-cell lymphoma: why so many types of subtypes? *Br J Haematol*. 2022;196(4):814–829.

238. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24(5):679–690.
239. Wright GW, Huang DW, Phelan JD, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell*. 2020;37(4):551-568.e14.
240. Phelan JD, Young RM, Webster DE, et al. A Multiprotein Supercomplex Controlling Oncogenic Signaling in Lymphoma. *Nature*. 2018;560(7718):387–391.
241. Younes A, Sehn LH, Johnson P, et al. Randomized Phase III Trial of Ibrutinib and Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Non-Germinal Center B-Cell Diffuse Large B-Cell Lymphoma. *J Clin Oncol*. 2019;37(15):1285–1295.
242. Wilson WH, Wright GW, Huang DW, et al. Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL. *Cancer Cell*. 2021;39(12):1643-1653.e3.
243. Yi S, Yan Y, Jin M, et al. Genomic and transcriptomic profiling reveals distinct molecular subsets associated with outcomes in mantle cell lymphoma. *J Clin Invest*. 132(3):e153283.
244. Bal E, Kumar R, Hadigol M, et al. Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. *Nature*. 2022;607(7920):808–815.
245. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods*. 2018;15(8):591.
246. Wilm A, Aw PPK, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012;40(22):11189–11201.
247. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 2013;31(3):nbt.2514.
248. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
249. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94–101.
250. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380–381.
251. Bolotin DA, Poslavsky S, Davydov AN, et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol*. 2017;35(10):908–911.
252. Nadeu F, Mas-de-les-Valls R, Navarro A, et al. IgCaller for reconstructing immunoglobulin gene rearrangements and oncogenic translocations from whole-genome sequencing in lymphoid neoplasms. *Nat Commun*. 2020;11(1):3390.
253. Lee J, Nguyen PT, Shim H-S, et al. EWSR1, a multifunctional protein, regulates cellular function and aging via genetic and epigenetic pathways. *Biochim Biophys Acta Mol Basis Dis*. 2019;1865(7):1938–1945.
254. Wang Y, Feswick A, Apostolou V, et al. Gammaherpesvirus-mediated repression reveals EWSR1 to be a negative regulator of B cell responses. *Proc. Natl. Acad. Sci. U.S.A.* 2022;119(32):e2123362119.
255. Kurihara Y, Watanabe H, Kawaguchi A, et al. Dynamic changes in intranuclear and subcellular localizations of mouse Prrp/DAZAP1 during spermatogenesis: the necessity of the C-terminal proline-rich region for nuclear import and localization. *Arch. Histol. Cytol.* 2004;67(4):325–333.

256. Wang Y, Ma M, Xiao X, Wang Z. Intronic Splicing Enhancers, Cognate Splicing Factors and Context Dependent Regulation Rules. *Nat Struct Mol Biol*. 2012;19(10):1044–1052.
257. Zhao WM, Jiang C, Kroll TT, Huber PW. A proline-rich protein binds to the localization element of Xenopus Vg1 mRNA and to ligands involved in actin polymerization. *EMBO J*. 2001;20(9):2315–2325.
258. Smith RWP, Anderson RC, Smith JWS, et al. DAZAP1, an RNA-binding protein required for development and spermatogenesis, can regulate mRNA translation. *RNA*. 2011;17(7):1282–1295.
259. de Leeuw RJ, Davies JJ, Rosenwald A, et al. Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes. *Hum. Mol. Genet.* 2004;13(17):1827–1837.
260. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer*. 2016;16(7):413–430.
261. Shuai S, Suzuki H, Diaz-Navarro A, et al. The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature*. 2019;574(7780):712–716.
262. Harbour JW, Roberson EDO, Anbunathan H, et al. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat Genet*. 2013;45(2):133–135.
263. Yoshimi A, Lin K-T, Wiseman DH, et al. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature*. 2019;574(7777):273–277.
264. Brooks AN, Choi PS, De Waal L, et al. A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events. *PLoS ONE*. 2014;9(1):e87361.
265. Ibrahimpasic T, Xu B, Landa I, et al. Genomic Alterations in Fatal Forms of Non-Anaplastic Thyroid Cancer: Identification of MED12 and RBM10 as Novel Thyroid Cancer Genes Associated with Tumor Virulence. *Clin Cancer Res*. 2017;23(19):5970–5980.
266. Anczuków O, Akerman M, Cléry A, et al. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol Cell*. 2015;60(1):105–117.
267. Karni R, de Stanchina E, Lowe SW, et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol*. 2007;14(3):185–193.
268. Park S, Brugiolo M, Akerman M, et al. Differential Functions of Splicing Factors in Mammary Transformation and Breast Cancer Metastasis. *Cell Rep*. 2019;29(9):2672-2688.e7.
269. Siculella L, Giannotti L, Di Chiara Stanca B, et al. A comprehensive understanding of hnRNP A1 role in cancer: new perspectives on binding with noncoding RNA. *Cancer Gene Ther*. 2022;
270. Han B, Liu Z, Hu X, Ling H. HNRNPU promotes the progression of triple-negative breast cancer via RNA transcription and alternative splicing mechanisms. *Cell Death Dis*. 2022;13(11):940.
271. Gallardo M, Lee HJ, Zhang X, et al. hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates Proliferation and Differentiation Programs in Hematologic Malignancies. *Cancer Cell*. 2015;28(4):486–499.
272. Zhang Y, Qian J, Gu C, Yang Y. Alternative splicing and cancer: a systematic review. *Sig Transduct Target Ther*. 2021;6(1):78.
273. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 2011;144(5):646–674.
274. Herviou P, Le Bras M, Dumas L, et al. hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nat Commun*. 2020;11(1):2661.

