

Trisomy 12 in Chronic Lymphocytic Leukaemia

by

Anya Kristen Hotinski

*Thesis
Submitted to Flinders University
for the degree of*

Doctor of Philosophy

College of Medicine & Public Health

22nd October 2021

CONTENTS

I.	ABSTRACT	V
II.	DECLARATION	VII
III.	ACKNOWLEDGEMENTS	VIII
IV.	LIST OF PUBLICATIONS	IX
V.	LIST OF COMMON ABBREVIATIONS	X
VI.	LIST OF FIGURES.....	XIII
VII.	LIST OF TABLES	XV
1	INTRODUCTION	17
1.1	NORMAL B CELL BIOLOGY.....	17
1.2	DIAGNOSIS OF CLL	18
1.2.1	<i>Morphology of CLL</i>	18
1.2.2	<i>Immunophenotype of CLL</i>	18
1.3	CLINICAL FEATURES & STAGING OF CLL.....	19
1.3.1	<i>Rai and Binet clinical staging systems</i>	20
1.4	CONDITIONS RELATED TO CLL	21
1.4.1	<i>Monoclonal B-lymphocytosis</i>	21
1.4.2	<i>Small Lymphocytic Lymphoma</i>	21
1.4.3	<i>Richter's syndrome</i>	21
1.5	AETIOLOGY OF CLL	21
1.5.1	<i>B cell receptor signalling in disease pathogenesis</i>	22
1.5.2	<i>The role of the microenvironment in disease pathogenesis</i>	25
1.6	TREATMENT OF CLL.....	27
1.6.1	<i>Chemoimmunotherapy</i>	27
1.6.2	<i>Small molecule inhibitors</i>	28
1.7	PROGNOSIS OF CLL.....	29
1.7.1	<i>Prognostic impact of the mutational status of the immunoglobulin heavy variable gene (IGHV)</i> . 30	
1.7.2	<i>Prognostic impact of CD38 and ZAP-70 expression</i>	31
1.7.3	<i>Prognostic impact of genomic lesions</i>	31
1.8	RECURRENT GENETIC LESIONS IN CLL.....	32
1.8.1	<i>Recurrent chromosomal aberrations in CLL</i>	33
1.8.2	<i>Recurrent somatic mutations in CLL</i>	34
1.9	TRISOMY 12 IN CLL.....	35
1.9.1	<i>Aneuploidy in cancer</i>	35
1.9.2	<i>Incidence</i>	35
1.9.3	<i>Phenotype</i>	36
1.9.4	<i>Treatment</i>	37
1.9.5	<i>Prognosis</i>	38
1.9.6	<i>Associated genetic lesions</i>	42
1.9.7	<i>Clonal frequency & evolution</i>	43
1.9.8	<i>Gene expression changes</i>	45
1.9.9	<i>B cell receptor signalling</i>	45
1.9.10	<i>Integrin signalling and interaction with the microenvironment</i>	46
1.9.11	<i>NOTCH signalling & mutations in trisomy 12 CLL</i>	48
1.10	CONCLUSION	49
1.11	HYPOTHESIS & AIMS	50
1.11.1	<i>Hypothesis</i>	50

1.11.2	Major Aim	50
1.11.3	Specific Aims	50
2	GENERAL MATERIALS AND METHODS	51
2.1	BUFFERS	51
2.2	PATIENT SAMPLES.....	51
2.2.1	Ethics & sample sources.....	51
2.2.2	Sample storage	51
2.3	CELL CULTURE.....	52
2.3.1	Cell thawing	52
2.3.2	Primary cell culture	52
2.4	DNA EXTRACTION	52
2.5	RNA EXTRACTION	53
2.6	COMPLEMENTARY DNA GENERATION	53
2.7	PRIMERS.....	54
2.8	PCR	54
2.9	SANGER SEQUENCING.....	54
2.10	REAL-TIME QUANTITATIVE PCR.....	54
3	CHARACTERISATION OF THE LOCAL COHORT OF TRISOMY 12 CLL	56
3.1	INTRODUCTION	56
3.2	METHODS.....	56
3.2.1	Patient identification, selection & cytogenetic analysis.....	56
3.2.2	Clinical annotation	57
3.2.3	IGHV analysis	57
3.2.4	Next-generation sequencing (5-gene panel).....	58
3.2.5	Sanger sequencing confirmation of variants.....	58
3.2.6	Immunophenotyping.....	59
3.3	RESULTS	60
3.3.1	Retrospective audit of trisomy 12 cohort.....	60
3.3.2	Prospective analysis of trisomy 12 cohort with available samples	63
3.4	DISCUSSION	72
4	INVESTIGATION OF TWO CASES OF TRISOMY 12 CLL WITH BIMODAL CD49D EXPRESSION	79
4.1	INTRODUCTION	79
4.2	METHODS.....	79
4.2.1	Flow cytometry.....	79
4.2.2	DNA extraction & storage for RNA extraction.....	80
4.2.3	Single Nucleotide Polymorphism array	81
4.2.4	Whole Exome Sequencing	81
4.2.5	Bioinformatics	81
4.2.6	Sanger sequencing	82
4.3	RESULTS	82
4.3.1	Results of IGHV analysis and SNP microarray: Patient 6.....	82
4.3.2	Results of WES: Patient 6	84
4.3.3	Confirmation of WES variants	90
4.3.4	Post-treatment changes: Patient 6	90
4.3.5	Results of SNP microarray and targeted mutation analysis: Patient 21	93
4.4	DISCUSSION	96
4.4.1	Patient 6.....	97
4.4.2	Patient 21	102
4.4.3	Conclusion	102

5	ANALYSIS OF GENE EXPRESSION IN TRISOMY 12 CLL	106
5.1	INTRODUCTION	106
5.2	METHODS	107
5.2.1	<i>RNA extraction</i>	107
5.2.2	<i>RNAseq</i>	107
5.2.3	<i>Bioinformatics</i>	108
5.2.4	<i>Quantitative RT-PCR confirmations</i>	108
5.2.5	<i>Bisulfite treatment and sequencing of ITGA4</i>	108
5.2.6	<i>Statistics</i>	108
5.3	RESULTS	109
5.3.1	<i>Gene expression changes in trisomy 12 CLL</i>	109
5.3.2	<i>Regulation of ITGA4 (CD49d): methylation analysis</i>	123
5.4	DISCUSSION	126
6	INVESTIGATION OF TOLL-LIKE SIGNALLING IN TRISOMY 12 CLL	133
6.1	INTRODUCTION	133
6.2	METHODS	135
6.2.1	<i>TLR4, IL8 and CD14 expression analysis</i>	135
6.2.2	<i>Statistics</i>	137
6.3	RESULTS	137
6.3.1	<i>Expression of TLR signalling components in RNAseq from patient 6</i>	137
6.3.2	<i>Expression of TLR4 pathway targets at baseline</i>	140
6.3.3	<i>Stimulating TLR4 with lipopolysaccharide (LPS)</i>	146
6.4	DISCUSSION	151
7	CONCLUSIONS AND FUTURE DIRECTIONS	155
7.1	CD49D IN CLL	155
7.2	CLL HETEROGENEITY AND SAMPLING ISSUES	156
7.3	ASSAY TECHNOLOGIES	157
7.4	PROGNOSTICATION IN CLL	159
7.5	GERMLINE PREDISPOSITION TO CLL	160
7.6	CLONAL EVOLUTION IN CLL	160
7.7	PATHOGENESIS OF TRISOMY 12 CLL	161
8	REFERENCES	163
9	APPENDIX	176
9.1	PRIMERS	176
9.2	SUMMARY OF PATIENTS: SAMPLES & CLINICAL DETAILS	178
9.3	SUMMARY OF PATIENTS: CYTOGENETICS RESULTS	182
9.4	SUMMARY OF PATIENTS: IGHV & NGS RESULTS	185
9.5	SUMMARY OF PATIENTS: TESTS PERFORMED	187
9.6	WES VARIANTS COMMON TO BOTH LEUKAEMIC CLONES (PATIENT 6)	189
9.7	WES VARIANTS IN CD49D+ CLONE ALONE (PATIENT 6)	194
9.8	WES VARIANTS IN CD49D- CLONE ALONE (PATIENT 6)	205
9.9	CURATED GENE LISTS FROM RNASEQ (PATIENT 6)	214
9.10	FLOW CYTOMETRY DATA	218
9.11	REAL-TIME PCR EXPRESSION DATA	220
9.12	STIMULATION ASSAY DATA	222

I. ABSTRACT

Chronic Lymphocytic Leukaemia (CLL) is a common incurable haematological malignancy of B lymphocytes. Trisomy of chromosome 12 is a recurrent genomic abnormality in CLL with unique associations and clinical response but an unclear molecular pathogenesis. Comprehensive immunogenetic characterisation of a local cohort of trisomy 12 CLL was performed and confirmed the known association with cell surface expression of the integrin and poor prognostic marker, CD49d. Two trisomy 12 CLL cases with bimodal expression of CD49d were identified and extensively interrogated after flow-cytometry cell sorting of the CD49d+ and CD49d- CLL cell populations. One of these cases was comprised of two completely unique leukaemic clones with different immunoglobulin heavy variable gene (IGHV) usage and mutational status, different karyotypes, and different exome sequencing variants. The CD49d+ clone harboured trisomy 12, a hypermutated IGHV4-34 and a lysine methyltransferase 2D (*KMT2D*) mutation on chromosome 12 that represents a putative novel driver of trisomy 12 CLL. The CD49d- clone had disomy 12, a sub-clonal deletion of chromosome 17p, an unmutated IGHV3-21 and a splicing factor 3b subunit 1 (*SF3B1*) mutation, a previously reported high-risk feature of CLL. Both clones harboured mutations in common that were not present in the non-malignant T cells which implies the existence of a pre-leukaemic progenitor cell prior to the IGHV gene rearrangement that occurs during lymphoid maturation and prior to the development of CLL. As well as the general implications for the clonal evolution of CLL, the case also demonstrated that the high-risk feature of CD49d expression does not necessarily associate with other high-risk features such as an unmutated IGHV and may not be a true driver of leukaemia (being normally expressed at high levels in mature B cells). Regulation of integrin subunit alpha 4, *ITGA4* (the CD49d gene), was shown not to be dependent on the methylation status of the promoter, at least in this case, in contrast to previous reports.

A comparison of the transcriptome of the trisomy 12 and disomy 12 clones in this biclonal case of CLL was also performed to identify pathways that are differentially regulated in trisomy 12 CLL. RNAseq of the clones implicated the potential importance of toll-like receptor signalling (via toll-like receptor 4, TLR4) in trisomy 12 CLL. Furthermore, tumour necrosis factor alpha induced protein 3 (TNFAIP3), an inhibitor of the TLR4 pathway, was identified in the gene-set expression analysis and is itself a target of the epigenetic regulator, *KMT2D*, which was also found to be mutated in the trisomy 12 clone. The TLR4 pathway was stimulated via addition of lipopolysaccharide (LPS, an immunogenic component of bacterial cell walls) to a cohort of 7 primary trisomy 12 and 4 disomy 12 CLL samples. Cell viability and RNA and protein expression of a range of intermediaries in the TLR4 pathway were measured after a 48-hour incubation. There was significantly higher expression of

TLR4 and CD14 (a component of the TLR4 complex) in the trisomy 12 group at baseline which approximated expression levels observed in normal B lymphocytes. No difference in cell viability or surface expression of TLR4, CD49d, CD14, however, was identified between the two groups following stimulation with LPS. In addition to this, there were no changes in mRNA expression of *TNFAIP3*, *KMT2D*, *TLR4* or interleukin 8, *IL8* (a downstream pro-inflammatory chemokine of TLR4 signalling), between the trisomy 12 and disomy 12 groups either at baseline or following stimulation. The stimulation assays were limited by low sample numbers of variable quality in terms of cell viability upon thawing, total viable cell numbers and RNA quality. A dependence of trisomy 12 CLL on TLR4 signalling could not be confirmed during this thesis.

In conclusion, this thesis presents evidence that trisomy 12 CLL cells more closely resemble normal B lymphocytes than their disomic CLL counterparts, demonstrates the importance of thorough investigation of unique individual cases to gain insights into clonal evolution and genomic complexity in CLL, and provides an avenue for future research in toll-like receptor signalling pathways to advance the understanding of early drivers of trisomy 12 CLL.

II. DECLARATION

I certify that this thesis:

1. Does not incorporate without acknowledgement any material previously submitted for a degree or diploma in the university; and
2. To the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

A handwritten signature in blue ink, consisting of stylized, overlapping loops and curves, positioned below the declaration text.

III. ACKNOWLEDGEMENTS

I would like to acknowledge the contribution of the Australian Government Research Training Program Scholarship (AGRTPS). A subset of the biospecimens utilised in this thesis were provided by the South Australian Cancer Research Biobank (SACRB), which is supported by the Cancer Council SA Beat Cancer Project, Medvet laboratories Pty Ltd and the Government of South Australia. Other biospecimens were obtained from the Australasian Leukaemia and Lymphoma Group (ALLG) CLL06 clinical trial with thanks.

I would like to sincerely thank all the patients and their families who consented for their samples to be used in this thesis – it would be truly impossible to work on this disease without their altruism and kind donations.

No professional editor was used in the preparation of this thesis.

I acknowledge the contribution of the following people who were of assistance with regards to technical support: Dr Sheree Bailey, Ms Rachell Hall, and Mr Athanasios (Arthur) Krotiris (all of SA Pathology, Southern Adelaide Local Health Network), and Ms Sarah Moore and Ms Chun Chun Young (SA Pathology, Central Adelaide Local Health Network).

I would also like to thank the academic members of the Molecular Medicine and Genetics Department for your advice and support – Dr Binoy Appukuttan, Dr Stephen Gregory, Dr Lauren Thurgood, and Dr Giles Best. Thank you especially to Dr Thurgood for her methodological assistance and coffee dates, and to Dr Best for his expert flow cytometry support. Thank you to the many medical science honours' and PhD students who shared my office space and kept the snacks drawer fully stocked. Thank you to Lara Escane, soon to be Dr Escane, for keeping me on track. Thank you to Ms Alana White for assistance with creating an introductory figure.

An enormous thank you is due to my supervisors – Associate Professors Bryone Kuss, Karen Lower and David Ross. In particular, Bryone and Karen were brilliant - not only on a scientific level, but also on a personal level and supported me through the trials and tribulations inherent to a PhD and through my maternity leave. Words cannot fully express my deep gratitude.

Finally, a big thanks to all my friends and family for your support through this “indulgence”! Thank you, James, Alexander, Ava, and George! You were all so enthusiastically supportive (mostly!), however, I dedicate this thesis to the baby of the family, “Furious George”.

IV. LIST OF PUBLICATIONS

Hotinski A, Best O, Thurgood L, Lower K, Kuss B. A biclonal cases of Chronic Lymphocytic Leukaemia with discordant mutational status of the Immunoglobulin Heavy Chain Variable region (*IGHV*) and bimodal CD49d expression. *Br J Haematol*. <https://doi.org/10.1111/bjh.17257>

Hotinski A, Best O, Kuss B. The future of laboratory testing in chronic lymphocytic leukaemia. *Pathology*. Online 5th March 2021. <https://doi.org/10.1016/j.pathol.2021.01.006>

V. LIST OF COMMON ABBREVIATIONS

+12	trisomy 12
ATM	ataxia telangiectasia mutated
BCL-2	B cell lymphoma 2
BCL11B	BAF chromatin remodeling complex subunit BCL11B
BCR	B cell receptor
BIRC3	baculoviral IAP repeat containing 3
BTK	Bruton's Tyrosine Kinase
CD	cluster of differentiation
cDNA	complementary DNA
CGH	comparative genomic hybridisation
CHIP	Clonal Haematopoiesis of Indeterminate Potential
CLL	Chronic Lymphocytic Leukaemia
CR	complete remission
Del(11q)/11q-	deletion of chromosome 11q
Del(13q)/13q-	deletion of chromosome 13q
Del(17p)/17p-	deletion of chromosome 17p
DLBCL	diffuse large B cell lymphoma
DNA	deoxyribonucleic acid
ERK	extracellular signal-regulated kinase
EZH2	enhancer of zeste homolog 2
FBXW7	F-box and WD repeat containing protein 7
FCR	fludarabine, cyclophosphamide, rituximab
FCS	fetal calf serum

FISH	fluorescence <i>in situ</i> hybridisation
Ig	immunoglobulin
IGH	immunoglobulin heavy locus
IGHV	immunoglobulin heavy variable gene
IGLV	immunoglobulin lambda variable gene
IL	interleukin
IRF4	interferon regulatory factor 4
<i>ITGA4</i>	Integrin subunit alpha 4 (CD49d gene)
iwCLL	International Workshop on CLL
KMT2D	histone-lysine N-methyltransferase 2D
LDH	lactate dehydrogenase
MBL	monoclonal B-lymphocytosis
M-CLL	subtype of CLL with a mutated IGHV
MDC1	mediator of DNA damage checkpoint protein 1
MEK	mitogen-activated protein kinase kinase
MFI	mean fluorescence intensity
mTOR	mechanistic target of rapamycin
MRD	minimal residual disease
NF-κB	nuclear factor-κB
NOTCH1	notch receptor 1
ORR	overall response rate
OS	overall survival
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PFS	progression-free survival

PI3K	phosphatidylinositol 3-kinase
qRT-PCR	real-time quantitative PCR
RIN	RNA integrity number
RNA	ribonucleic acid
sd	standard deviation
SF3B1	splicing factor 3b subunit 1
SLL	Small lymphocytic lymphoma
SNP	single nucleotide polymorphism
SNV	single nucleotide variation
TET2	tet methylcytosine dioxygenase
TLR	toll-like receptor
TNFAIP3	tumour necrosis factor alpha induced protein 3
TP53	tumour protein P53
TTFT	time to first treatment
U-CLL	subtype of CLL with an unmutated IGHV
VLA-4	very late antigen 4
WES	whole exome sequencing

VI. LIST OF FIGURES

FIGURE 1-1. MORPHOLOGY OF CLL.....	19
FIGURE 1-2. LYMPHADENOPATHY AT DIAGNOSIS IN A PATIENT WITH CLL.	20
FIGURE 1-3. B CELL RECEPTOR SIGNALLING [31].....	23
FIGURE 1-4. THE CLL MICROENVIRONMENT [30].	26
FIGURE 1-5. PROGNOSTIC IMPACT OF GENOMIC LESIONS IN CLL [78]	32
FIGURE 1-6. GENETIC ABNORMALITIES IN CLL [29].	33
FIGURE 1-7. KAPLAN-MEIER TIME TO FIRST TREATMENT (TTT) CURVE COMPARING TRISOMY 12 CLL (BLUE LINE) TO NON-TRISOMY 12, NON-DEL(11Q), NON-DEL(17P) CLL (GREEN LINE) AS PUBLISHED BY STRATI ET AL. [124]	40
FIGURE 1-8. RECURRENT MUTATIONS REFINE PROGNOSIS IN TRISOMY 12 CLL.....	41
FIGURE 1-9. TRISOMY 12 IS HYPOTHESISED TO BE AN EARLY EVENT IN THE EVOLUTION OF CLL.....	44
FIGURE 3-1. KAPLAN-MEIER CURVES OF OVERALL SURVIVAL AND TIME TO FIRST TREATMENT FOR LOCAL COHORT OF TRISOMY 12 CLL.	61
FIGURE 3-2. NUMBER OF CYTOGENETIC ABNORMALITIES IN ADDITION TO TRISOMY 12 (A) AND CLONAL FREQUENCY OF TRISOMY 12 (B) IN COHORT AS DETERMINED BY SNP MICROARRAY (BLUE), CBA (ORANGE) OR FISH (GREY).	62
FIGURE 3-3. CYTOGENETIC ABNORMALITIES IN ADDITION TO TRISOMY 12 IN THE LOCAL COHORT (AS DETECTED BY ANY METHOD)	63
FIGURE 3-4. EXAMPLE OF HIERARCHICAL GATING STRATEGY (PATIENT 6).....	68
FIGURE 3-5. CD38, CD11B AND CCR7 EXPRESSION IN TRISOMY 12 AND DISOMY 12 CLL.....	69
FIGURE 3-6. CD49D EXPRESSION IN TRISOMY 12 AND DISOMY 12 CLL.....	70
FIGURE 3-7. TWO TRISOMY 12 CLL CASES EXHIBIT BIMODAL EXPRESSION OF CD49D.....	71
FIGURE 4-1. FLOW SORTING OF PATIENT 6.	83
FIGURE 4-2. SNP MICROARRAY AND IGHV RESULTS OF THE SORTED FRACTIONS FROM PATIENT 6.....	84
FIGURE 4-3. CNV ANALYSIS FROM WES DATA OF SORTED FRACTIONS FROM PATIENT 6	85
FIGURE 4-4. TYPES OF VARIANTS DETECTED ON WHOLE EXOME ANALYSIS OF FRACTIONATED SAMPLES OF PATIENT 6.	87
FIGURE 4-5. NUMBER OF PROTEIN-ALTERING VARIANTS DETECTED ON WHOLE EXOME ANALYSIS OF FRACTIONATED SAMPLE FROM PATIENT 6.	87
FIGURE 4-6. SANGER SEQUENCING CONFIRMATION OF VARIANTS DETECTED BY WES ON SORTED AND GERMLINE DNA FROM PATIENT 6.....	92
FIGURE 4-7. POST-TREATMENT CHANGE IN CD49D+ CLL CELLS IN PATIENT 6	92
FIGURE 4-8. FLOW SORT OF PATIENT 21	94
FIGURE 4-9. PURITY CHECK FROM FLOW SORT OF PATIENT 21.....	95
FIGURE 4-10. NOTCH1 AND BIRC3 SANGER SEQUENCING RESULTS ON UNSORTED AND SORTED SAMPLES FROM PATIENT 21.	96
FIGURE 4-11. PROPOSED EVOLUTION OF CLL IN PATIENT 6 (SEE TEXT).	104
FIGURE 4-12. PROPOSED EVOLUTION OF CLL IN PATIENT 21 (SEE TEXT).	105
FIGURE 5-1. NORMALISED READ COUNT PLOT OF RNASEQ OF TWO CLL CLONES FROM PATIENT 6.....	110
FIGURE 5-2. HISTOGRAM OF CHANGES IN GENE EXPRESSION LEVELS BETWEEN THE TWO CLL CLONES FROM PATIENT 6.	110
FIGURE 5-3. GENE-SET ENRICHMENT ANALYSIS OF RNASEQ DATA (PATIENT 6) USING PANTHER VIA THE ENRICHR PLATFORM.	112
FIGURE 5-4. RNASEQ ANALYSIS OF KMT2D GENE TRANSCRIPT IN PATIENT 6	114
FIGURE 5-5. ITGA4 MRNA EXPRESSION (RELATIVE TO THE HOUSEKEEPING GENE GUSB) IN TRISOMY 12 AND DISOMY 12 CLL (A) AND IN CD49D+ AND CD49D- CLL (B).	117
FIGURE 5-6. EXPRESSION OF KMT2D AND TNFAIP3 (RELATIVE TO THE HOUSEKEEPING GENE GUSB) IN A COHORT OF TRISOMY 12 AND DISOMY 12 CLL (A AND B RESPECTIVELY)	119

FIGURE 5-7. EXPRESSION OF IRF4 AND EZH2 (RELATIVE TO THE HOUSEKEEPING GENE GUSB) IN A COHORT OF CD49D+ AND CD49D- CLL (A AND B RESPECTIVELY)	120
FIGURE 5-8. RNA QUALITY IS AFFECTED BY THE METHOD OF SAMPLE PROCESSING BUT NOT THE TIME TO PROCESS (WITHIN 30-HOURS)	123
FIGURE 5-9. MAP OF ITGA4 AMPLICONS	124
FIGURE 5-10. BISULFITE SEQUENCING OF ITGA4 FROM SORTED SAMPLES (PATIENT 6)	125
FIGURE 5-11. HYPOTHESISED ROLE OF KMT2D IN TRISOMY 12 CLL	130
FIGURE 6-1. KMT2D TARGETS OF KEY B CELL REGULATORY PATHWAYS	134
FIGURE 6-2. RNASEQ COVERAGE OF TLR4 FOR THE CD49D+ AND CD49D- CLONES OF PATIENT 6	138
FIGURE 6-3. THE KEGG TOLL-LIKE RECEPTOR SIGNALLING PATHWAY [244]	139
FIGURE 6-4. RELATIVE MRNA EXPRESSION OF TLR4 AND IL8 AT BASELINE	141
FIGURE 6-5. TLR4 EXPRESSION ON THE CD49D+ AND CD49D- CLL CLONES IN PATIENT 6	143
FIGURE 6-6. TLR4 AND CD14 CELL SURFACE EXPRESSION	144
FIGURE 6-7. CORRELATIONS BETWEEN CD49D, TLR4 AND CD14 PROTEIN EXPRESSION AS DETERMINED BY FLOW CYTOMETRY	145
FIGURE 6-8. CORRELATION BETWEEN TLR4 MRNA AND PROTEIN EXPRESSION	146
FIGURE 6-9. CELL VIABILITY AFTER 48-HOURS STIMULATION WITH LPS	148
FIGURE 6-10. CHANGE IN TLR4, CD49D AND CD14 PROTEIN EXPRESSION AFTER 48-HOURS OF LPS-STIMULATION	149
FIGURE 6-11. CHANGE IN TLR4, ITGA4, KMT2D, TNFAIP3 AND IL8 MRNA EXPRESSION AFTER 48-HOURS OF LPS STIMULATION	150

VII. LIST OF TABLES

TABLE 1-1. BINET AND MODIFIED RAI CLINICAL STAGING SYSTEMS OF CLL	20
TABLE 1-2: TWO MAIN SUBTYPES OF CLL ARE DISTINGUISHED BY THE MUTATIONAL STATUS OF THE IGHV .	24
TABLE 1-3. OVERVIEW OF THE FOUR MOST COMMON STEREOTYPED SUBSETS OF CLL	30
TABLE 3-1. DETAILS OF ANTIBODY PANEL USED IN IMMUNOPHENOTYPIC CHARACTERISATION OF TRISOMY 12 COHORT	59
TABLE 3-2. SUMMARY OF RESULTS FOR 22 TRISOMY 12 CLL PATIENTS.....	64
TABLE 3-3. IGHV RESULTS FOR THE LOCAL TRISOMY 12 CLL COHORT.....	65
TABLE 3-4. MUTATIONS IN ATM, BIRC3, NOTCH1, SF3B1 AND TP53 IN TRISOMY 12 CLL COHORT	67
TABLE 4-1. IGHV AND SNP MICROARRAY RESULTS OF THE UNSORTED AND THREE SORTED FRACTIONS FROM PATIENT 6.	83
TABLE 4-2. NUMBER OF VARIANTS DETECTED ON WHOLE EXOME ANALYSIS OF FRACTIONATED SAMPLES FROM PATIENT 6.	86
TABLE 4-3. LIST OF POTENTIAL DRIVER MUTATIONS (WITH A VAF OF GREATER THAN 45%) COMMON TO BOTH CLONES (BUT ABSENT IN T CELLS) AND PRESENT SOLELY IN THE CD49D+ OR CD49D- CLL CLONES.	88
TABLE 4-4. SUMMARY OF IMMUNOGENETIC FEATURES OF TWO DIFFERENT CLL CLONES IN PATIENT 6.	91
TABLE 4-5. SNP ARRAY RESULTS OF SORTED FRACTIONS FROM PATIENT 21.....	95
TABLE 5-1. NUMBER OF GENES UP- OR DOWN-REGULATED BY AT LEAST 2-FOLD IN THE CD49D+ CLONE WITH RESPECT TO THE CD49D- CLONE.	111
TABLE 5-2. EXPRESSION LEVELS AND READ COUNTS OF MDC1, KMT2D, BCL11B AND SELECTED KMT2D TARGETS IN THE RNASEQ OF CD49D+ AND CD49D- CLONES OF PATIENT 6.	113
TABLE 5-3. EXPRESSION OF EZH2 AND IRF4 IN RNASEQ OF CD49D+ AND CD49D- CLONE OF PATIENT 6	115
TABLE 5-4. COMPARISON OF RNASEQ AND QRT-PCR DATA FOR PATIENT 6	116
TABLE 5-5. RNA QUALITY.	122
TABLE 5-6. NUMBER OF METHYLATED CPGS IN ITGA4 AMPLICONS OF SORTED SAMPLES FROM PATIENT 6 AND 21.	124
TABLE 6-1. ANTIBODY COCKTAIL FOR TLR4 AND CD14 EXPRESSION ANALYSIS.....	136
TABLE 6-2. EXPRESSION LEVELS AND READ COUNTS OF GENES IN TOLL SIGNALLING PATHWAY IN THE RNASEQ OF PATIENT 6	138
TABLE 9-1. PRIMER SEQUENCES FOR SANGER SEQUENCING.	176
TABLE 9-2. PRIMER SETS USED TO CONFIRM LISTED BIRC3 MUTATIONS IN CHAPTER 3.....	176
TABLE 9-3. PRIMER SEQUENCES FOR QRT- PCR.	177
TABLE 9-4. PRIMER SEQUENCES FOR BISULFITE-TREATED DNA	177
TABLE 9-5. SUMMARY OF ENTIRE PATIENT COHORT: IDENTIFICATION, SAMPLE INFORMATION AND CLINICAL DETAILS	178
TABLE 9-6. SUMMARY OF CYTOGENETICS RESULTS FOR ENTIRE TRISOMY 12 COHORT & NON-TRISOMY 12 CLL CONTROLS.	182
TABLE 9-7. SUMMARY OF IGHV AND NGS PANEL RESULTS FOR TRISOMY 12 COHORT WITH AVAILABLE SAMPLES (PATIENTS 1-22).....	185
TABLE 9-8. SUMMARY OF TESTS PERFORMED ON PATIENT SAMPLES.....	187
TABLE 9-9. VARIANTS COMMON TO BOTH CD49D+ AND CD49D- CLL CLONES DETECTED IN WES OF SORTED FRACTIONS OF PATIENT 6	189
TABLE 9-10. VARIANTS PRESENT IN THE CD49D+ CLL CLONE ALONE AS DETECTED IN WES OF SORTED FRACTIONS OF PATIENT 6.	194
TABLE 9-11. VARIANTS PRESENT IN THE CD49D- CLL CLONE ALONE AS DETECTED IN WES OF SORTED FRACTIONS OF PATIENT 6	205
TABLE 9-12. GENES DOWNREGULATED 10 OR MORE-FOLD (WITH ABSOLUTE READS OF >100) IN THE CD49D+ CLONE IN PATIENT 6	214

TABLE 9-13. GENES UPREGULATED 10 OR MORE-FOLD (WITH ABSOLUTE READS OF >100) IN THE CD49D+ CLONE IN PATIENT 6	216
TABLE 9-14. FLOW CYTOMETRY DATA FROM CHAPTERS 3 AND 6	218
TABLE 9-15. RELATIVE MRNA EXPRESSION OF ITGA4, IRF4, EZH2, KMT2D AND TNFAIP3 FROM A COHORT OF TRISOMY 12 AND DISOMY 12 CLL (CHAPTER 5).	220
TABLE 9-16. RELATIVE MRNA EXPRESSION OF TLR4 AND IL8 FROM A COHORT OF TRISOMY 12 AND DISOMY 12 CLL (CHAPTER 6).....	221
TABLE 9-17. LPS STIMULATION ASSAY DATA: VIABILITY AND IMMUNOPHENOTYPING (CHAPTER 6).....	222
TABLE 9-18. LPS STIMULATION ASSAY DATA: GENE EXPRESSION (CHAPTER 6)	223

1 INTRODUCTION

Chronic Lymphocytic Leukaemia (CLL) is a lymphoid neoplasm resulting from the proliferation of small mature B cells (B lymphocytes) that accumulate in the blood, bone marrow and lymphoid tissues. CLL is the most common leukaemia in Australia with approximately 1000 new cases diagnosed every year [1]. It is more frequent in Western countries, with an incidence rate of <1-5.5 cases per 100,000 people per year [2]. The incidence increases with age and 80% of new cases in Australia are diagnosed in people over the age of 60 [1]. It is extremely rare in children and the median age of diagnosis is variably reported between the 6th and 7th decades of life. There is a slight male preponderance and a familial predisposition to the disease [2, 3].

Normal B cell biology will be discussed at first in brief, followed by an overview of the diagnosis, aetiology, treatment, prognosis, and genetic lesions in CLL. The chapter will culminate in a discussion of aspects specific to CLL with trisomy of chromosome 12, the focus of this thesis.

1.1 NORMAL B CELL BIOLOGY

B lymphocytes are key cells of the vertebrate immune system and express diverse surface immunoglobulin (Ig) molecules that are receptors for specific antigens [4, 5]. They derive from haematopoietic stem cells in the fetal liver and bone marrow and undergo maturation in secondary lymphoid tissue (for example, the lymph nodes and spleen) following exposure to their cognate antigen. Activated B cells terminally differentiate into plasma cells and produce antibodies (immunoglobulins) of the same structure as the surface receptor that function to bind and neutralise antigen.

The immunoglobulin molecule is the key component of the B cell receptor (BCR). It is a heterodimeric protein comprised of two heavy and two light chains that both contain variable domains (that recognise antigen) and constant domains (that determine effector function) [6]. The light chains are either of the kappa or lambda type. There are five main isotypes of the immunoglobulin molecule determined by the heavy chain constant domains (IgA, IgD, IgE, IgG and IgM). The variable domains are responsible for antigen specificity and are created by a series of complex immunoglobulin gene rearrangement events that culminate in a vast repertoire of BCR structures with differing antigen specificities [5].

The immunoglobulin heavy locus (IGH) is located on chromosome 14q and contains numerous heavy variable (IGHV), heavy diversity (IGHD) and heavy joining (IGHJ) genes that are recombined during B cell development and encode the variable domains of the protein heavy chain.

The variable domain is comprised of three hypervariable complementarity-determining regions (CDRs) interspaced between four framework regions. The immunoglobulin kappa locus is located on chromosome 2q, and the immunoglobulin lambda locus is on chromosome 22q: both are comprised of numerous IGLV and IGLJ genes that are also rearranged during B cell development [6].

Further diversity of immunoglobulin is achieved during the germinal centre reaction in lymphoid tissues following exposure to antigen. During this process, somatic hypermutation of the immunoglobulin heavy variable genes occurs and results in further affinity maturation of the immunoglobulin. Class-switch recombination also occurs leading to a switch of Ig isotype. Clonality of the immunoglobulin heavy variable gene (IGHV) is a feature of CLL and will be discussed in sections 1.5.1 and 1.7.1.

Most circulating B cells in healthy adults are naïve B cells expressing surface IgM and IgD that have not been previously exposed to antigen [7]. They comprise approximately 9% of peripheral blood lymphocytes [8]. Total blood lymphocytes measure between 1.5-4.0 x10⁹/L in healthy adults, however, the reference range varies between laboratories [9]. Elevated levels of peripheral blood B lymphocytes are the cornerstone feature of CLL.

1.2 DIAGNOSIS OF CLL

Standardised diagnostic criteria for the diagnosis of CLL have been published by the World Health Organisation and the International Workshop on Chronic Lymphocytic Leukemia [10, 11]. In both criteria, CLL requires the presence of B lymphocytes at levels of greater than 5 x 10⁹/L in the peripheral blood. The malignant B lymphocytes in CLL are identified by both their morphology and immunophenotype by flow cytometry.

1.2.1 Morphology of CLL

The peripheral blood smear shows increased numbers of largely monomorphic small mature lymphocytes that are characterised by a clumped chromatin pattern, absence of nucleoli, and a thin rim of cytoplasm (see Figure 1-1). “Smudge” or “smear” cells are commonly observed. Larger more immature lymphocytes, known as prolymphocytes, may be present but are usually rare and by definition must account for less than 55% of lymphoid cells [2].

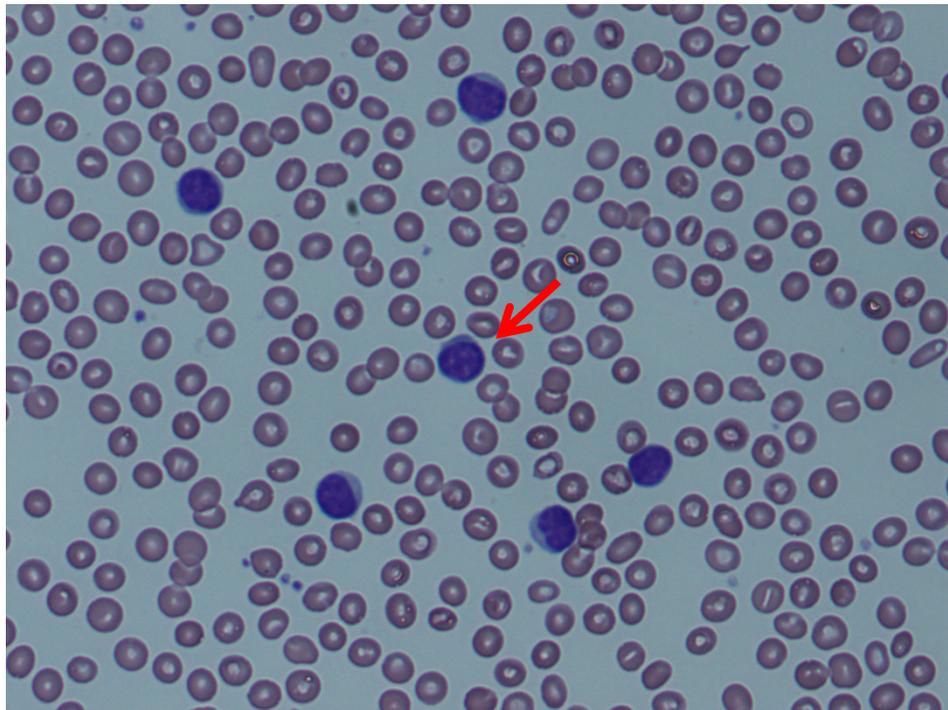
1.2.2 Immunophenotype of CLL

CLL cells co-express the surface T cell marker CD5 aberrantly along with B cell surface antigens CD19, CD20 and CD23 [2, 11]. CD20 expression is usually weak compared to normal B cells and other B cell lymphoid malignancies [12]. There is also dim surface expression of CD79b (a component of the BCR)

and the surface immunoglobulin (IgM and/or IgD) [2]. There is kappa or lambda immunoglobulin light chain restriction. CD10 is negative and FMC7 (an epitope of CD20) is also typically negative [2].

Differential expression of several other cell surface markers and intracellular proteins have demonstrated prognostic significance in CLL. These include CD38, ZAP-70 (see section 1.7.2) and the integrin CD49d (see section 1.9.10.1) and will be discussed in more detail later.

Figure 1-1. Morphology of CLL. The peripheral blood smear demonstrates increased numbers of monomorphic small mature lymphocytes. The red arrow points to a prototypical mature lymphocyte.



1.3 CLINICAL FEATURES & STAGING OF CLL

The clinical features of CLL vary widely. Many patients are asymptomatic at diagnosis and have an incidental finding of a high lymphocyte count (lymphocytosis) on a routine blood test. A minority of patients can present with fatigue, recurrent infection, enlarged lymph nodes (lymphadenopathy, see Figure 1-2), an enlarged liver or spleen (hepatosplenomegaly), autoimmune phenomena or peripheral blood cytopenias (anaemia, neutropenia or thrombocytopenia) from extensive bone marrow infiltration. Two widely recognised staging systems are currently in use, both routinely in the clinic and in clinical trials, and encapsulate the wide variation of patient presentations.

Figure 1-2. Lymphadenopathy at diagnosis in a patient with CLL. Painless, symmetric enlargement of lymph nodes in the cervical chain is demonstrated (red arrows). Photo is shared with patient consent.