275. Wang M, Jain P, Lee HJ, et al. Ibrutinib Plus Rituximab and Venetoclax (IRV) Followed By Risk-Stratified Observation or Short Course R-Hypercvad/MTX in Young Patients with Previously Untreated Mantle Cell Lymphoma - Phase-II Window-2 Clinical Trial. *Blood*. 2021;138(Supplement 1):3525–3525.
276. Tam CS, Anderson MA, Pott C, et al. Ibrutinib plus Venetoclax for the Treatment of Mantle-Cell Lymphoma. *N. Engl. J. Med.* 2018;378(13):1211–1223.
277. Leonard JP, LaCasce AS, Smith MR, et al. Selective CDK4/6 inhibition with tumor responses by PD0332991 in patients with mantle cell lymphoma. *Blood*. 2012;119(20):4597–4607.
278. Martin P, Ruan J, Furman R, et al. A phase I trial of palbociclib plus bortezomib in previously treated mantle cell lymphoma. *Leuk. Lymphoma*. 2019;60(12):2917–2921.
279. Martin P, Bartlett NL, Blum KA, et al. A phase 1 trial of ibrutinib plus palbociclib in previously treated mantle cell lymphoma. *Blood*. 2019;133(11):1201–1204.
280. Morschhauser F, Bouabdallah K, Stilgenbauer S, et al. Clinical Activity of Abemaciclib (LY2835219), a Cell Cycle Inhibitor Selective for CDK4 and CDK6, in Patients with Relapsed or Refractory Mantle Cell Lymphoma. *Blood*. 2014;124(21):3067–3067.
281. Seftel MD, Kuruvilla J, Kouroukis T, et al. The CDK inhibitor AT7519M in patients with relapsed or refractory chronic lymphocytic leukemia (CLL) and mantle cell lymphoma. A Phase II study of the Canadian Cancer Trials Group. *Leuk. Lymphoma*. 2017;58(6):1358–1365.
282. Bock C, Datlinger P, Chardon F, et al. High-content CRISPR screening. *Nat Rev Methods Primers*. 2022;2(1):8.
283. Camacho FI, Algara P, Rodríguez A, et al. Molecular heterogeneity in MCL defined by the use of specificV H genes and the frequency of somatic mutations. *Blood*. 2003;101(10):4042–4046.

Appendix A.