1.3.1 Rai and Binet clinical staging systems

The modified Rai and Binet staging systems are summarised in Table 1-1. Both systems recognise more advanced disease by peripheral blood cytopenias owing to bone marrow infiltration by CLL.

Table 1-1. Binet and modified Rai clinical staging systems of CLL. Nodal group refers to a lymph node conglomerate in one anatomic region. Hb = haemoglobin, Plt = platelet count.

Binet staging system [13]		Modified Rai staging system [14]	
Stage	Feature	Stage	Feature
A	≤2 nodal groups affected	Low risk	Lymphocytosis alone
B	≥3 nodal groups affected, and/or hepatomegaly and/or splenomegaly	Intermediate risk	Lymphocytosis and lymphadenopathy (any site), and/or hepatomegaly and/or splenomegaly
C	Hb <100g/L and/or Plt <100 x 10 ⁹ /L	High risk	Hb <100g/L and/or Plt <100 x 10 ⁹ /L

1.4 CONDITIONS RELATED TO CLL

1.4.1 Monoclonal B-lymphocytosis

Monoclonal B-lymphocytosis (MBL) is characterised by a circulating clonal B cell population of less than $5 \times 10^9/L$ in the peripheral blood and the absence of lymphadenopathy, splenomegaly, and disease-related symptoms or cytopenias [15]. Most cases of MBL have the same immunophenotype as CLL, but only some progress to overt CLL. Progression depends partly on the size of the clone: low-count MBL ($\leq 0.5 \times 10^9/L$ clonal B cells) rarely progresses to symptomatic CLL [16]. High-count MBL ($>0.5 \times 10^9/L$ clonal B cells) can progress to symptomatic CLL requiring treatment but at a rate of less than 2% per year [17, 18]. In one series of 185 patients with MBL, overt CLL requiring chemotherapy occurred in 7% after a median follow-up of 6.7 years [17].

1.4.2 Small Lymphocytic Lymphoma

Small lymphocytic lymphoma (SLL) is a disease on the same continuum as CLL and is treated virtually identically to CLL in routine practice and clinical trials. It is comprised of cells with the same morphology and immunophenotype of the malignant cells in CLL, however, the clinical phenotype is different with most of the disease burden confined to the lymph nodes rather than the blood. By definition, there is lymphadenopathy and/or splenomegaly but the level of peripheral blood clonal B lymphocytes is $<5 \times 10^9/L$ [11].

1.4.3 Richter's syndrome

In a minority of patients (5-10%) an aggressive lymphoma can develop on the background of CLL and is termed the Richter's transformation or syndrome [19]. It can be clonally related or unrelated to the underlying CLL and portends a dismal prognosis with a median overall survival of less than 1 year [2, 20].

1.5 AETIOLOGY OF CLL

There is little evidence for the role of environmental factors, such as diet or lifestyle, in the pathogenesis of CLL. There is a positive association between the risk of developing CLL and exposure to the herbicide mixture Agent Orange used during the Vietnam war [21]. Occupational and environmental exposure to other chemicals (for example, in farmers) has been suggested to contribute to the development of CLL, but the exact nature of the specific chemicals and clear causation has not been demonstrated [22].

There are genetic polymorphisms that contribute to disease susceptibility in a minority of patients with CLL [23-27]. Swedish population-based registry data has demonstrated an 8.5-fold

relative risk of CLL and a 1.9-fold relative risk of non-Hodgkin lymphoma in first-degree relatives of CLL patients [28]. Despite this, most cases of CLL are sporadic and there is no single causative genetic lesion responsible for disease [29]. There is strong evidence, however, to support the aetiological role of a limited set of auto- and/or allo-antigens that engage the BCR, leading to dysregulated B cell signalling and contribute to disease pathogenesis. This will be discussed below.

1.5.1 B cell receptor signalling in disease pathogenesis

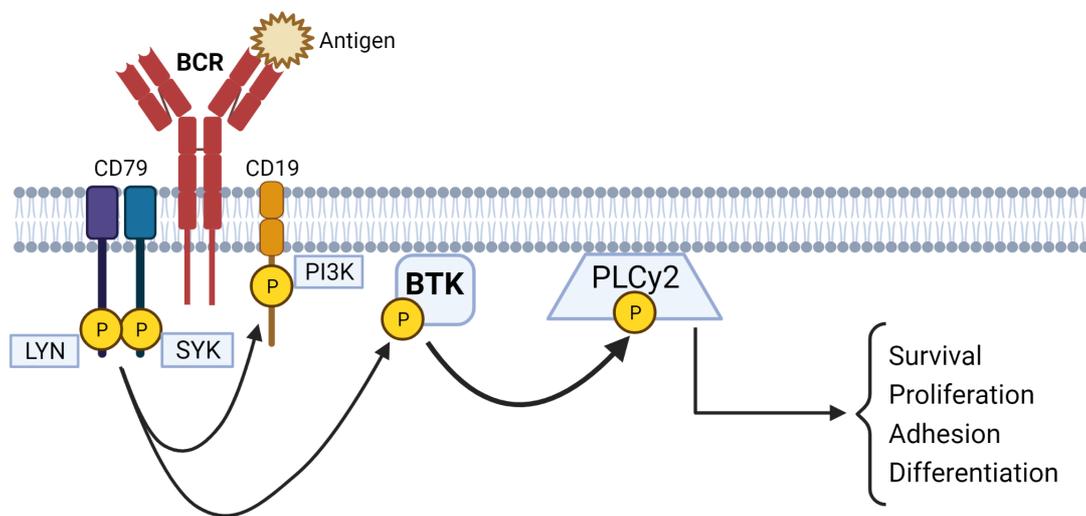
The BCR is a complex of two main components. The first component is an antigen-specific ligand-binding transmembrane immunoglobulin molecule comprising two heavy and two light chains. The second component is a signalling heterodimer comprising the subunits Ig α and Ig β (also known as CD79a and CD79b) that contain immunoreceptor tyrosine-based activation motifs (ITAMs) [30-32]. Upon ligand-binding, there is cross-linking of the BCR and activation of the receptor-associated tyrosine kinases LYN and SYK that act to phosphorylate the ITAMs of CD79. This initiates a series of downstream events that ultimately lead to the activation of B cell proliferation, survival, adhesion, and differentiation pathways (see Figure 1-3). Critical molecules in the pathway include Bruton's Tyrosine Kinase (BTK) and phosphatidylinositol 3-kinase (PI3K). There is also crosstalk between this pathway and chemokine and toll-like receptor signalling pathways. BTK plays a central role in all three pathways and is the target of an effective drug in CLL, ibrutinib (see section 1.6.2.1).

The structure of the BCR is also important in CLL pathogenesis. Two main subtypes of CLL can be identified based on the mutational status of the heavy chain immunoglobulin component of their BCR. Unmutated CLL (U-CLL) is characterised by an immunoglobulin heavy variable gene (IGHV) sequence that is greater than or equal to 98% homologous to the closest germline sequence. Conversely, mutated CLL (M-CLL) has an IGHV sequence that differs by more than 2% to the germline sequence due to somatic hypermutation (a normal process that occurs in the germinal centres of lymphoid tissues during the adaptive immune response). U-CLL and M-CLL exhibit differences in BCR responsiveness to antigen, downstream signalling pathway activation and IGHV gene usage. This will be discussed in brief below and the differences between both important subtypes of CLL is summarised in Table 1-2.

BCRs in CLL have a restricted, biased use of immunoglobulin genes (encoding both heavy and light chain variable domains of the immunoglobulin molecule) differing from the broad diversity observed in the B cell repertoire of healthy individuals [33]. Indeed, approximately 1 in 75 unrelated patients will have CLL cells with virtually identical surface immunoglobulin and approximately 30% will have a highly similar structure of the immunoglobulin heavy chain complementarity-determining region 3 (VH CDR3), known as stereotypy [33-35]. Stereotypy of the BCR implicates selective

pressure by a limited set of auto- and/or allo-antigens involved in clonal expansion and disease ontogeny. Many stereotyped subsets of CLL have been identified to date with highly similar BCR structures and define different groups with distinct biologic and clinical features (see also section 1.7.1) [36].

Figure 1-3. B cell receptor signalling [31]. Following antigen binding of the B cell receptor (BCR), a cascade of signalling events is initiated leading to B cell activation. Dysregulation of signalling is critical in CLL pathogenesis. Adapted from and reprinted by permission from Springer Nature: Nature Publishing Group. *Oncogene* vol. 34, pp.2426-36 (BTK inhibitors in chronic lymphocytic leukemia: a glimpse to the future, Spaargaren et al.). © Macmillan Publishers Limited, 2014.



Despite these findings, the exact nature of the antigens implicated in disease pathogenesis is largely unknown. There is evidence that neo-autoantigens developed through the apoptotic process (either from the relocation of antigens normally confined within cells to the cell surface, or through oxidative changes leading to neo-epitopes) are involved. Around 60% of U-CLL cases express an immunoglobulin molecule that demonstrates low-affinity recognition and polyreactivity with apoptotic cells [37]. Interestingly, some of these neo-epitopes recognised have remarkable similarity to microbial antigens. In subset 6 CLL, the BCR monoclonal antibody recognises nonmuscle myosin heavy chain IIA that is exposed during apoptosis, and in another stereotyped subset of M-CLL (V3-7Sh) the cognate antigen has been identified as a yeast wall component [38-40]. In other cases of M-CLL, there is left often stereotypy of the BCR receptor which is oligo- or mono-reactive (rather than polyreactive as seen in U-CLL) and largely unknown high affinity autoantigens are thought to be involved in tonic antigen-dependent B cell signalling [29].

Table 1-2: Two main subtypes of CLL are distinguished by the mutational status of the IGHV. The differences between each group are summarised in the table below. IGHV = immunoglobulin heavy variable gene. Ig = immunoglobulin.

	Unmutated IGHV (U-CLL)	Mutated IGHV (M-CLL)
Putative cell of origin	Naïve pre-germinal centre B cell	Post-germinal centre B cell after somatic hypermutation
Clinical outcome	Poor	Good
B cell receptor (BCR)		
Structure	Biased Ig repertoire More often stereotyped	Biased Ig repertoire Less often stereotyped
Reactivity	Low affinity poly- or self-reactivity	Oligo- or mono-reactivity
Putative antigen	Neo-autoantigens (from oxidation or apoptosis)	Unknown
BCR signalling	Increased BCR signalling competence Proliferative phenotype	Heterogeneous Anergic phenotype
Genetic lesions	Enriched for high-risk lesions	Enriched for low-risk lesions
Clonal evolution	Higher degree of evolution	Lower degree of evolution

CLL cells also demonstrate autonomous BCR signalling, driven in part by recognition of epitopes within the BCR itself [41]. The end result of activation of the BCR signalling cascade differs between U-CLL and M-CLL cases, and can result in an anergic unstimulated or proliferative B cell [42]: U-CLL cases have lesser down-regulation of surface immunoglobulin (IgM) and increased BCR signalling competence [43, 44]; M-CLL cases have a heterogeneous response to surface IgM stimulation but a predominantly anergic phenotype [42]. The main pathways leading to cell survival and proliferation following BCR activation include the nuclear factor- κ B (NF- κ B) pathway, the MEK-extracellular signal-regulated kinase (ERK) pathway and the mechanistic target of rapamycin (mTOR) pathway [30, 31].

The differing end-result of BCR activation between the two subtypes of CLL likely relates to their differing cells of origin. The U-CLL subtype appears to arise from naïve pre-germinal centre B cells prior to the somatic hypermutation of IGHV that occurs within the germinal centre of lymph nodes upon exposure to antigen as part of the normal immune response [29]. This hypothesis has been supported by the differing gene expression profiles and epigenomes of both subtypes,

however, the exact cell or origin of CLL remains a subject of debate and some hypothesise that it arises earlier in haematopoietic stem cells [29, 45-48].

Therefore, signalling via the BCR in a tonic (antigen-dependent) and constitutive (antigen-independent) manner is critically important in the pathogenesis of CLL with survival and proliferation signals received from the microenvironment integrated by the BCR [49]. The critical nature of BCR signalling in CLL is also evidenced by the remarkable clinical efficacy of signalling pathway inhibitors including ibrutinib and idelalisib (see section 1.6.2).

1.5.2 The role of the microenvironment in disease pathogenesis

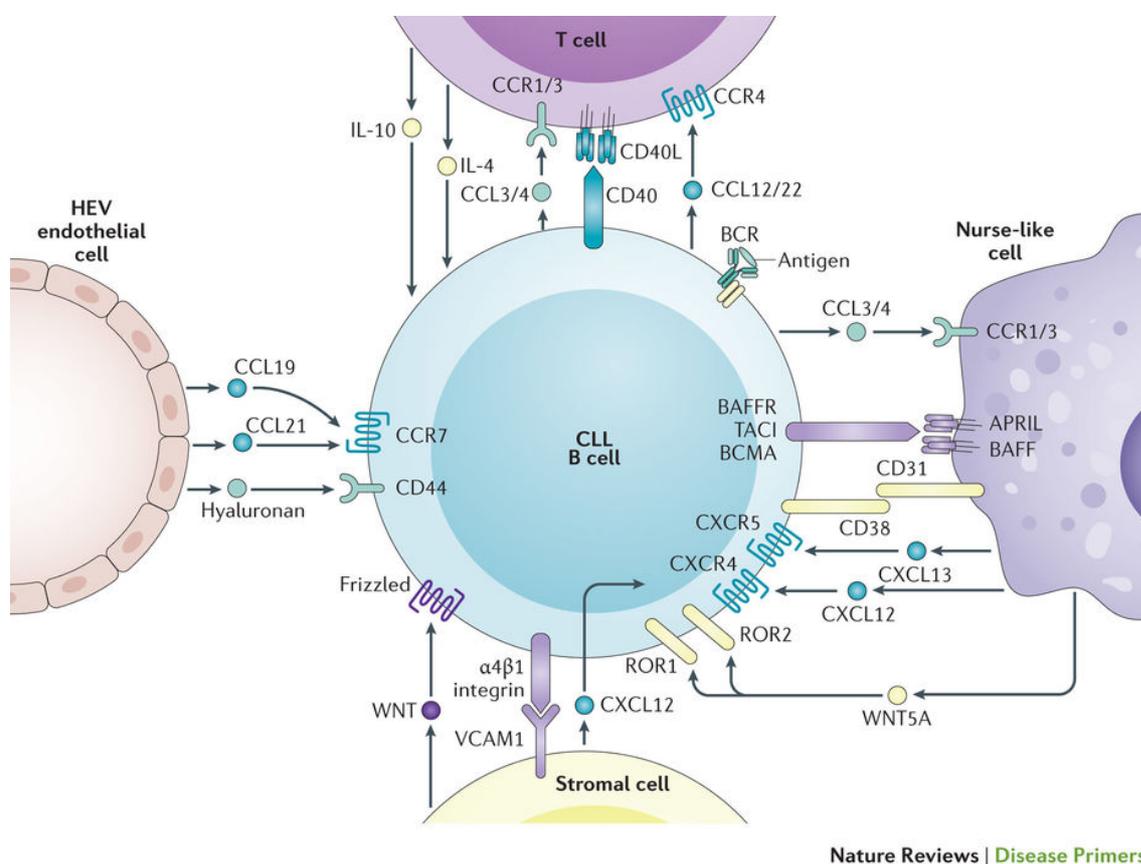
Given the strong evidence for the role of antigenic-drive in CLL pathogenesis, and early observations that CLL cells undergo apoptosis when cultured alone *in vitro*, it has become increasingly clear that the cancer microenvironment is important in CLL cell survival and proliferation [29, 50]. CLL cells circulating in the blood stream enter lymphoid tissues (the “microenvironment”) down chemokine gradients where they receive survival signals from non-malignant cells (such as nurse-like cells, T cells and stromal cells) (see Figure 1-4). CLL is therefore often considered a disease of two compartments (the blood and the tissue microenvironment) with gene expression profiling studies show stronger BCR and NF- κ B signalling from lymph-node derived CLL cells (where they form proliferation centres) as compared to those in the blood [51].

Recruitment of CLL cells to the tissue microenvironment occurs primarily through the interaction between the cell surface receptor, CXC-chemokine receptor 4 (CXCR4), and the chemokine CXC-chemokine ligand 12 (CXCL12). CXCL12 is secreted by nurse-like cells and stromal cells in lymph nodes. Recruitment can also occur via the interaction between CC-chemokine receptor 7 (CCR7) and CC-chemokine ligands 19 and 21 (CCL19 and CCL21) secreted by the endothelial lining of high endothelial venules (HEVs). After recruitment to the lymph nodes, proliferative signals are received by non-malignant cells and include CXCL12, B cell activating factor (BAFF) and tumour necrosis factor ligand superfamily member 13 (known as APRIL), which activate NF- κ B signalling. In tissues, exposure of the BCR to auto- or allo-antigens (as discussed in section 1.5.1) may lead to proliferation and enhance the responsiveness of B cells to microenvironmental signals [30]. Furthermore, it is postulated that the CLL cells themselves may help shape the microenvironment that further supports their survival and proliferation [29].

CD38 is another important player in the interaction between CLL cells and the microenvironment and is variably expressed in CLL [52]. CD38 is a transmembrane protein that functions as an adhesion molecule and receptor in normal B lymphocytes. Interaction with its receptor CD31 (expressed on nurse-like cells in CLL patients) results in remodelling of the CLL-

membrane, enhanced proliferation, up-regulation of the CD100 receptor on proliferating cells and down-modulation of CD72, a negative regulator of immune responses [53]. The interaction of CD100 on CLL cells and plexin-B1 on nurse-like cells also promotes survival and proliferation of CLL cells [54]. Overall, CLL clones with higher number of cells expressing CD38 and more responsive to signalling via the BCR, exhibit enhanced migratory capacity, and reflect an active, proliferative B cell [52].

Figure 1-4. The CLL microenvironment [30]. CLL cells integrate signals from endothelial cells, T cells, nurse-like cells, and stromal cells in the lymphoid tissue microenvironment. See text for details. Reprinted by permission from Springer Nature: Nature Publishing Group. *Nature Reviews Disease Primers*, vol. 3, p.16096 (Chronic lymphocytic leukaemia, Kipps et al.). © Macmillan Publishers Limited, 2017.



Other microenvironmental interactions that support CLL cell survival and proliferation include crosstalk between T cells and stromal cells (see Figure 1-4). T cell support appears to play a role in the pathogenesis of CLL: in an adoptive-transfer xenograft mouse model, disease will only develop in the presence of autologous T cells [55]. T cells are recruited through several chemokines secreted by CLL cells and contribute proliferative signals to CLL cells through the CD40-CD40L (CD40

ligand) axis and secretion of cytokines such as IL-4 and IL-10 [30]. Interactions between CLL cells and mesenchymal-stromal cells also promote cells survival and are mediated by the interaction between CLL cell surface $\alpha 4\beta 1$ integrins (such as CD49d) and stromal vascular cell adhesion protein 1 (VCAM1). CD49d will be discussed in detail in section 1.9.10.1.

In addition to BCR-signalling, other key signalling pathways that are activated by interactions with the microenvironment include the Wnt, Hedgehog, Notch and toll-like signalling pathways [56, 57]. These pathways are implicated in many malignancies, including both solid organ and haematological neoplasms. Recurrent chromosomal aberrations and somatic mutations are also implicated in the pathophysiology of CLL. These will be discussed in section 1.8.

1.6 TREATMENT OF CLL

Not all patients with CLL will require treatment at diagnosis and only those with “active” or symptomatic disease are considered for therapy. The International Workshop on Chronic Lymphocytic Leukaemia has published guidelines with seven indications for treatment [11]. These include evidence of progressive bone marrow failure, massive or symptomatic splenomegaly, massive or symptomatic lymphadenopathy, progressive blood lymphocytosis (with a rapid doubling time), autoimmune cytopenias poorly responsive to standard treatment with corticosteroids, constitutional symptoms, or symptomatic extranodal involvement. A high blood lymphocyte count alone is not a sole indicator for treatment. In those patients not meeting indications for treatment, a “watch and wait” approach is standard of care. In those patients requiring treatment, the choice of therapy is dictated by several factors including disease biology, treatment history, patient fitness for therapy, patient goals, and drug cost and availability (which varies markedly throughout the world). A brief non-exhaustive overview of treatment is presented below.

1.6.1 Chemoimmunotherapy

The treatment paradigm for CLL has evolved rapidly in recent years with the advent of targeted small molecule inhibitors (see section 1.6.2). However, treatment recommendations differ around the world due to the availability and cost of these new agents, and their use in treatment-naïve disease is not yet well established. The gold-stand frontline therapy for fit patients (without deletion of chromosome 17p) is widely accepted to be a chemo-immunotherapy combination known as FCR. This regime includes the two cytotoxic agents fludarabine (F) and cyclophosphamide (C) combined with the anti-CD20 monoclonal antibody, rituximab (R). The CLL08 study documented superiority of the addition of rituximab to the FC backbone and established FCR as the first line treatment of choice for fit patients [58].

Long-term outcomes of the CLL08 trial have been published and demonstrate that prolonged remissions for a subset of treatment naïve patients [59]. The patients who fared best were those randomised to receive FCR (as opposed to FC alone) and who had the favourable prognostic indicator of a mutated IGHV (M-CLL). Of this group, 86.3% were alive at 5 years with a median overall survival (OS) that had not been reached at the time of publication after a median follow up of 5.9 years. Importantly, very few relapses were observed after 7 years suggesting FCR may prove to be a “functional” cure for patients with this disease subtype. Similar outcomes were reported in The University of Texas MD Anderson Cancer Center (MD Anderson) series in which a plateau on the progression-free survival (PFS) curve was observed for M-CLL patients post-FCR with no relapses after 10.4 years [60].

An alternative regimen for patients unfit for FCR due to comorbidity is the combination of the alkylating agent chlorambucil with the glyco-engineered anti-CD20 monoclonal antibody, obinutuzumab. In a study of 781 previously untreated CLL patients, the combination of chlorambucil-obinutuzumab resulted in a median PFS of 26.7 months (compared to 16.3 months with chlorambucil-rituximab and 11.1 months with chlorambucil alone) and prolonged overall survival compared to chlorambucil alone [61].

1.6.2 Small molecule inhibitors

The outcome with FCR for patients with the chromosomal aberration, deletion 17p (see section 1.8.1), is poor with a response rate of 68% and a median progression-free survival (PFS) of 12 months [58]. As such, novel agents (such as ibrutinib) are beginning to replace FCR as upfront therapy for these patients with high-risk disease. Furthermore, second-line treatment of CLL following relapse post-FCR usually involves introduction of a novel agent such as ibrutinib, idelalisib or venetoclax, rather than retreatment with chemoimmunotherapy.

1.6.2.1 Ibrutinib

Ibrutinib is an orally bioavailable, irreversible inhibitor of BTK, an integral component of the B cell receptor signalling pathway. It has shown impressive single-agent efficacy in relapsed and refractory CLL/SLL [62], and also as front-line therapy in patients with deletion 17p where it appears to substantially improve the negative prognostic effect of this high-risk cytogenetic feature [63, 64]. Ibrutinib acts to release CLL cells from the microenvironment (where they receive pro-survival signals) and results in a characteristic worsening of peripheral blood lymphocytosis on initiation which parallels egress of CLL cells from nodal compartments into the bloodstream and a decrease in lymphadenopathy.

1.6.2.2 Idelalisib

Idelalisib is an oral inhibitor of PI3K, another critical component of the B cell receptor signalling pathway upstream of BTK. It has been shown to confer increased PFS and OS to patients with relapsed/refractory CLL when used in combination with rituximab, as opposed to single-agent rituximab [65]. It causes a similar redistribution lymphocytosis to ibrutinib.

1.6.2.3 Venetoclax

Venetoclax is an oral inhibitor of B cell lymphoma 2 (BCL-2), an anti-apoptotic protein constitutively overexpressed in CLL, and directly induces apoptosis of CLL cells [66]. An overall response rate (ORR) of 77% was observed in the first-in-human study of single agent venetoclax in a cohort of relapsed/refractory CLL patients [67]. An equally high ORR of 79.4% was seen in a high-risk cohort of patients with deletion 17p treated with venetoclax alone [68]. Given these impressive results, its use in the upfront setting and in combination with alternative agents is being actively pursued. Recently, the four-year analysis of the randomised Phase III MURANO study comparing venetoclax-rituximab to bendamustine-rituximab in relapsed/refractory CLL showed sustained benefit for the former regime in terms of PFS with a 4-year PFS of 57.3% versus 4.6% after a median follow-up of 22 months [69]. A combination of venetoclax and obinutuzumab has been studied as upfront treatment in patients with untreated CLL and coexisting conditions where it showed a significant improvement in PFS at 24 months (88.2%) compared to the combination of chlorambucil-obinutuzumab (64.1%) [70]. The benefit was also demonstrated in patients with *TP53* mutations and/or U-CLL. Furthermore, achievement of minimal residual disease (MRD) negativity in blood (as determined by an allele-specific oligonucleotide PCR assay with a cut off of 10^{-4}) was more frequent in the venetoclax-obinutuzumab arm compared to the chlorambucil-obinutuzumab arm (75.5% versus 35.2%) at three months post-treatment completion. This was also true in the bone marrow and was consistently observed across all subgroups, including del(17p) CLL.

1.7 PROGNOSIS OF CLL

The prognosis of CLL varies widely and there are several well validated prognostic indicators that are in routine use to aid in predicting disease course. Some of these indicators (such as deletion of chromosome 17p) are also useful predictive markers of response to therapy as mentioned in the preceding section. The prognostic impact of the IGHV gene sequence, expression of two flow cytometric markers (CD38 and ZAP-70) and recurrent genomic aberrations will be discussed.

1.7.1 Prognostic impact of the mutational status of the immunoglobulin heavy variable gene (IGHV)

One of the most powerful independent biologic prognostic indicators in CLL is the mutational status of the IGHV [71, 72]. U-CLL is characterised by a clinically more aggressive course and poorer prognosis [71, 72]. In the landmark study by Hamblin *et al.*, the median OS for patients with U-CLL was significantly worse than M-CLL for patients with early-stage disease (95 versus 293 months respectively; $p=0.0008$) [72]. In the CLL08 trial, U-CLL was predictive of worse outcome to both FC and FCR chemotherapy [58]. In the cohort of patients randomised to FCR, the U-CLL group had a 5-year PFS of 33.1% versus 66.6% in the M-CLL group [59].

Table 1-3. Overview of the four most common stereotyped subsets of CLL. Data has been collated from the study by Jaramillo *et al.* [73]. *frequency refers to % of 1861 patients with advanced -stage CLL (that is, those requiring treatment) in the CLL08, CLL10 and CLL11 clinical trials. TTFT = time to first treatment. † refers to % of patients with the subset that developed Richter’s syndrome.

	Subset #1	Subset #2	Subset #4	Subset #8
Frequency*	2.0%	3.3%	0.6%	0.9%
Clinico-biologic associations	Associated with del(11q)	-	Younger patients	Associated with trisomy 12
Clinical outcome	Poor – similar to U-CLL	Shorter TTFT compared to other M-CLL	Favourable – better than other M-CLL	Poor
Richter’s syndromet†	2.6%	3.3%	0%	6.3%
IGHV gene usage	IGHV1-5-7	IGHV3-21	IGHV4-34	IGHV4-39

Certain stereotyped B cell receptor subsets also have prognostic relevance in CLL (see also section 1.5.1). For example, subset 2 (utilising the IGHV3-21 and IGVL3-21 genes) has a more aggressive prognosis than other M-CLL with a time to first treatment of 22 months. Subset 8 (characterised by IGHV4-39 gene usage and an association with trisomy 12) has a uniformly aggressive course and a 17-fold increase in the relative risk of Richter’s transformation [74, 75]. A summary of the four most common stereotyped subsets in CLL are presented in Table 1-3. The data presented in the table was collated from a study by Jaramillo *et al.* analysing the prognostic impact

of BCR-stereotypes in 2453 patients with early or advanced-stage CLL enrolled in the German CLL Study Group CLL01, CLL08, CLL10 and CLL11 clinical trials [73].

1.7.2 Prognostic impact of CD38 and ZAP-70 expression

CD38 expression is a poor prognostic marker and correlates with IGHV mutational status [71]. Expression of CD38 in $\geq 30\%$ of CLL cells has been defined as CD38-positive (CD38+) disease and portends inferior OS and shorter TTFT. The clinical behaviour may be explained by the biological role of CD38 as an integrator of environmental signals that promote survival of B lymphocytes (see section 1.5.2). One of the disadvantages of the use of CD38 as a prognostic marker is that its level may change over the course of the disease [76].

ZAP-70 is an intracellular protein normally expressed in T lymphocytes that is aberrantly expressed in a subset of CLL and is part of an intracellular signalling pathway activated by CD38. It was originally identified as differentially expressed in U-CLL and M-CLL and has been proposed as a surrogate marker for mutational status of the IGHV [77]: higher levels of ZAP-70 expression are associated with U-CLL and are indicative of worse patient outcome [78, 79]. The technical and logistic limitations of performing this test have made it a less robust assay for most laboratories and therefore it is not widely used.

1.7.3 Prognostic impact of genomic lesions

Approximately 80% of CLL patients will harbour one of four recurrent chromosomal aberrations at diagnosis as detected by fluorescence *in situ* hybridisation (FISH) of interphase nuclei [80]. These abnormalities include deletion of chromosome 13q14 (del(13q)), deletion of chromosome 11q (del(11q)), deletion of chromosome 17 (del(17p)) and trisomy of chromosome 12 (+12). These alterations have established prognostic significance and divide patients into 5 different outcome groups based on their karyotype. The median overall survival of each group observed in the landmark paper by Döhner *et al.* was 133 months (del(13q)), 114 months (+12), 111 months (normal karyotype), 79 months (del(11q)) and 32 months (del(17p)) (see Figure 1-5) [80]. The survival curves largely reflect outcomes with treatment: all patients in the del(17p) and del(11q) arms received treatment during the observation period, over 80% of the normal karyotype and +12 groups were treated, and approximately two-thirds of patients with del(13q) were treated. The type of treatment was not elucidated; however, it is inferred to be chemotherapy as this study was published prior to the introduction of the novel agents. These genomic lesions will be discussed in further detail in the following section.

1.8 RECURRENT GENETIC LESIONS IN CLL

There is no unifying genetic aberration underlying CLL. Along with the four common chromosomal aberrations identified by FISH, there are a small number of recurrently mutated putative driver genes that have been identified across several next-generation sequencing studies in CLL (see Figure 1-6 and section 1.8.2) [81-87]. A longer “tail” of more rarely mutated genes is also identified, clustering in commonly disrupted pathways in CLL such as NF- κ B and NOTCH signalling.

Figure 1-5. Prognostic impact of genomic lesions in CLL [80]. Probability of survival from date of diagnosis is shown in the Kaplan-Meier curve below and separates five different outcome subgroups based on cytogenetic aberrations. See text for details. Reproduced with permission from Döhner et al. [80], Copyright Massachusetts Medical Society.

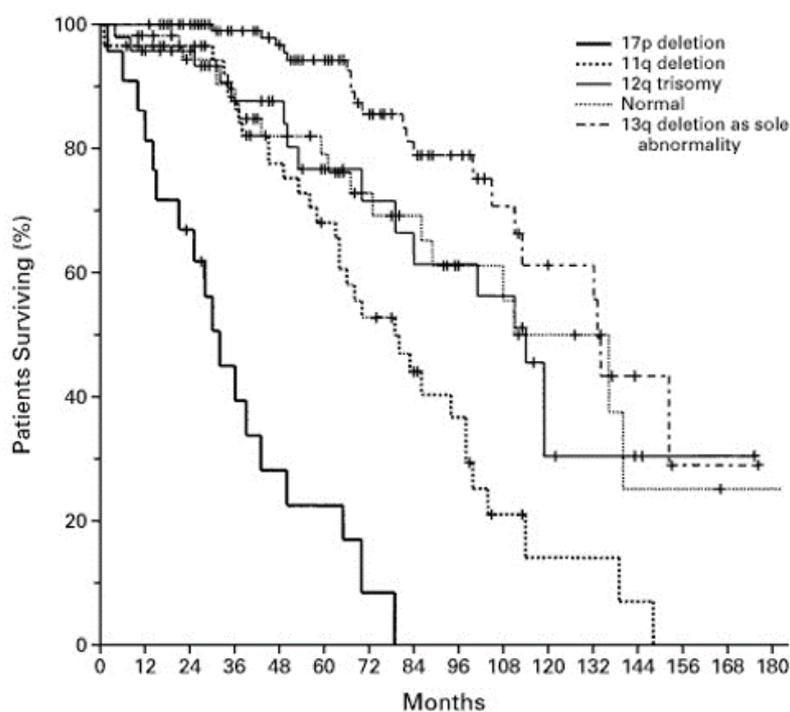
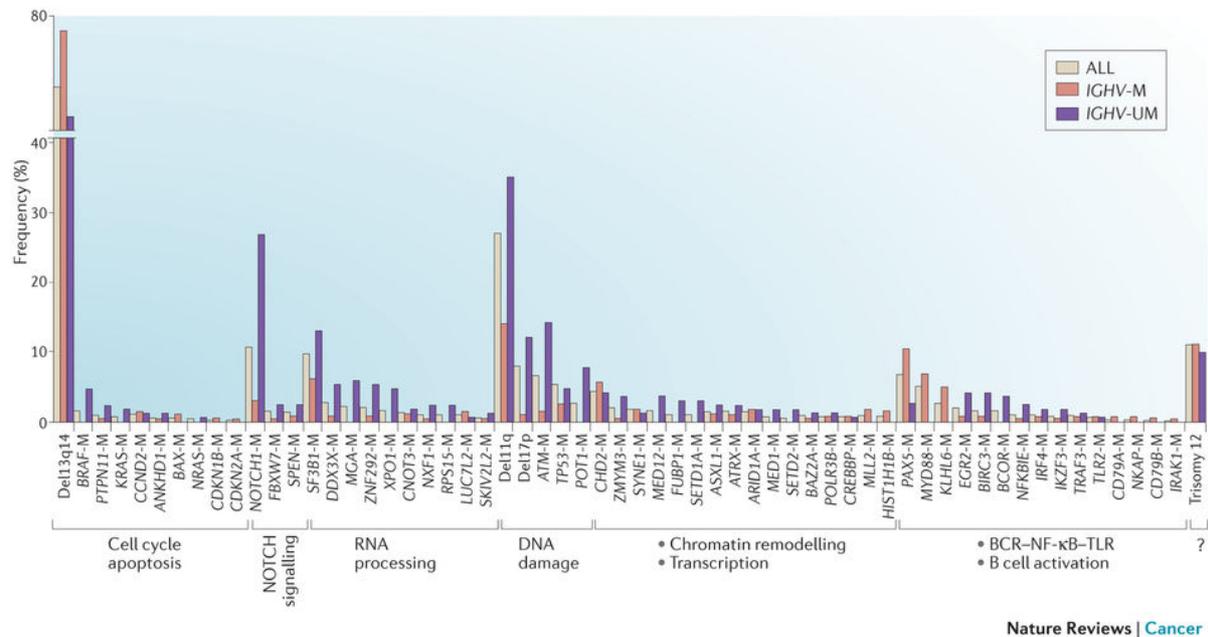


Figure 1-6. Genetic abnormalities in CLL [29]. Recurrent chromosomal and genetic abnormalities detected in CLL through massive-parallel sequencing studies [81, 86-88]. A small number of the same chromosomal aberrations and putative driver genes are identified across studies and cluster in commonly disrupted pathways such as NOTCH signalling. The frequencies of each aberration in M-CLL (IGHV-M) and U-CLL (IGHV-UM) is shown below. Reprinted by permission from Springer Nature: Nature Publishing Group. *Nature Reviews Cancer*, vol. 16, p.145-162 (The molecular pathogenesis of chronic lymphocytic leukaemia, Fabbri et al.) © Macmillan Publishers Limited, 2016.



1.8.1 Recurrent chromosomal aberrations in CLL

Of the four recurrent chromosomal abnormalities in CLL (del(13q), del(11q), del(17p) and trisomy 12), del(13q) is the most common, present in 55% of cases [80]. It is present as a sole cytogenetic aberration in 66% of cases (as interrogated by FISH alone) and is present at high clonal frequency in the majority of cases, suggesting a role early in disease pathogenesis [80, 83]. It is proposed as one of two early evolutionary events in CLL, along with trisomy 12, with which it has a significantly low rate of co-occurrence [83]. Del(13q) CLL has the most favourable outcome with the longest median overall survival of all CLL [80].

The role of del(13q) in disease pathogenesis is well established. Whilst the size of the deletion varies, the minimally deleted region contains the microRNA cluster *miR15a/16-1* and the deleted in lymphocytic leukaemia 2 (*DLEU2*) gene which encodes a regulatory long non-coding RNA [89]. This *DLEU2*/microRNA cluster has tumour suppressive effects, inhibiting expression of proteins involved in cell cycle progression and inhibitors of apoptosis [89].

Del(11q) and del(17p) CLL are features of high-risk disease. Del(11q) is found in approximately 20% of CLL patients at diagnosis, is strongly associated with U-CLL and most often results in deletion of the tumour suppressor gene, ataxia telangiectasia mutated (*ATM*) [29, 90].

Another gene located on the long arm of chromosome 11 is the baculoviral IAP repeat containing 3 (*BIRC3*) gene which is a negative regulator of the non-canonical NF- κ B signalling pathway and may also be disrupted in del(11q) CLL [91]. Del(17p) is infrequent in CLL at diagnosis (present in <10% of cases) but is enriched in chemo-refractory disease and portends a poor prognosis [92-95]. It is strongly associated with U-CLL and results in deletion of the tumour protein P53 (*TP53*) tumour suppressor gene [83, 92].

The molecular pathogenesis of trisomy 12 CLL, the other putative founder lesion in CLL alongside del(13q), is unknown. This is the focus of this thesis and will be expanded upon in detail in section 1.9.

1.8.2 Recurrent somatic mutations in CLL

There are less somatic mutations in the CLL genome compared to most solid tumours and other mature lymphoid malignancies at approximately 0.9 mutations per mega base with 10-30 non-silent events per patient [29]. Different driver mutations have been found to preferentially associate with different chromosomal aberrations suggesting they co-operate to drive leukaemia and reflect different pathogenic mechanisms that may in part explain the clinical variability of the disease. The most frequently mutated genes include *ATM*, *BIRC3*, notch receptor 1 (*NOTCH1*), splicing factor 3b subunit 1 (*SF3B1*) and *TP53*.

Mutations of the *ATM* gene are associated with del(11q), occur almost exclusively in U-CLL, and are associated with a shortened TTFT and chemoresistance [90, 92, 96]. The *ATM* gene is large spanning 146kb of genomic DNA and encodes a protein that is involved in DNA damage repair. The exact role of *ATM* mutations in CLL has been difficult to elucidate given the large size of the gene and the difficulty in distinguishing pathogenic mutations from population polymorphisms.

Mutations of *BIRC3* are associated trisomy 12 and fludarabine-resistant CLL and have recently been demonstrated to predict for progressive disease following upfront FCR independent of *TP53* or IGHV mutational status [91, 97, 98]. Disruption of the *BIRC3* protein leads to constitutive activation of NF- κ B signalling [91].

Mutations of *SF3B1* occur in approximately 10% of cases, are associated with aggressive disease and are predictors for both shortened TTFT and worse OS [98-102]. *SF3B1* encodes a critical component of the cellular splicing machinery and most mutations localise to a conserved C-terminus part of the protein, however, their functional consequences are not yet fully understood [29].