Chapter 2 Supplemental Figures

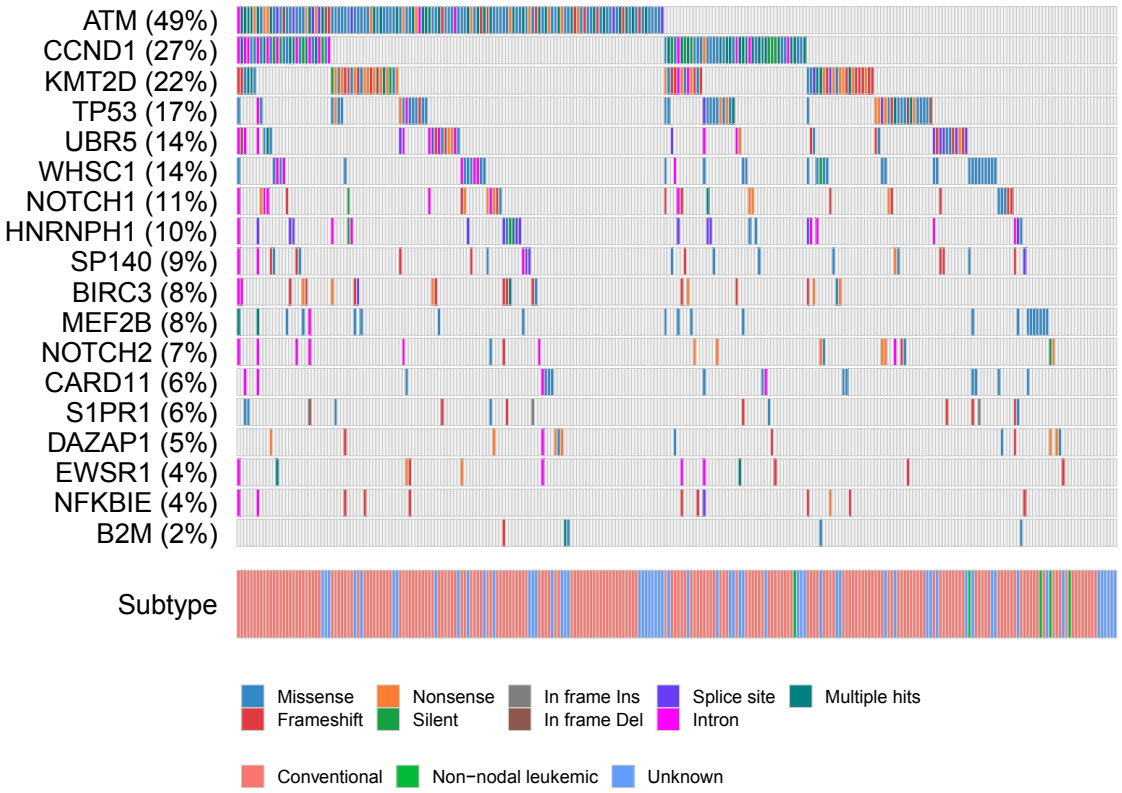

Figure A.1 Recurrent mutations are identified in MCL exomes
Non-synonymous, silent, intronic and splicing mutations observed across MCL tumour exomes in 18 candidate MCL genes. When considering coding and non-coding mutations in all candidate MCL genes, *HNRNPH1* was mutated in 10% of cases and were predominately splicing and intronic mutations within or immediately flanking exon 4.