Incompetence of *TP53* via mutation or deletion remains a hallmark of resistant disease, genomic complexity, and overall poor prognosis. Eighty percent of patients with del(17p) harbour a

TP53 mutation on the remaining allele but mutations of *TP53* are predictors of poor prognosis irrespective of the presence of concomitant del(17p) [103-106]. Even small subclonal *TP53* mutations are associated with poor outcome. In a study of 309 patients with untreated CLL, subclonal *TP53* mutations (present in 9% of patients with a median mutated allele frequency of 2.1%) predicted for poor survival, similar to outcomes observed with clonal *TP53* mutations [107].

Mutations of *NOTCH1*, and more rarely the mutated F-box and WD repeat domain containing 7 (*FBXW7*) gene are enriched in trisomy 12 CLL and will be discussed in section 1.9.6.

1.9 TRISOMY 12 IN CLL

Trisomy of chromosome 12 is a recurrent genomic alteration in CLL and defines a subgroup of CLL with distinct biological and clinical features. However, despite its unique phenotype, recurrence, and prognostic significance in CLL, its underlying pathogenesis is unclear and warrants further investigation. This will be the focus of this research. A brief introduction to aneuploidy and trisomy 12 in cancer will be presented, followed by an overview of trisomy 12 CLL, including its significant co-occurrence with mutations of *NOTCH1* and association with CD49d expression.

1.9.1 Aneuploidy in cancer

Alterations in chromosome number (either gains or losses) are found in almost all types of human cancers [108]. It is well established that chromosomal alterations in cancer can occur due to abnormal segregation of chromosomes during mitosis (non-disjunction), and that this occurs at a higher rate than in normal diploid cells [109]. The molecular basis of aneuploidy remains undefined in most cancers; however, it is likely heterogeneous. Alterations in genes involved in the cellular processes of replication, chromosome segregation during mitosis and cell-cycle checkpoints (such as the spindle checkpoint) can lead to chromosomal instability [110]. Furthermore, it has been proposed that aneuploidy itself predisposes to further chromosomal instability, however, this is not universally accepted [109]. Whilst the non-disjunction event leading to aneuploidy initially may be a random event, cancer karyotypes are not completely random and selection for particular aneuploidies may be influenced by the cancer microenvironment [109].

1.9.2 Incidence

Trisomy 12 is variably reported as the second or third most common recurrent chromosomal abnormality in CLL and is present in around 15-20% of patients at diagnosis [80, 111]. Trisomy 12 is also present at a similar frequency in monoclonal B-lymphocytosis [112-114]. There is limited data to determine if its incidence changes with disease progression, but it appears largely stable over time,

at least in the absence of therapy. Using FISH, trisomy 12 was acquired in none of 41 CLL cases over a 4-year period and in only 2 of 77 cases in another series [115, 116].

Trisomy of chromosome 12 is also reported in other haematological neoplasms including the two mature B cell lymphoproliferative disorders, diffuse large B cell lymphoma and follicular lymphoma, where it is present in 10-20% of cases [108, 117]. Renal cancers are the most frequent non-haematological cancers to contain trisomy 12 [108]. Thirty-five percent of papillary renal cell carcinomas and 20% of Wilms tumours (a rare paediatric renal malignancy) contain trisomy 12, however, it does not appear to be associated with a distinct clinical outcome in this setting and the reason underlying the association is not clear [118, 119].

1.9.3 Phenotype

1.9.3.1 Morphology

Trisomy 12 CLL tends to have “atypical” morphology characterised by the presence of increased numbers of lymphocytes with cleaved nuclei or plasmacytoid features [120, 121]. Circulating prolymphocytes are also more prominent than in non-trisomy 12 CLL, in itself a poor prognostic feature [122]. The biology underpinning the presence of circulating prolymphocytes is unclear and it has been postulated that either these cells represent a release into the blood of CLL cells from lymph node proliferative centres that are in the mitotic phase of the cell cycle or they represent a subclone with a survival advantage [123].

1.9.3.2 Immunophenotype

Trisomy 12 CLL has also been consistently demonstrated to have an “atypical immunophenotype” with a modified Matutes score of less than 4, whereas typical CLL has a score of 5 out of 5 [120, 124-127]. The Matutes score is based on the surface expression of 5 markers: CD5, CD23, FMC7, surface immunoglobulin and CD79b. Low CD23 expression correlates with trisomy 12 [128]. Trisomy 12 also has a positive association with CD79b expression (which is usually negative or weakly expressed in CLL) and also is correlated with bright expression of surface immunoglobulin and CD20 (which are both usually dim) [124, 127].

Furthermore, trisomy 12 CLL cases are also more often CD38 positive, defined as the presence of >30% of cells positive for this marker [126, 129, 130]. CD38 positivity is a poor prognostic marker in CLL and is involved in cell homing to lymphoid tissues (the CLL microenvironment) [71, 131]. There is also over-expression of other cell homing molecules (such as the integrin CD49d) that suggest a unique interaction and/or dependence on the microenvironment (see section 1.9.9).

1.9.3.3 Clinical features

Trisomy 12 is more common in SLL than other cytogenetic aberrations [132]. There is a higher incidence of Richter's transformation in patients with trisomy 12 CLL (2%) compared to non-trisomy 12 CLL (0.4%) [126]. There is also a high frequency of trisomy 12 in patients with Richter's syndrome where it is present in approximately one third of cases [10, 126]. Both observations again highlight the importance of the lymph node microenvironment in trisomy 12 CLL, given both diseases have a more nodal phenotype than CLL.

The two largest cohorts published to date specifically exploring the features of trisomy 12 CLL patients in comparison to non-trisomy 12 CLL consist of 250 and 322 cases respectively [126, 130]. Clinical features at presentation were investigated in the first study by Strati *et al.* and found a higher incidence of thrombocytopenia (platelet count $<100 \times 10^9/L$) in the trisomy 12 group compared to 516 normal karyotype CLL controls. No other features (including gender, age, clinical stage, presence of bulky lymphadenopathy or other cytopenias) were significantly different between the groups. Both studies also investigated clinical outcomes which will be discussed in section 1.9.5.

The only other dedicated study of a cohort of trisomy 12 CLL patients was a retrospective analysis of 188 patients by the French Innovative Leukemia Organisation working group, however, there was no non-trisomy 12 comparator arm and clinical features were not the focus of the study [133]. This cohort was subjected to a more detailed molecular analysis with the aim to further refine prognostication and will be discussed in section 1.9.5.2.

1.9.4 Treatment

There is no current alteration to the treatment paradigm for CLL patients with trisomy 12 who are devoid of concurrent *TP53* mutations or del(17p): they are offered the standard upfront chemo-immunotherapy regime, FCR, if fit enough for therapy. There is clear superiority of FCR compared to FC (a regimen lacking the anti-CD20 monoclonal antibody) for trisomy 12 patients on subgroup analysis of the CLL08 trial [58]. Furthermore, along with M-CLL, trisomy 12 independently associates with MRD negativity following FCR in an MD Anderson series [134]. Accordingly, in the Strati *et al.* cohort, overall response rates to first-line treatment (which was mostly FCR; 64% of patients) was very good at 94% with a complete response (CR) rate of 52% [126].

Observational data has demonstrated an association with an increased risk of infusional reactions with the anti-CD20 monoclonal antibody obinutuzumab and both an abbreviated and attenuated redistribution lymphocytosis following commencement of ibrutinib monotherapy [135, 136]. The reasons behind each observation are unclear. It has been postulated that higher CD20 expression levels in trisomy 12 CLL underlie the increased risk of reaction to obinutuzumab and that

an enhanced dependence on microenvironmental interactions in trisomy 12 CLL are responsible for the altered pattern of response to ibrutinib.

In pooled data from three separate clinical trials evaluating ibrutinib treatment, trisomy 12 was the only genomic factor associated with a decreased overall response rate to the drug (84% vs. 93% for non-trisomy 12 CLL, $p=0.03$) but there was a trend to a higher rate of complete response (33% vs. 22%, $p=0.07$) [137]. This did not, however, translate to an increased PFS or OS for trisomy 12 CLL and the reason for the observations are unclear. It is possible that a subset of trisomy 12 CLL is particularly responsive to ibrutinib (accounting for the trend to higher CR), however, the distinguishing features of this putative group of responders are not known. Ibrutinib is not currently recommended as first-line treatment for trisomy 12 patients, although treatment approaches to CLL are rapidly evolving.

Ex vivo drug responses of 184 primary CLL samples have demonstrated that trisomy 12 is an important factor in determining response to inhibitors of the BCR pathway including ibrutinib [138]. Trisomy 12 CLL has increased sensitivity to BTK, PI3K, MTOR, and MEK inhibitors *ex vivo*, suggesting a specific signalling signature leading to altered drug dependencies [138]. This has not yet been borne out in prospective clinical trials and remains a subject of research. It is possible that the observed differences in *ex vivo* and *in vivo* responses are the result of microenvironmental protection from the effects of BTK inhibition in the latter, however, this is unknown and has not been published to date.

Despite good responses to first-line therapy, trisomy 12 CLL is incurable and will invariably relapse (outside of a subset of patients with M-CLL that may have long-term remissions following FCR). There is thus a requirement for re-treatment with diminishing complete responses. The timing and sequence of therapies is not clear, and there is a need for better prognostication and more personalised, targeted approaches to therapy. This is true of all CLL but is typified in the trisomy 12 subgroup of “intermediate” prognosis. So, there is a requirement to better predict which patients in this subgroup will require treatment and to target the critical pathways responsible for leukaemic progression. Understanding the biology of trisomy 12 CLL is therefore of importance to improve prognostication and treatment outcomes for patients.

1.9.5 Prognosis

1.9.5.1 Time to treatment, overall survival and causes of death

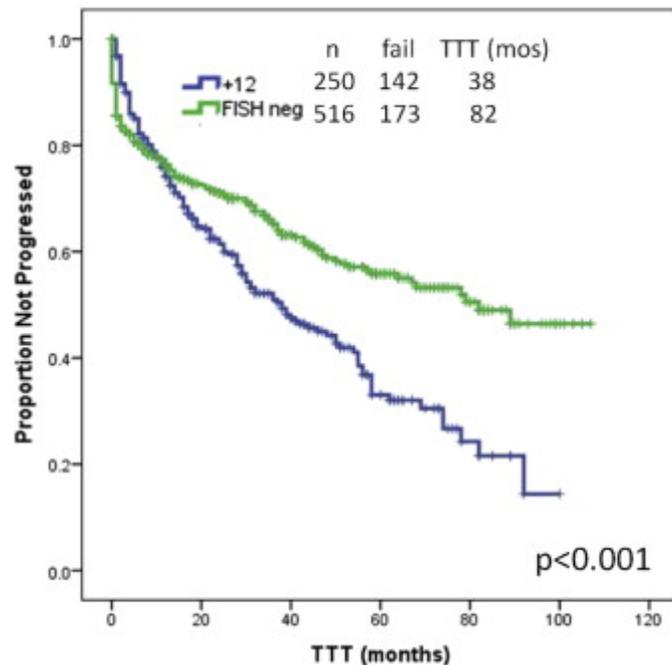
In the two largest trisomy 12 cohorts published to date by Strati *et al.* and Bulian *et al.*, TTFT was 38 months and 3.1 years with median follow ups of 51 months and 3.5 years respectively, both significantly shorter than non-trisomy 12 comparators (excluding del(17p) and del(11q)) [126, 130].

Fifty-seven percent (57%) and 67% of the Strati and Bulian trisomy 12 cohorts respectively required treatment during the period of observation, again significantly more than the comparator arms. The TTFT Kaplan-Meier curve comparing trisomy 12 and non-trisomy 12 CLL patients from the Strati publication is shown in Figure 1-7.

In the Strati cohort, median OS for trisomy 12 patients was not reached with a median follow up of 51 months [126]. Eighteen of the 251 patients (7.2%) had died at the time of publication. In the Bulian cohort, the median OS was 15.6 years (with a median follow up of 3.5y), significantly shorter than the comparator group (non-del(17p), non-del(11q), non-TP53 mutants) [130]. Deaths were observed in 17% of the cohort during the period of observation.

The leading causes of death in patients with trisomy 12 were secondary malignancies and Richter's transformation found in 9% (22/250) and 2% (6/250) of patients respectively after a median follow up of 51 months [126]. Both secondary malignancy and Richter's transformation were also significantly more frequent in the trisomy 12 compared to non-trisomy 12 patients: 9% versus 1% incidence of secondary malignancy and 2% versus 0.4% incidence of Richter's transformation [126]. However, the total number of affected patients overall was small and needs confirmation in a larger series.

Figure 1-7. Kaplan-Meier time to first treatment (TTT) curve comparing trisomy 12 CLL (blue line) to non-trisomy 12, non-del(11q), non-del(17p) CLL (green line) as published by Strati et al. [126] Reprinted by permission from Elsevier. *Clinical Lymphoma Myeloma and Leukemia*, vol. 15, p.420-427 (Second cancers and Richter transformation are the leading causes of death in patients with trisomy 12 chronic lymphocytic leukemia, Strati et al.). © Elsevier, 2015.

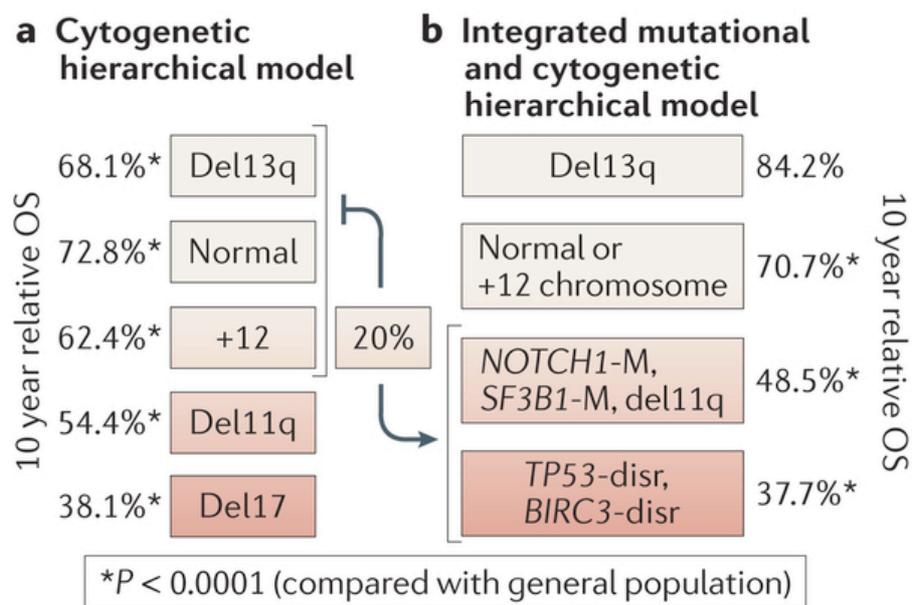


1.9.5.2 Prognostic markers

Several biological markers have refined the traditional “intermediate” prognosis ascribed to trisomy 12 CLL. Firstly, incorporating molecular information leads to more accurate prognostication and a model integrating cytogenetic and molecular data has been proposed (see Figure 1-8) [139]. Patients with trisomy 12 that also harbour somatic *NOTCH1*, *SF3B1*, *TP53* or *BIRC3* mutations have inferior survival compared to those devoid of these mutations, in part explaining the variability of clinical outcomes and refining the “intermediate” prognosis. For example, 10-year relative OS for patients with trisomy 12 alone is around 70%, whereas it declines to 48.5% for those with *NOTCH1* mutations and to 37.7% for those with *BIRC3* or *TP53* mutations [29]. It should be noted, however, that there is some conflicting data about the association of *NOTCH1* mutation with inferior overall survival: in one study of 188 trisomy 12 patients, only *TP53* disruption confirmed worse OS (HR for death 5.0; $P < 0.001$) and *NOTCH1* mutational status did not independently predict for worse OS [133]. Finally, the prognostic impact of the molecular aberrations should be considered in the context of the treatment delivered. The integrated model presented in Figure 1-8 reflects the outcomes of patients receiving a range of chemotherapy-based treatments (ROSSI paper). There is evidence that specific treatments may overcome the poor prognosis of certain high-risk lesions: for example, the

combination of venetoclax-obinutuzumab overcomes the poor prognosis conferred by *TP53* aberrations [70].

Figure 1-8. Recurrent mutations refine prognosis in trisomy 12 CLL. The model below proposed by Rossi et al. reclassifies one fifth of low- and intermediate-risk patients (including those with trisomy 12) into higher risk categories based on concomitant somatic mutations [139]. Reprinted by permission from Springer Nature: Nature Publishing Group. *Nature Reviews Cancer*, vol. 16, p.145-162 (*The molecular pathogenesis of chronic lymphocytic leukaemia*, Fabbri et al.). © Macmillan Publishers Limited, 2016.



Nature Reviews | Cancer

Clonal frequency (that is, the proportion of cells bearing trisomy 12 within an individual patient's tumour) has also been suggested to play a role in prognostication. In a series of 289 patients with trisomy 12, those patients with a higher percentage of trisomy 12 cells (>60% of cells; representing 60% of all cases) had a shorter TTFT (30 vs. 49 months) and OS (96 vs. 159 months) compared to those with clonal frequencies of <60% [140]. This finding could not be recapitulated in the cohort by Bulian et al. or Roos-Weil et al. whereby higher percentages of trisomy 12 nuclei did not correlate with an inferior OS [130, 133].

Additional chromosomal abnormalities have also been linked to prognosis. The presence of additional trisomies (such as trisomy 18 and 19; see section 1.9.6) have consistently been associated with a good prognosis, with a longer time to next treatment and lower relapse risk [133, 141]. The

presence of deletion of the long arm of chromosome 14 (del(14q)) has been associated with an inferior prognosis for unclear reasons [142].

It has been proposed that the single most powerful predictor of prognosis in trisomy 12 CLL is the mutational status of IGHV [143]. Trisomy 12 is not associated with the IGHV mutational status (neither being enriched in U-CLL nor M-CLL) and in the largest published trisomy 12 group to date of 322 patients, the mutational status of IGHV was the sole independent factor that correlated with both OS and TTFT in multivariate analysis [130]. The hazard ratio for death was 2.13 in the trisomy 12 patients with U-CLL compared to those with M-CLL ($p=0.0112$). In the same data set, there was a high frequency of positivity for CD49d and CD38 (>30% of positive cells), and mutations of *NOTCH1* and *BIRC3* in line with previous studies, however, none of these markers were able to stratify trisomy 12 patients in terms of both OS and TTFT. CD49d positivity correlated with shorter TTFT in the trisomy 12 group, but not inferior OS. This was in comparison to the case control cohort in which inferior OS was correlated to CD49d and CD38 positivity, unmutated IGHV, and mutations of *NOTCH1* and *SF3B1*, as per previously published data. The lack of independent association of worse OS or TTFT with mutations of *NOTCH1* or *SF3B1* in the trisomy 12 cohort is unexpected and the authors concede confirmation in a prospective series is required. This strong dependence and “peculiar clinical relevance” of the IGHV mutational status alone to the prognosis of trisomy 12 CLL has been postulated to relate to a stronger dependence on signals from the microenvironment integrated by the B cell receptor, although this has yet to be confirmed [130].

A retrospective analysis of 188 trisomy 12 CLL patients confirmed the prognostic importance of the IGHV mutational status in this subtype of CLL, at least in terms of time to next treatment [133]. Trisomy 12 with U-IGHV was associated with a shorter time to next treatment in multivariate analysis (HR 9.09; $p=0.001$), however, U-IGHV did not retain independent prognostic significance with regards to overall survival in this analysis.

Despite the addition of molecular information to help refine prognostication of trisomy 12 in CLL, it remains a heterogeneous group with some conflicting reports about survival outcomes for some subtypes (such as *NOTCH1* mutants and those with U-IGHV).

1.9.6 Associated genetic lesions

The co-occurrence of trisomy 12 with other chromosomal abnormalities in CLL varies from study to study depending on the methodology used but it is observed in combination with all the other three common genomic aberrations (del(13q), del(11q), del(17p)), none of which are mutually exclusive. Trisomy 12 does, however, have a low rate of co-occurrence and negative correlation with del(13q) in keeping with their hypothesised roles as separate founder lesions [83].

Trisomy 12 is repeatedly observed in association with additional trisomies. Approximately 12% of patients with trisomy 12 will harbour an additional trisomy (usually of chromosome 19) which appears to alter the clinical phenotype and ameliorate the prognosis to some extent for unclear reasons [141]. This does not appear to be a random association: whilst the acquisition of additional trisomies may be a random event (due to chromosomal mis-segregation during mitosis), it is also possible that there is an underlying unknown mechanism predisposing the trisomic 12 CLL cell to additional trisomies (such as, defects in the mitotic spindle). No such mechanism has been previously published. Furthermore, there are likely selective environmental pressures selecting for clones that harbour specific trisomies (such as trisomy 12 and trisomy 19) and resulting in a distinct pattern of clonal evolution. This is also supported by the phenotypic differences observed between CLL cases harbouring trisomies 12 and 19, and trisomy 12 and another non-19 trisomy [144].

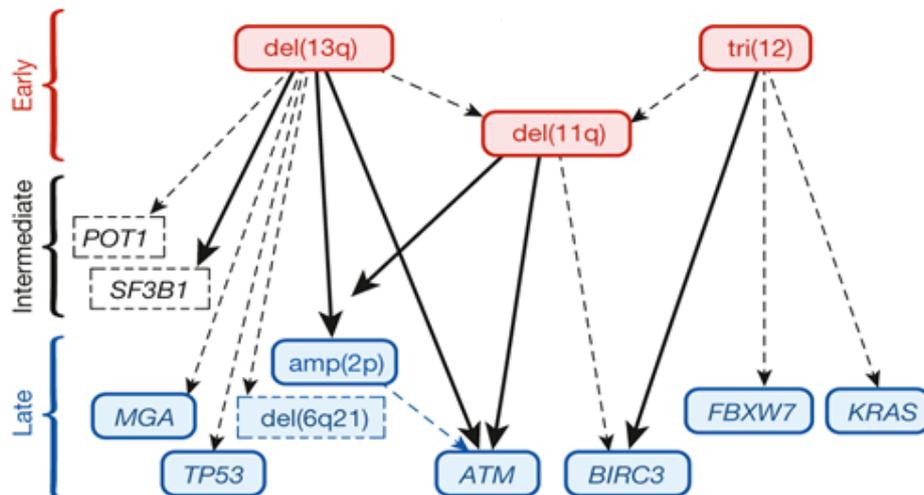
Trisomy 12 is also associated with del(14q). There is also a significant co-occurrence of del(14q) with trisomy 12: 45-47% of del(14q) cases also harbour trisomy 12 and in the majority of these cases, there are no other additional cytogenetic aberrations [145, 146]. Del(14q) is a rare aberration in mature B cell neoplasms (including CLL) present in 1.5% of cases [146]. The reason for the underlying association has not been elucidated but it appears to worsen prognosis [126, 142].

Trisomy 12 also associates with different recurrent somatic mutations as well as the gross chromosomal aberrations mentioned above, suggesting co-operation to drive leukaemia. Mutations of *SF3B1* (involved in the spliceosome) are negatively associated with trisomy 12, and mutations of *BIRC3*, *BCOR*, *FBXW7* and *NOTCH1* are significantly associated with trisomy 12 [83]. The latter two are involved in NOTCH pathway signalling and will be discussed in more detail in section 1.9.11. There is also an association between trisomy 12 and mutations of *NRAS*, *KRAS* and *BRAF*, members of the Ras/MAPK signalling pathway which are not further discussed [147].

1.9.7 Clonal frequency & evolution

The acquisition of trisomy 12 appears to be an early event in the evolution of a patient's CLL. This has been inferred from whole exome sequencing data from 538 patients with CLL [82, 83]. The proportion of the tumour cells (the cancer cell fraction, CCF) that harboured a mutation in any one patient was calculated using the variant allelic fraction, local copy number and purity of the sample. Mutations were then called clonal or subclonal depending on the CCF, and early and late drivers of CLL were inferred from aggregate frequencies. That is, if in most patients, a mutation is clonal, the likelihood that it is an early event in the evolution of CLL is higher. A map of drivers with temporally directed edges was created (see Figure 1-9) and trisomy 12 appeared to be one of two early (potentially founder) events in CLL evolution, along with del(13q).

Figure 1-9. Trisomy 12 is hypothesised to be an early event in the evolution of CLL. A map of the inferred evolution of CLL from analysis of the clonal frequencies of genetic abnormalities in a cohort of 538 patients [82, 83]. Reprinted by permission from Springer Nature: Nature Publishing Group. *Nature*, vol. 526, p.525-530 (*Mutations driving CLL and their evolution in progression and relapse*, Landau et al.). © Macmillan Publishers Limited, 2015.



In line with these findings, the median clonal prevalence of trisomy 12 cells within the cancer cell fraction (CD5/CD19 positive CLL cells) was 94% in the Bulian cohort, again suggesting that trisomy 12 occurs early and is a clonal event in CLL [130]. Furthermore, this hypothesis is supported by the low rate of co-occurrence observed between del(13q) and trisomy 12 in another cohort [82, 83]. In the Bulian cohort, del(13q) was observed within 26% of trisomy 12 patients (it was an isolated chromosomal abnormality in the remaining 74%) but the percentage of nuclei observed with each abnormality was randomly distributed with the same mean range in all patients, that may suggest the independent co-occurrence of these two abnormalities from potentially two different founder clones [130]. The high clonal frequency, however, does not necessarily imply that trisomy 12 is a founder event: it is also possible that its acquisition imparts such a significant survival advantage over other clones, that the trisomic 12 clone quickly expands to represent the majority of the tumour at the time of diagnosis and/or sampling.

The clonal frequency of trisomy 12 does not appear to change over time, although there is some conflicting data. Whole-exome sequencing data from matched pre-treatment and disease relapse samples in 59 patients demonstrate stability of the trisomy 12 clone despite branched evolution [82, 83]. In another study of 41 patients, there was no change in trisomy 12 clonal frequency as detected by FISH over a 4-year period regardless of the presence or absence of clinical disease progression [115]. However, some patients in this study had clonal frequencies of <2%,

lower than the reported sensitivity of routine FISH in most diagnostic laboratories. Other studies have demonstrated an increase in trisomy 12 clonal frequency (as determined by FISH) with progressive disease [148, 149].

1.9.8 Gene expression changes

A specific gene expression signature of trisomy 12 CLL has been established using DNA microarrays and more recently RNAseq [138, 150, 151]. As expected, there is a gene-dosage effect with most overexpressed genes mapping to chromosome 12, however, not all genes on chromosome 12 are uniformly upregulated. There is also a unique profile of gene expression in trisomy 12 CLL compared to the other genomic subgroups which appear to cluster together. Overexpressed genes include Huntingtin interacting protein-1-related (*HIP1R*) and Myogenic factor 6, herculin (*MYF6*) [150] but neither of these genes have been explored further and are not known to be involved in B cell lymphoproliferative disorders. Other groups have demonstrated overexpression of insulin growth factor 1 receptor (*IGF1R*) [152], glioma-associated oncoprotein 1 (*GLI1*), and protein patched homolog 1 (*PTCH1*) [153] in trisomy 12 CLL in more targeted analyses. The latter two are key components of the Hedgehog signalling pathway and *GLI1* maps to chromosome 12. Recent RNAseq data has been published with a gene set enrichment analysis for differentially expressed genes in trisomy 12 CLL. Significant enrichment in genes in the PI3K-AKT-mTOR, chemokine and BCR signalling pathways, and genes involved in regulation of the actin cytoskeleton was observed [138]. Other key signalling pathways that are upregulated in trisomy 12 CLL including integrin signalling (of which integrin subunit alpha 4, *ITGA4*, the gene encoding CD49d, is a member) and NFAT (Nuclear Factor of Activated T cells) signalling.

Despite these findings, the critical genes that lead to and/or propagate leukaemogenesis in trisomy 12 CLL remain largely unexplored and there are clearly other mechanisms apart from the gene-dosage effect at play, as several of the overexpressed genes do not map to chromosome 12. Important pathways that are likely involved in pathogenesis including BCR signalling, integrin signalling/microenvironmental interactions and NOTCH signalling. These will each be discussed in the following sections.

1.9.9 B cell receptor signalling

Putative antigens that may play a role in trisomy 12 CLL and downstream pathways of BCR activation have not been studied specifically in the trisomy 12 subgroup. Although, the mutational status of IGHV in the trisomy 12 cohort is not skewed to either subtype (approximately half of patients with trisomy 12 CLL have an unmutated configuration of IGHV in several series), certain BCR stereotypes are over-represented. Sixty percent (60%) of patients with BCRs that belong to subset 8 have trisomy

12 and U-CLL, significantly greater than any other subset [36]. This subset has an extremely poor prognosis and increased risk of Richter's transformation and fits with the high incidence of trisomy 12 in Richter's syndrome [10, 74]. In addition to this, patients with trisomy 12 and M-CLL are over-represented in subset 201, utilising IGHV4-34 [36]. The implication of these findings is unknown but may suggest that factors specific to trisomy 12 CLL give rise to clones that respond to certain (unknown) antigenic pressures.

1.9.10 Integrin signalling and interaction with the microenvironment

The ability for CLL cells to migrate to the lymph node and bone marrow microenvironments is important in pathogenesis as it is here that the cells receive pro-survival signals and can be protected from chemotherapy. Several findings suggest that trisomy 12 CLL is particularly dependent on microenvironmental signals. These include the over-representation of trisomy 12 in SLL and Richter's transformation, the enhanced CR rate and abbreviated classical redistribution lymphocytosis with ibrutinib (a disruptor of key microenvironmental pathways), and the independent prognostic importance of the mutational status of IGHV. The mechanisms that underlie this apparent dependence are unclear, however, key regulators of CLL cells homing to the lymph node microenvironment are more often expressed in trisomy 12 CLL. The most well described of these regulators is the cell surface molecule, CD49d.

1.9.10.1 CD49d in trisomy 12 CLL

CD49d is an integrin involved in the leukocyte adhesion cascade, a process that controls leukocyte migration from the blood to lymphoid tissues. It is encoded by *ITGA4*. Integrins are heterodimeric cell surface transmembrane proteins involved in the inducible adhesion of leukocytes to the vascular wall and include CD11a/CD18, CD11b/CD18, CD49d/CD29 and CD49d/ITGB7 [129]. CD49d is the alpha chain of the integrin heterodimer named Very Late Antigen 4 (VLA-4) that interacts with vascular cell adhesion molecule 1 (VCAM1) expressed on endothelial cells and is highly expressed on normal B lymphocytes [154].

CD49d expression is not limited to trisomy 12 CLL, however, a higher proportion of trisomy 12 CLL cases demonstrate CD49d expression compared to non-trisomy 12 CLL. In one of the initial studies of CD49d expression in CLL, 39% of a diverse cohort of 1200 CLL patients demonstrated increased levels of expression of CD49d as investigated by flow cytometry using a 30% cut-off (CD49d-positive) [155]. Increased expression of CD49d was present in most cases of trisomy 12 CLL (89.4%, 184/206 cases) and the percentage of cases with increased CD49d was significantly higher in trisomy 12 compared to any other cytogenetic category. Furthermore, amongst CD49d-positive cases, those with trisomy 12 expressed CD49d at the highest levels as determined by the mean

fluorescence intensity. There were no differences in the proportion of cases with unmutated IGHV or *NOTCH1* mutations between the CD49d-positive or CD49d-negative trisomy 12 cases. However, a later study has demonstrated a positive association with *NOTCH1* mutations and CD49d expression and postulated regulation of CD49d expression via NOTCH1 pathway activation [156].

The percentage of CD49d-positive cells weakly correlates with *ITGA4* mRNA levels ($r^2=0.6$), possibly alluding to regulation at the post-transcriptional level [155]. *ITGA4* is the gene encoding CD49d and is located on chromosome 2. Regulation of *ITGA4* itself has not been fully elucidated – it has been shown to be hypomethylated in CD49d-positive trisomy 12 CLL with the level of methylation being inversely correlated to CD49d expression [155]. The biological mechanism underpinning the relationship between trisomy 12 and *ITGA4* hypomethylation has not been established, and it has been hypothesised that differentially expressed genes in trisomy 12 are involved in DNA methylation and chromatin remodelling [155].

Increased expression of CD49d on circulating malignant cells in trisomy 12 CLL has been confirmed in later studies [129, 157]. The functional consequence of increased CD49d expression includes upregulation of genes of downstream effector molecules (RAS guanyl releasing protein 2, *CALDAG-GEFI*, and *RAP1B* located on chromosome 12), increased ligand (VCAM-1) binding and enhanced VLA-4 directed adhesion and motility [129, 158]. There is also a dependence on an alternative signalling pathway in trisomy 12 CLL (that is even more pronounced in CD49d-positive cases): the CCR7-CCL21 axis as opposed to the CXCR4-CXCL12 axis [158].

The reason for over-expression of CD49d in trisomy 12 has not been fully resolved, however, interferon regulatory factor 4 (*IRF4*) and Ikaros also appear to play a role. *IRF4* encodes a transcription factor that is B cell differentiation stage specific and is involved in cell fate. In a study of 223 treatment naïve patients with CLL, *IRF4* expression was significantly reduced in CD49d-positive cases, of which trisomy 12 CLL (39 cases) had the lowest expression of all subsets [159]. Furthermore, it was shown that low levels of *IRF4* (particularly in trisomy 12 CLL) resulted in increased expression of Ikaros (a target gene of *IRF4*) which in turn increases CD49d expression and enhances ability to bind VCAM1.

Clinically, increased expression of CD49d (using a 30% cut-off) is associated with shorter time to first treatment and associates with other biomarkers of aggressive disease including increased LDH and beta-2-microglobulin, ZAP-70 and CD38 expression, unmutated IGHV, and *NOTCH1* and *SF3B1* mutations [157]. There is also an association with a more lymphoma-like presentation (with lower lymphocyte counts and more lymphadenopathy) and an over-representation in SLL, consistent with its role in cell homing to lymph nodes [157, 160]. CD49d

expression is also a predictor of shorter PFS in patients treated with ibrutinib and correlates with reduced redistribution lymphocytosis and an inferior lymph node response (independent of trisomy 12 status) [161]. Positivity for CD49d is also an independent prognosticator of worse overall survival (hazard ratio for death = 1.88, $P < 0.0001$), even after accounting for *BIRC3*, *IGHV*, *NOTCH1*, *SF3B1* and *TP53* mutational status [162]. This is a surprising feature given that normal B cells express CD49d at high levels [5, 154].

The observed changes in integrin signalling may explain at least some of the clinical differences in trisomy 12 CLL with the shift from the leukaemic phase to a more lymphoma-like phase and altered pattern of response to ibrutinib. It is also clear that other mechanisms are at play in the pathogenesis of trisomy 12 CLL as this group does not have a uniformly poor outcome, as would be expected with the majority overexpressing CD49d, a marker of poor prognosis and modulator of CLL cell migration to the protected microenvironment. In addition to this, it appears that *NOTCH1* and *IRF4* may have roles in regulating integrin signalling through largely unknown mechanisms, and that epigenetic regulation of the *ITGA4* promoter is also involved in pathogenesis via an unclear mechanism.

1.9.11 NOTCH signalling & mutations in trisomy 12 CLL

NOTCH1 mutations are strongly associated with trisomy 12 suggesting a functional synergy [111, 163-166]. Between 30-45% of trisomy 12 cases also harbour mutations of *NOTCH1* and the association is strongest in trisomy 12 cases with an unmutated *IGHV* [163, 165, 166]. Most mutations in *NOTCH1* in CLL occur in the final exon of the gene, exon 34, leading to disruption of the regulatory PEST domain of the protein. The end-result is a more stable *NOTCH1* protein that resists degradation and sustains activation of the *NOTCH* signalling pathway [167]. *NOTCH1* is a cell surface receptor involved in development, cell differentiation and proliferation processes (reviewed in [168]). Upregulation of *NOTCH* signalling is important in all subtypes of CLL and is observed in both wild-type and mutant *NOTCH1* cases [169, 170].

The prototypical mutation in approximately 75% of cases is a 2-base pair deletion in exon 34 (c. 7544_7545delCT), resulting in a frameshift and premature stop codon (p.P2514Rfs*4) [166]. Mutations of *FBXW7*, a negative regulator of *NOTCH1*, are rare (2.5% of untreated patients) and are also associated with trisomy 12, supporting the role of altered *NOTCH* signalling in the pathogenesis of this subgroup of CLL [86, 111].

Clinically, *NOTCH1* mutations are predictors of decreased overall survival in a number of series but not all [111, 165, 166, 171]. The presence of *NOTCH1* mutations in cases with trisomy 12 appears to refine the “intermediate” prognosis traditionally ascribed to these cases (see Figure 1-8).

Patients with CLL and *NOTCH1* mutations have also been associated with resistance to the anti-CD20 monoclonal antibody rituximab in the CLL08 trial [172]. A recent study suggests that this is likely due to lower levels of CD20 (the target of rituximab) in these cases [173].

In summary, NOTCH signalling is important in the disease biology of CLL, and the microenvironment is critical in activating the pathway. *NOTCH1* mutations are enriched in trisomy 12 CLL and have a stabilising effect on the signalling pathway leading to a plethora of downstream events that are still being elucidated, but include regulation of CD20, CD49d and the cell cycle [156, 165, 167, 173, 174].

1.10 CONCLUSION

Trisomy 12 is a recurrent chromosomal abnormality in CLL with a unique clinical and molecular phenotype. There is a strong correlation between trisomy 12 and mutations of *NOTCH1*, a critical player in CLL oncogenesis. There is also a strong dependence on microenvironmental signals as evidenced by the overexpression of the homing integrin CD49d, over-representation in SLL, the strong prognostic relevance of the mutational status of IGHV, and the altered clinical response to ibrutinib. Despite this unique phenotype and although trisomy 12 is identified recurrently in CLL and has clear prognostic significance, the underlying pathogenic mechanisms through which it contributes to CLL leukaemogenesis are unknown. The unfolding new evidence regarding differential response to ibrutinib compared to other CLL subtypes has also yet to be explored and may be of immediate clinical utility.

Presently, trisomy 12 is hypothesised to be an early event in the evolution of CLL given its high clonal frequency in most cases but no key oncogenic or regulatory mechanisms have been discovered to account for its recurrent role in CLL. Furthermore, the interaction between trisomy 12, NOTCH signalling and CD49d expression has not previously been explored in depth and understanding these relationships may lead to a greater understanding of pathogenesis of trisomy 12 in CLL. In conclusion, trisomy 12 CLL is poorly understood and further research into this subtype of this yet incurable disease has the potential to provide insights into novel biomarkers and therapeutic targets, as well as key biological pathways leading to cancer.

1.11 HYPOTHESIS & AIMS

1.11.1 Hypothesis

There are critical factors upregulated on chromosome 12 that lead to activation of downstream pathways responsible for leukaemogenesis and the positive selection of trisomy 12 clones which respond to specific environmental signals integrated by the B cell receptor.

1.11.2 Major Aim

1. To investigate the pathogenesis of trisomy 12 in CLL

1.11.3 Specific Aims

1. To characterise the local cohort of trisomy 12 CLL patients (see Chapter 3)
2. To identify differentially expressed pathways in trisomy 12 CLL (see Chapters 4, 5, 6)
3. To investigate the relationship between CD49d expression and trisomy 12 (see Chapter 5)

2 GENERAL MATERIALS AND METHODS

The following chapter outlines general materials and methods employed throughout this thesis. Specific methods are presented later in their relevant chapters.

2.1 BUFFERS

Phosphate buffered saline (PBS) was made by diluting one PBS tablet (Medicago AB, Sweden) in 1L of deionised water. The 50:50 mix comprised RPMI-1640 media (Life Technologies Australia Pty Ltd.) and fetal calf serum (Life Technologies Australia Pty Ltd.) in a 1:1 ratio. The cryopreservation “freezing mix” comprised 30% dimethyl sulfoxide (DMSO; Fisher Scientific) and 70% RPMI-1640 (Life Technologies Australia Pty Ltd.). The flow cytometry wash buffer comprised PBS with 1% fetal calf serum (Life Technologies Australia Pty Ltd.) and 0.02% sodium azide (Sigma-Aldrich).