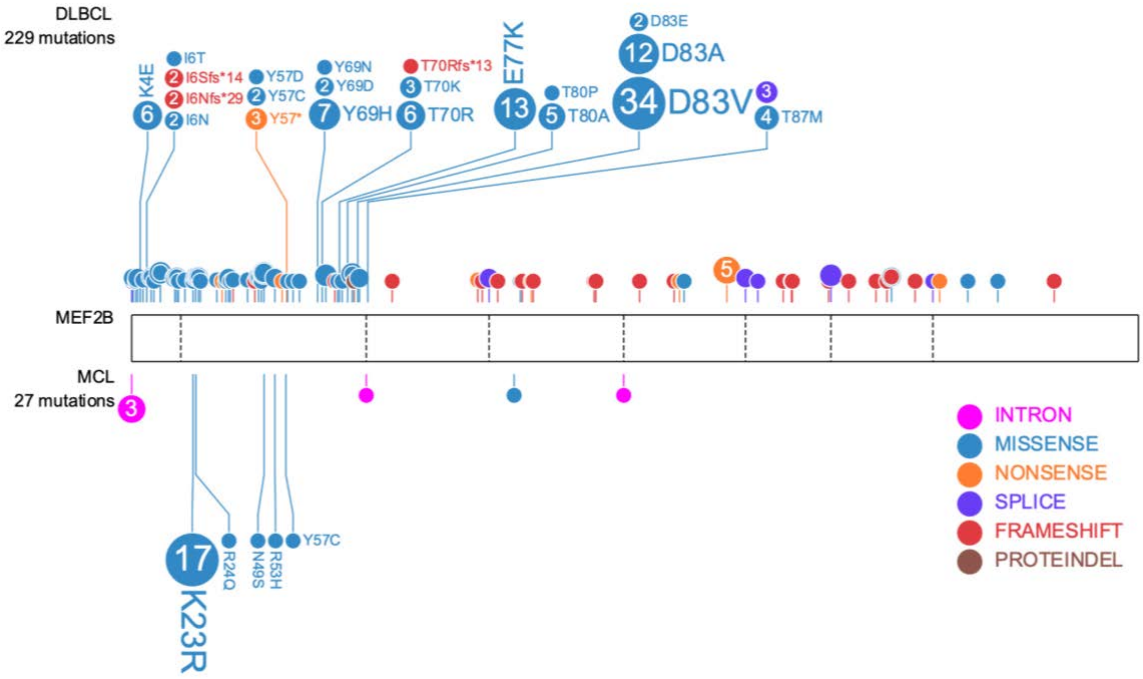

Figure A.2 *MEF2B* Mutation Pattern in MCL and DLBCL exomes
The pattern of mutations in *MEF2B* is distinct between MCL and DLBCL. In MCL, K23R was the predominant mutation whereas this mutation was only observed in a single DLBCL case.

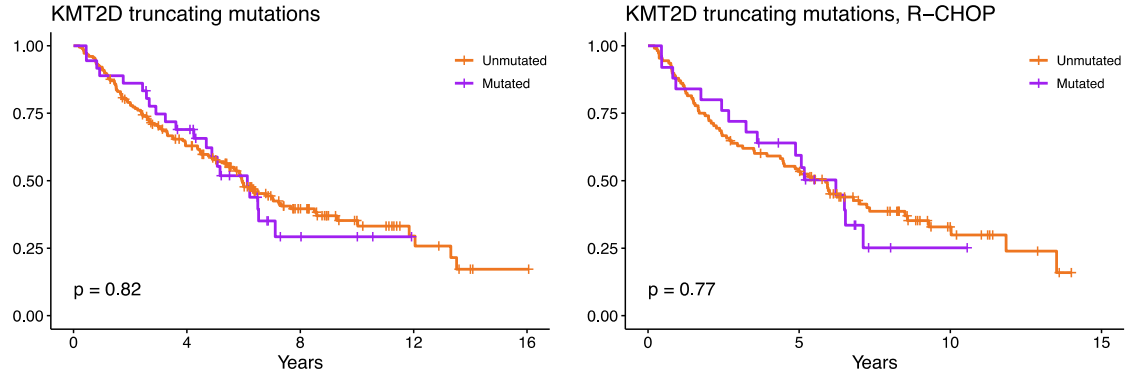

Figure A.3 Overall survival of *KMT2D* truncated MCL tumours
Left: overall survival of MCL tumours stratified by *KMT2D* truncating mutations. *Right*: overall survival of MCL tumours stratified by *KMT2D* truncating mutations restricted to only R-CHOP treated patients.

Appendix B.

Supplemental Data File

Description:

Supplementary Data 1. Mutation Frequency of MCL genes in tumour exomes. Mutation counts of panel genes. “Paired Total” is the number of mutations found in tumour-normal paired analyses (i.e., exomes and/or genomes).

Supplementary Data 2. *HNRNPH1* mutations identified in newly sequenced cases (BC) and published cohorts (Wu et al., Bea et al., Agarwal et al., Khodadoust et al., Rule et al.).

Supplementary Data 3. Digital PCR primers and probes for *HNRNPH1* (canonical and alternative isoforms), *TBP*, *YWHAZ*, *UBC*, *ATCB*.

Supplementary Data 4. HNRNPH1 plasmid primers.

Supplementary Data 5. Site-directed mutagenesis primers for *HNRNPH1*.

Supplementary Data 6. Patient metadata of MCL genomes. MCLs from Nadeau et al. or Pararajalingam, Koyle, et al. not re-published here.

Supplementary Data 7. *CCND1* structural variations in MCL genomes. Obtained from Manta and converted to BEDPE format.

Supplementary Data 8. Copy number variations in MCL genomes in SEG format. Derived from Battenberg and Control-FREEC.

Supplementary Data 9. Differentially expressed genes (FDR < 0.1) within GISTIC2.0 wide peaks.

Supplementary Data 10. Structural variations in MCL genomes. Obtained from Manta and converted to BEDPE format.

Supplementary Data 11. Simple somatic mutations in MCL genomes in MAF format. Mutations restricted to coding and splicing mutations.

Supplementary Data 12. aSHM/SHM regions derived from BL/DLBCL genomes and including putative *CCND1* hypermutation region.

Filename:

PararajalingamPrasath_Supplemental_data.xlsx