2.2 PATIENT SAMPLES

2.2.1 Ethics & sample sources

Primary CLL samples were obtained from three sources: the local Flinders Medical Centre tissue bank, the South Australian Research Cancer Biobank (SACRB) and the CLL06 Australia Leukaemia and Lymphoma group (ALLG) national clinical trial (stored at Flinders Medical Centre). This study was approved by the Southern Adelaide Clinical Human Research Ethics Committee, SACHREC EC00188, approval number OFR # 216.056. Details of all patients presented in this thesis are included in Table 9-5, Table 9-6, Table 9-7, and Table 9-8 of the Appendix.

Three healthy controls were also utilised (see Table 9-5): B cells were obtained by Dr Lauren Thurgood by CD19+ bead enrichment of peripheral blood mononuclear cell (PBMC) preparations from whole blood venesection samples. Venesections were performed therapeutically for hereditary haemochromatosis or Polycythaemia Vera. The patients had no known B cell disorders at the time of venesection and were used in lieu of controls without any haematological disorders as large volume samples were readily available. There is no current evidence that B lymphocytes in either condition are abnormal. Collection of samples was approved by the Southern Adelaide Clinical Human Research Ethics Committee (SACHREC, project number 237896).

2.2.2 Sample storage

Peripheral blood or bone marrow mononuclear cells (PBMCs or BMMCs) were isolated from fresh peripheral blood or bone marrow by the Ficoll-Hypaque density gradient method using sterile

technique. In brief, blood or bone marrow was diluted 1:1 in PBS and layered over 10mL of cold Lymphoprep™ (Axis-Shield, Oslo, Norway). The buffy coat was collected using a Pasteur pipette following centrifugation at 800 x *g* for 20-30 minutes. Mononuclear cells were washed in PBS, centrifuged at 300 x *g* for 5min and resuspended in 5mL of prewarmed 50:50 mix (see buffers). The cell count was determined using a 0.4% w/v trypan blue (Bio-Rad Laboratories, Hercules, CA, USA) exclusion assay on the Bio-Rad Laboratories TC20™ automated cell counter. Cells were diluted to the desired concentration in 50:50 mix and then cryopreserved with the addition of 1mL of freezing mix (see buffers) per 1mL of cell suspension in a controlled rate freezing container. Each cryovial contained 4×10^7 cells as standard and was placed in liquid nitrogen for long term storage following short term storage at -80°C for 1-30 days. Additionally, prior to dilution in freezing mix, aliquots of the cell suspension were washed again in PBS, pelleted, and resuspended in TRI Reagent® (Sigma-Aldrich) or TRIzol™ (ThermoFisher). Samples were incubated at room temperature for at least 5 minutes and then stored at -80°C until required.

2.3 CELL CULTURE

2.3.1 Cell thawing

Cryopreserved primary cells were thawed by adding 10mL supplemented media (see section 2.3.2) pre-warmed to 37°C in a dropwise manner. Cells were then pelleted (300 x *g* for 5 min), washed in 10mL of PBS or RPMI-1640 media (Life Technologies Australia Pty Ltd.), re-pelleted (300x *g* for 5 min) and resuspended in supplemented media at the desired concentration. Cell viability and number were determined using a 0.4% w/v trypan-blue (Bio-Rad Laboratories, Hercules, CA, USA) exclusion assay on the Bio-Rad Laboratories TC20™ automated cell counter unless otherwise stated.

2.3.2 Primary cell culture

Cells were maintained in RPMI-1640 media (Life Technologies Australia Pty Ltd.) supplemented with 2mM L-glutamine (Sigma-Aldrich), 1% penicillin-streptomycin (Sigma-Aldrich) and 10% fetal calf serum (FCS; Life Technologies Australia Pty Ltd.) in a 37°C 5% CO₂ incubator. Aseptic technique was employed, and cells were handled in a class II Biosafety cabinet.

2.4 DNA EXTRACTION

Genomic DNA was extracted from total cell suspensions using the Qiagen™ DNeasy® Blood & Tissue Kit (Cat. No 69504) as per the manufacturer's recommendations, eluted in 50-200µL of AE buffer or H₂O and stored at -20°C. The CD5/CD19+ CLL cell fraction was not sorted prior to DNA extraction due

to quantity and availability of samples. DNA was quantitated using the NanoDrop™ 2000 spectrophotometer (ThermoFisher Scientific) or the Qubit fluorometer (ThermoFisher Scientific).

2.5 RNA EXTRACTION

RNA extractions were performed using a TRI Reagent® (Sigma-Aldrich) separation method and carried out with the use of RNase Zap (ThermoFisher) on the working surface and gloved hands. RNA was extracted from thawed cryopreserved PBMC preparations. 1mL of TRI Reagent® (Sigma-Aldrich) or TRIzol™ (ThermoFisher) was added to cell preparations, mixed, and incubated at room temperature for 5 minutes before storing at -70°C until required. Due to variable RNA quality following thawing of samples, RNA was also extracted from PBMCs that had been directly stored in TRI Reagent® (Sigma-Aldrich) or TRIzol™ (ThermoFisher) after blood collection (if available). Samples were then thawed and an additional 1mL of TRI Reagent® (Sigma-Aldrich) or TRIzol™ (ThermoFisher) was added. 400µL of chloroform was added, the samples were gently shaken and incubated at room temperature for 5 minutes. They were then centrifuged at 3200 x *g* for 10 minutes at 4°C and the clear top layer containing the RNA was transferred to a clean falcon tube. The RNA was precipitated with 1mL of isopropanol, inverted and incubated at room temperature for 10 minutes. The RNA was pelleted at 3200 x *g* for 40 minutes at 4°C. The pellet was then washed in ice cold 75% ethanol and centrifuged again at 3200 x *g* (20 minutes, 4°C). The pellet was air dried and resuspended in 30µL of sterile water. The samples were quantitated using a NanoDrop™ (ThermoFisher Scientific) spectrophotometer and stored at -70°C until required. RNA quality was assessed using the Agilent 2100 Bioanalyzer according to the manufacturer's instructions. Only RNA samples with an RNA integrity number (RIN) of ≥8.0 were utilised.

2.6 COMPLEMENTARY DNA GENERATION

Complementary DNA (cDNA) was generated from RNA samples as follows: 9µL of RNA (maximum of 1µg) was incubated with 2µL of 50ng/µL random hexamers and 1µL of 10mM dNTPs (Invitrogen™, ThermoFisher Scientific) for 5 minutes at 65°C. Samples were incubated on ice for 2 minutes, and then 4µL of 5xbuffer (Invitrogen™, ThermoFisher Scientific), 2µL of 0.1M DTT and 1µL of diethyl pyrocarbonate (DEPC) H₂O (Invitrogen™, ThermoFisher Scientific) were added. Samples were then incubated at 25°C for 2 minutes and 1µL of Superscript IV reverse transcriptase (Invitrogen™, ThermoFisher Scientific) was added. Samples were then cycled at 25°C for 10 minutes, 42°C for 50 minutes and 70°C for 15 minutes. Samples were diluted with 30-100µL of DEPC H₂O and stored at -20°C prior to use.

2.7 PRIMERS

All primers were ordered from Integrated DNA Technologies (IDT®, Iowa, USA), resuspended in FG3 buffer (Qiagen™) or sterile H₂O to a final concentration of 100μM and stored at -20°C in aliquots. 6.25μM or 10μM working stock solutions of primers were made by adding 6.25μL or 10 μL of each of the forward and reverse primers to 87.5μL or 100μL of FG3 buffer or sterile H₂O respectively. A full list of primer sequences is presented in the Appendix (see section 9.1). 6.25μM working stock solutions were used for standard PCR and the 10μM solutions were used for qRT-PCR.

2.8 PCR

Each PCR reaction contained 20-100ng of DNA template, 2.5μL of 10xPlatinum buffer, 0.75μL of 50mM MgCl₂, 5μL of 4mM dNTPs, 2μL of 6.25μM primer working stock, 0.2μL of 5U/μL Platinum Taq DNA polymerase and were made to a final volume of 25μL with sterile water. All products were supplied by Invitrogen™, ThermoFisher Scientific. Standard PCR cycling conditions were as follows: 96°C for 5 minutes, then 35 cycles of 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 30 seconds, and finished with 10 minutes extension at 72°C. The annealing temperature for all PCR primers was 58°C, except for the *NOTCH1* product which had an annealing temperature of 60°C. PCR products were visualised on a 1.5-2% agarose gel.

2.9 SANGER SEQUENCING

PCR products were cleaned up prior to sequencing with the addition of 0.25μL of exonuclease (NEB Biolabs), 1μL of shrimp alkaline phosphatase (USB®, Affymetrix) and 1.25μL of shrimp alkaline phosphatase buffer (USB®, Affymetrix) per 5μL of PCR product. Reactions were incubated at 37°C for 60 minutes, 80°C for 20 minutes and then 25°C for 2 minutes, then diluted in 20-50μL of sterile water. Sanger sequencing of the diluted product was performed at the Flinders Sequencing Facility (Department of Genetics & Molecular Pathology, SA Pathology, Adelaide, South Australia). Sequencing results were visualised using SnapGene® software (versions 4.0.7 and 5.0.7).

2.10 REAL-TIME QUANTITATIVE PCR

10μL real-time quantitative PCR (qRT-PCR) reactions were set up as follows: 1μL of 10μM forward and reverse primers, 5μL of 2xSybr Green PCR master mix (Applied Biosystems), 1μL of cDNA template and 3L of H₂O. Primers were optimised on universal cDNA (generated from universal RNA, Invitrogen), or FH9 or Raji cell line total cDNA. Primers were utilised if a single product was amplified on the melt curve and serial dilutions yielded a standard curve with a slope of -3.2 (±0.3) and a r² value of ≥0.99. Each sample was run in triplicate with reverse transcriptase (RT) negative controls.

PCR was performed on the Applied Biosystems Viaa7 instrument with the following conditions: ramp to 50°C at 1.6°C/s, hold for 2 minutes, ramp to 95°C at 1.6°C/s, hold for 10 minutes, 40 cycles of 95°C x 15 seconds and 60°C x 60 seconds. Melt curves were performed at the end of the PCR. Gene expression was normalised to the housekeeping gene *GUSB* and relative expression was determined using the following equation: $2^{-(\text{target gene expression} - \text{GUSB expression})}$. Any technical replicate that crossed a threshold of ≥ 0.5 cycles earlier or later than the other two replicates was excluded from analysis. Any sample in which non-specific products were amplified on the melt curve were also excluded from analysis.

3 CHARACTERISATION OF THE LOCAL COHORT OF TRISOMY 12 CLL

3.1 INTRODUCTION

As outlined in the introduction, trisomy 12 CLL has a unique phenotype characterised by an increased frequency of *NOTCH1* mutations and increased expression of the homing integrin CD49d compared to other subtypes of CLL. These findings have been reported numerous times, and two large international cohorts have also described the clinical outcomes of trisomy 12 CLL [126, 130]. The first specific aim of this thesis was to comprehensively characterise the local cohort of patients with trisomy 12 CLL and obtain a complete clinical, genetic and immunophenotypic profile of the group.

3.2 METHODS

3.2.1 Patient identification, selection & cytogenetic analysis

Sixty patients with trisomy 12 CLL were identified through the centralised diagnostic cytogenetic laboratory database by Ms Sarah Moore (SA Pathology, Adelaide) and are listed in Table 9-5. This included all patients with CLL in South Australia and the Northern Territory that had undergone cytogenetic analysis (by any method) as requested by their treating physician from June 2012 until June 2017. Methods used to identify trisomy 12 included conventional karyotype by chromosomal banding analysis (CBA), fluorescence *in situ* hybridisation (FISH), and single nucleotide polymorphism (SNP) microarray (implemented from August 2014). The probes used for FISH were directed to 13q14.3, centromere 12 (12p11.1-q11), 11q21.3 and 17p13.1, and the IGH locus on chromosome 14 in a subset of patients.

A subset of the identified patients (n=15) had available cryopreserved PBMCs and had previously consented to tissue banking (either through the Department of Haematology at Flinders Medical Centre or the South Australian Cancer Research Biobank) (patients 1-8, and 16-22). A further 7 samples were obtained from the on-site tissue bank of Australian CLL patients enrolled on the Australasian Leukaemia & Lymphoma Group (ALLG) CLL06 clinical trial (patients 9-15). This resulted in a final cohort of 22 trisomy 12 CLL patients with available samples for prospective genetic and immunophenotypic characterisation (patients 1-22). Five patients with normal karyotype/disomy 12 CLL were chosen randomly as comparators from samples stored locally in 2016 (DIS1-5) but DIS2 was later excluded as it did not meet the diagnostic criteria for CLL. Patients without available samples underwent audit of clinical and cytogenetic data alone (patients 23-67).

This study was approved by the Southern Adelaide Clinical Human Research Ethics Committee, SAC HREC EC00188, approval number OFR # 216.056.

The initial analysis was performed in 2017 and 2018. Six further cases of trisomy 12 CLL (patients 68-73) were identified in the Flinders Medical Centre tissue bank in February 2020 and were included here in the CD49d expression analysis only. A further 9 samples of disomy 12 CLL (DIS6-15) were also added in February 2020 and are also used in the CD49d expression analysis. DIS6 and DIS8 were excluded after audit as they were post-treatment remission samples.

Details of all patients presented in this thesis are included in Table 9-5, Table 9-6, Table 9-7, and Table 9-8 of the Appendix.

3.2.2 Clinical annotation

Clinical information and diagnostic cytogenetic results were obtained from case-note review and from interrogation of the electronic OACIS Clinical Information System for those patients treated in a public health facility. Information on patients treated in a private setting was obtained from OACIS where possible, but additional information from case-note review was not pursued. Clinical information (other than age and gender) on the 7 patients from the CLL06 trial was not available. Metrics collected from the 60 local patients included age, gender, date of diagnosis, time to first treatment (TTFT) and overall survival (OS). If the exact date of diagnosis was not available in the case notes, the date at which a CLL clone of $>5 \times 10^9/L$ was identified on OACIS was used. If the date was recorded as a month (for example, "June 2014"), the date of diagnosis was set to the first of the month as a default (that is, 1/06/14 in this example). TTFT was defined as the time from the date of diagnosis until the date of the first day of treatment. OS was defined as the time from the date of diagnosis until death from any cause. Kaplan-Meier curves for TTFT and OS were constructed using GraphPad Prism software (version 7).

3.2.3 IGHV analysis

Determination of the somatic hypermutation status of the rearranged IGHV gene was performed on extracted DNA using the Lymphotrack® assay on the Illumina™ next-generation sequencing platform. Library preparation and pooling was performed by Ms Rachel Hall in the Department of Genetic Pathology (SA Pathology, Flinders Medical Centre, South Australia) and sequencing was performed at the Australia Cancer Research Foundation Cancer Genomics Facility (Adelaide, South Australia). IGHV sequences were aligned using IMGT/V-QUEST (www.imgt.org/IMGT_vquest/vquest) [175]. Samples were defined as mutated if $\geq 2\%$ of base pairs differed from the germline sequence as originally described [71, 72]. Stereotypes were assigned using ARResT/Assign Subsets (<http://tools.bat.infspire.org/arrest/assignsubsets/>) [176].

3.2.4 Next-generation sequencing (5-gene panel)

A customised NGS gene panel (GeneRead kit, Qiagen) targeting the most frequently mutated genes in CLL (*ATM*, *BIRC3*, *NOTCH1*, *SF3B1* and *TP53*) was designed and incorporated into CLL diagnostics at the SA Pathology Genetic Pathology laboratory (Flinders Medical Centre, South Australia). It was designed to cover exonic regions (and 20bp intronic regions either side of exons) at a depth of at least 100x with paired-end reads and an amplicon length of 225bp. Amplicon library preparation, sample pooling and purification was performed according to the manufacturer's instructions. In brief, genomic DNA was diluted to 2.5ng/μL and 4μL of solution was used in 4 separate PCR reactions (for a total of 40ng) per patient. PCR was performed using the GeneRead primer mix pool to generate amplicons. The 4 reactions were then pooled and purified using AMPure XP beads and the products quantified using the Qubit flurometer. Library construction for Illumina sequencing was then performed. Firstly, one-step end-repair and adapter ligation (adapters from Adapter Plate 96-plex Illumina) was performed using a thermocycler and adapter-ligated DNA was again cleaned up using Agencourt AMPure XP beads. Finally, the purified library was amplified using the HiFi PCR Master Mix and primer mix (GeneRead), purified with AMPure XP beads and quantified. The samples were then pooled into a single tube in equimolar ratios for a final concentration of 4nM. Library quality was assessed using the Agilent Bioanalyzer. The author performed the library preparation for 11 cases. The remaining 11 were prepared by Ms Rachel Hall in the Department of Genetic Pathology (SA Pathology, Flinders Medical Centre, South Australia). Sequencing was performed at the Australia Cancer Research Foundation Cancer Genomics Facility (Adelaide, South Australia) on an Illumina MiSeq platform.

Fastq sequencing files were aligned to the Genome Reference Consortium Human Build 37 (GRCh37) using NextGENe® software (SoftGenetics®) and a variant call file (vcf) created. The vcf was analysed using Illumina VariantStudio software and filtered using the following criteria: variants present in greater than 5% of reads, variants present with a population global frequency of <5%, and variants predicted to be deleterious by PolyPhen and SIFT. Variants were then further analysed manually. Any variant present disproportionately (usually >90%) in either the forward or reverse reads was discounted as sequencing error. In addition, single nucleotide variants (SNVs) that appeared solely at the last base of a read were discarded as sequencing error. Homopolymers were also discounted as

3.2.5 Sanger sequencing confirmation of variants

All variants detected by NGS (excepting variants in *ATM*) were validated using Sanger sequencing. Primers were designed for each variant using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and optimised on genomic DNA extracted from

the OSU-CLL cell line [177]. See Table 9-1 for a list of primer sequences. Four different sets of *BIRC3* primers were used to detect different mutations within the gene and are listed in Table 9-2.

3.2.6 Immunophenotyping

A 6-colour CLL antibody panel including the integrins CD49d and CD11b was designed (see Table 3-1). The cocktail was made up to a total of 600µL in FACS buffer. 15µL of the antibody cocktail was added to 50µL of thawed, washed patient PBMCs (containing 5×10^5 cells; see also section 2.3.1). An additional 2.5 µL of CCR7-APC (BD) was added to each sample. The reaction was incubated for 30 minutes in the dark and then washed in FACS buffer. Flow cytometry was performed on the BD FACS Canto™ II (BD Biosciences) and the data analysed using BD FACSDiva™ (BD Biosciences). Application settings were employed to standardise results across different time points of data acquisition.

Table 3-1. Details of antibody panel used in immunophenotypic characterisation of trisomy 12 cohort. BD = Becton Dickinson company.

Antibody	Fluorophore	Titre	Volume added to cocktail (µL)	Brand (clone)
CD5	V450	1:4	50	BD (L17F12)
CD38	FITC	1:2	100	BD
CD49d	PE-Cy5	1:2	100	BD
CD11b	PE	1:4	50	BD (D12)
CD19	PE-Cy7	1:2	100	BD (SJ25C1)
CD45	APC-H7	1:8	25	BD (2D1)

For each sample, 50 000 events were collected. The hierarchical gating strategy was as follows: lymphocytes were initially gated based on CD45 expression and their side scatter properties. Doublets were then excluded using forward scatter area versus side scatter area. CLL cells were gated based on co-expression of CD5 and CD19. Mean fluorescence intensity (MFI) of CD38, CD49d, CD11b and CCR7 was determined from the CLL cell gate. MFI of the lymphoid gate (based on forward and side scatter) of a matched unstained aliquot of each sample was subtracted from the MFI of the stained sample's CLL cell gate to account for autofluorescence. CD38 positivity was defined as >30% of CLL cells expressing CD38 as compared to internal negative isotype controls (CD5+ T cells) as previously defined [71]. CD49d positivity was similarly defined as >30% of CLL cells expressing CD49d compared to internal positive isotype controls (CD5+ T cells) as previously defined

[155]. MFI of the four investigational markers between the trisomy 12 and disomy 12 cohorts were compared using an unpaired t-test with Welch's correction (not assuming equal standard deviations). Fisher's exact test was used to compare the frequency of CD38 and CD49d positivity between trisomy 12 and disomy 12 groups. For the comparison of CD49d MFI between the three groups (trisomy 12 CLL, disomy 12 CLL and healthy controls), a one-way ANOVA was performed. Significant findings were defined as those with p values < 0.05.

3.3 RESULTS

The first part of the results sections deals with a retrospective case note audit of the entire cohort of trisomy 12 CLL cases as of 2017. The second part deals with the prospective analysis of these 22 trisomy 12 cases which had samples available for analysis. An additional 6 trisomy 12 samples added in 2020 were subjected to CD49d expression analysis alone.

3.3.1 Retrospective audit of trisomy 12 cohort

3.3.1.1 Clinical information

Sixty local patients with trisomy 12 CLL were identified through a centralised database (patients 1-8, and 16-67). A further 7 patients with trisomy 12 CLL from the national CLL06 clinical trial were identified (patients 8-15), resulting in a total cohort of 67 patients (see Appendix, Table 9-5). Only age, gender, and FISH results were available for the 7 CLL06 patients, however, all patients recruited on to the CLL06 trial were treatment naive as per the study protocol. Limited clinical information was available for 19 of 60 patients in the local group. In total, 43 of 67 (64%) of patients were male and the median age was 70 years. As of September 2017, 41 of the 47 (87%) patients with information available had undergone at least one line of therapy since diagnosis. The median follow-up time was 9.0 years. Within the local cohort, the median TTFT was 3.3 years (n=37) and OS was 15.6 years (n=39) (see Figure 3-1). The last date of observation for this analysis was the 21st of September 2017. Eleven deaths (11/39; 28% of patients) occurred during the period of observation. Two of the eleven deceased patients (18%) died of Richter's transformation and seven died of complications of progressive CLL (64%). Two patients (18%) died of metastatic carcinoma; one of the two patients also had progressive CLL at the time of death. Only 10 of the deaths are recorded in the survival curve as the date of diagnosis for one of the deceased patients is unknown.

3.3.1.2 Cytogenetic results

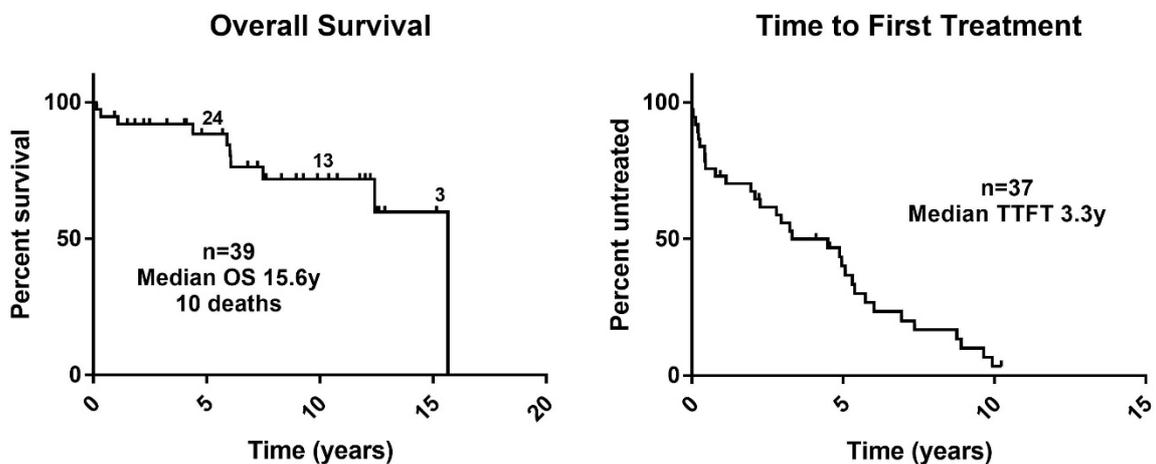
Cytogenetic reports for the 67-patient cohort were analysed and are summarised in Table 9-6 of the Appendix. Many patients had undergone serial testing, however, the results obtained as close to the

initial diagnosis and prior to the first treatment were interrogated wherever possible. Cytogenetic analysis was undertaken by either:

1. Fluorescence *in situ* hybridisation (FISH) (n=15);
2. Chromosomal Banding Analysis (CBA) (n=9); and
3. SNP microarray (n=43)

The clonal frequency of trisomy 12 was first established and divided into three categories: low frequency (<20% of nuclei with trisomy 12), medium frequency (20-60%) and high frequency (60%) as these values have previously been linked to clinical outcome [140]. Frequency data was available for 51 of 67 patients (15/15 by FISH, 9/9 by karyotype and 27/43 by array) and is shown in Figure 3-2. The number of additional cytogenetic abnormalities along with trisomy 12 is also presented in Figure 3-2.

Figure 3-1. Kaplan-Meier curves of overall survival and time to first treatment for local cohort of trisomy 12 CLL. Median overall survival was 15.6 years with 10 deaths observed during the period of observation. The number of patients at risk at 5, 10 and 15 years is indicated above the curve in left hand panel. Median time to first treatment was 3.3 years.

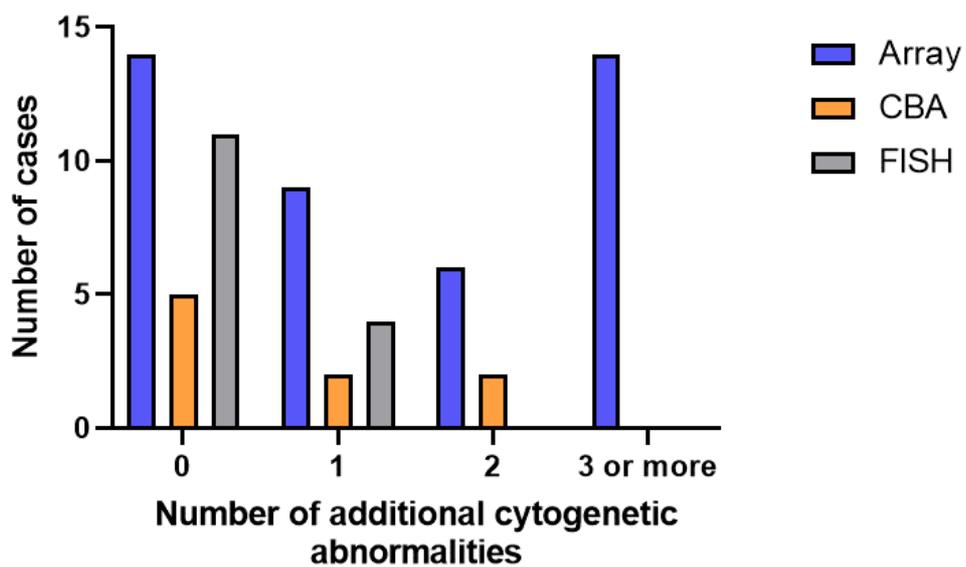


The type of abnormalities observed in addition to trisomy 12 are presented in Figure 3-3. Del(13q) was the most frequent additional aberration present in 7/37 (19%) of cases followed by del(14q) present in 6/37 (16%) of cases. Del(11q) was present in 4/37 (11%) of cases but never as a sole additional abnormality. Additional trisomies were present in 2 of 37 (5%) of cases: one case harboured three trisomies (12, 18 and 19) and the other harboured trisomy 19 as part of a complex

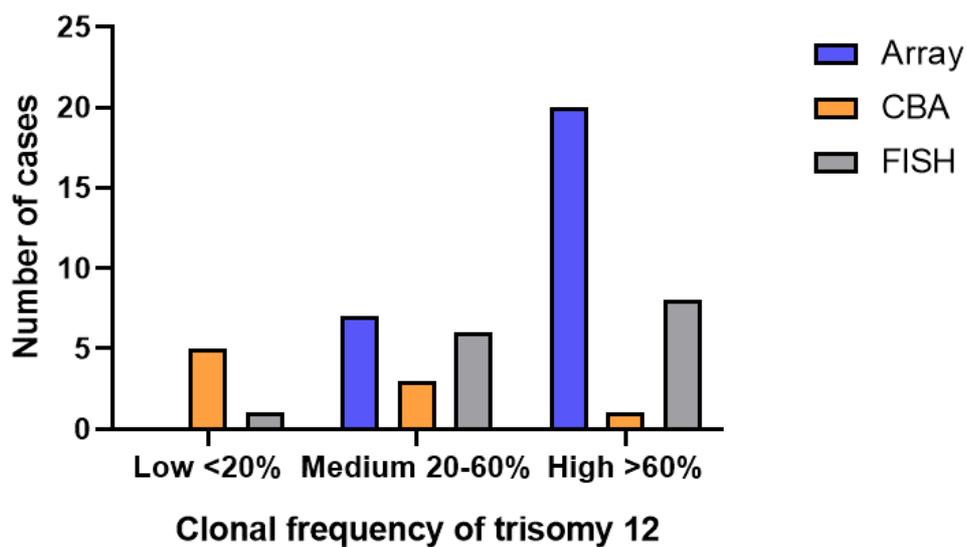
karyotype with trisomy 12. Two cases of del(17p) were observed, both in patients with complex karyotypes. Trisomy 12 was part of a complex karyotype (as defined by ≥ 3 chromosomal abnormalities) in 38% of cases.

Figure 3-2. Number of cytogenetic abnormalities in addition to trisomy 12 (A) and clonal frequency of trisomy 12 (B) in cohort as determined by SNP microarray (blue), CBA (orange) or FISH (grey). SNP = single nucleotide polymorphism; CBA = chromosomal banding analysis; FISH = fluorescence in situ hybridisation.

A

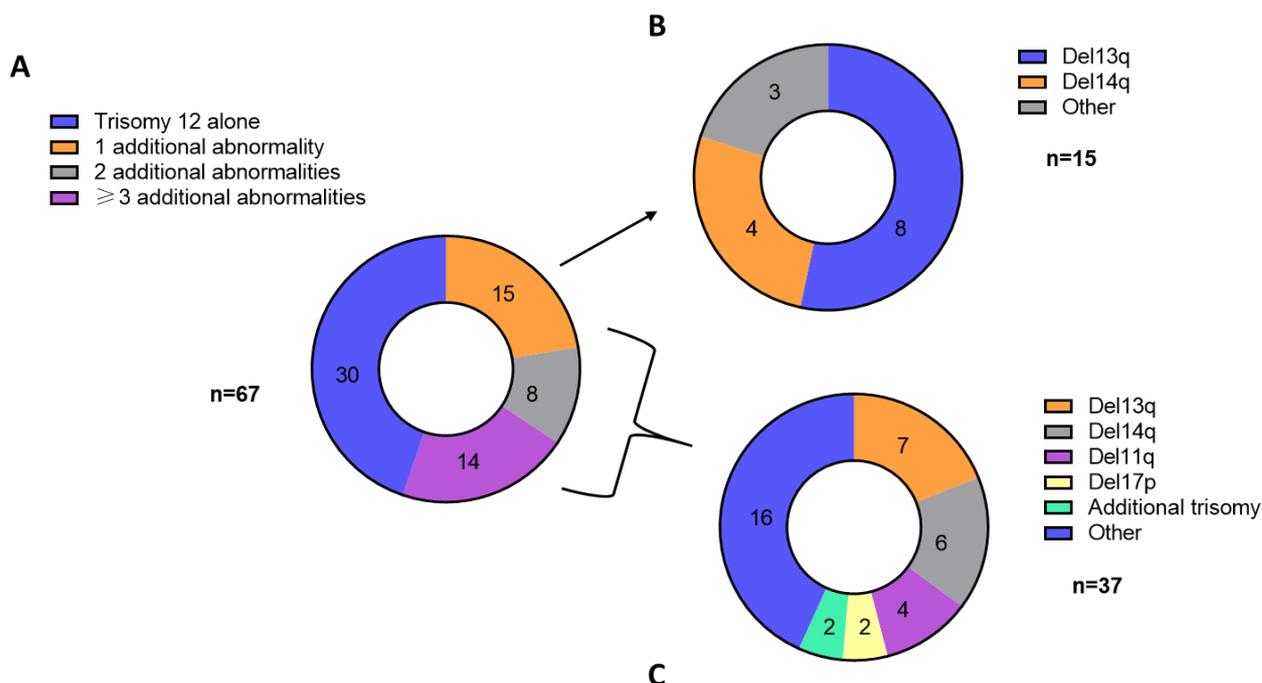


B



In all cases subjected to SNP microarray, there was complete duplication of an entire copy of chromosome 12. In one case (patient 34) there was also loss of heterozygosity at the following coordinates on chromosome 12: 42,304,125-48,713,787. This is an area of 6,409,663bp in length and contains the gene *NELL2*. The homozygosity at this locus was part of a complex karyotype with chromothripsis and is of uncertain significance.

Figure 3-3. Cytogenetic abnormalities in addition to trisomy 12 in the local cohort (as detected by any method). The pie graph in (A) shows the number of cytogenetic abnormalities for the whole cohort (n=67). The number in each pie section refers to the number of patients affected. The pie graphs in (B) and (C) show the nature of the additional chromosomal abnormalities. (B) shows the breakdown for patients with only one additional abnormality (n= 15), and (C) shows the breakdown for patients with 1, 2, 3 or more additional abnormalities (n=37).



3.3.2 Prospective analysis of trisomy 12 cohort with available samples

The remaining sections describe results obtained prospectively on 22 of 67 patients with available cryopreserved samples (patients 1-22). Each sample underwent targeted IGHV analysis, extended immunophenotyping (including CD49d and CD38), and mutational analysis of 5 commonly mutated genes in CLL. Pre-treatment samples were selected where possible: 16 of the 22 samples were taken prior to any anti-leukaemia treatment, 3 samples were following 1 line of therapy (patients 1, 2 and

Table 3-3. IGHV results for the local trisomy 12 CLL cohort. U = unmutated; M = mutated.

Patient ID	IGHV		
	U/M	IGHV gene usage	Stereotyped (subset)?
1	M	IGHV3-07	no
2	U	IGHV1-08	no
3	U	IGHV2-05*10	no
4	U	IGHV5-05*01	no
5	M	IGHV3-09*01	no
6	1 clone U 1 clone M	IGHV3-21*01 & IGHV4-34*01	unknown
7	U	IGHV7-4*01	yes (subset #1)
8	U	IGHV6-01*01	no
9	U	IGHV3-07*03	no
10	U	IGHV4-39*07	no
11	U	IGHV4-39*01	yes (subset #8)
12	M	IGHV3-07*03	no
13	M	IGHV3-73*03	no
14	M	IGHV4-39*07	no
15	U	IGHV1-69*01	no
16	U	IGHV3-49*03	no
17	U	IGHV4-39*01	yes (subset #8)
18	U	IGHV1-08	no
19	U	IGHV4-59*01	no
20	U	IGHV1-69*01	no
21	U	IGHV3-30*03	no
22	U	IGHV1-69*01	no

3.3.2.2 Mutational profile

Twelve of 22 (55%) of patients had an identifiable protein-coding mutation within *ATM*, *BIRC3*, *NOTCH1*, *SF3B1* or *TP53* (see Table 3-4). Seven patients (32%) had one mutation and five patients (23%) had two mutations in the same or different genes. *BIRC3* mutations were most common (27%) followed by mutations of *ATM* (18%), *NOTCH1* (9%), *SF3B1* (5%) and *TP53* (5%). Both *NOTCH1* mutations identified were the most common pathogenic mutation reported in CLL

(c.7541_7542delCT). One of the *BIRC3* mutations (c.1639delC; patient 2) is a known recurrent mutation in CLL, although its functional consequence is yet to be elucidated. Interestingly, the other five identified *BIRC3* mutations have not been reported in CLL and are not entered in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (<https://cancer.sanger.ac.uk/cosmic>). It is possible that these variants represent germline polymorphisms rather than true mutations, however, this could not be confirmed as constitutional DNA was not available for analysis. The *BIRC3* mutation in patient 12 belongs in a mutation hotspot for haematopoietic/lymphoid cancers, and the mutation in patient 11 has been reported in a patient with bladder cancer in the COSMIC database. The *TP53* mutation in patient 19 has been previously reported [178], but the *SF3B1* mutation in patient 14 has not. All variants were verified by Sanger sequencing with the exception of the *ATM* variants, and the *BIRC3* variant in patient 1 due to insufficient DNA.

3.3.2.3 Immunophenotyping

An example of the gating strategy employed for immunophenotyping is presented in Figure 3-4. Expression levels (MFI and percentage positive cells) of CD38, CD11b and CCR7 were determined in 18 of the 22 cases in the cohort from 2017 (see Table 9-14 and Figure 3-5). The remaining 4 cases did not have enough stored viable cryopreserved cells for immunophenotypic analysis (these samples were used for DNA extraction alone; patients 1, 4, 13 and 16). Five disomy 12 CLL controls were used as comparators (DIS1-5) in the analysis but DIS2 was later excluded as it did not meet the diagnostic criteria for CLL. There was no statistically significant difference in MFI of CD38, CD11b or CCR7 between the trisomy 12 and disomy 12 CLL groups (see Figure 3-5A,C,D). No cases were CD38 positive (>30% of cells) and there was no difference between CD38 expression in terms of percentage positive cells between trisomy 12 and disomy 12 CLL (see Figure 3-5B).

CD49d expression was first analysed in 2017 on these same cases (data not shown), and again in 2020 on 14 trisomy 12 cases and 11 disomy 12 cases (see Table 9-14 and Figure 3-6). CD49d expression was significantly higher in the trisomy 12 group compared to the disomy 12 group when comparing MFI or percentage positive cells ($p=0.048$ and $p=0.00004$ respectively; see Figure 3-6A,B). All 14 trisomy 12 cases (100%) were CD49d positive compared to 3/11 (27%) of the disomy 12 group ($p=0.0002$; see Figure 3-6C). Two of the trisomy 12 cases (and no disomy 12 cases) demonstrated bimodal expression of CD49d with two distinct positive and negative populations (patients 6 and 21) (see Figure 3-7). CD49d expression on the three healthy controls (mean MFI \pm sd = 11827 \pm 1513) was significantly higher than on either the trisomy 12 group (mean \pm sd = 4290 \pm 2421; $p=0.0004$) or the disomy 12 group (mean \pm sd = 1884 \pm 3201; $p<0.0001$).

Table 3-4. Mutations in ATM, BIRC3, NOTCH1, SF3B1 and TP53 in trisomy 12 CLL cohort. VAF = variant allele frequency. SNP = single nucleotide polymorphism. "Consequence" of uncertain significance or benign SNP is as reported from ClinVar. *insufficient DNA for confirmation. ATM variants were not confirmed with Sanger sequencing.

Patient ID	Gene	Mutation	Predicted protein change	VAF (%)	Read Depth	Consequence	Sanger confirmation
1	BIRC3	c.784T>C	p.Phe262Leu	21	3964	not reported	No*
2	BIRC3	c.1639delC	p.Gln547Asn fsTer21	43	16371	recurrent in CLL	Yes
9	ATM	c.6919C>T	p.Leu2307Phe	-	-	uncertain	No
11	NOTCH1	c.7541_ 7542delCT	p.Pro2514Arg fsTer4	44	382	known pathogenic	Yes
11	BIRC3	c.73G>A	p.Asp25Asn	47	2104	not reported	Yes
12	BIRC3	c.1665_ 1666delAA	p.Arg555Ser fsTer3	31	8482	not reported	Yes
13	ATM	c.2572T>C	p.Phe858Leu	48	5918	uncertain	No
13	ATM	c.3161C>G	p.Pro1054Arg	48	4045	benign SNP	No
14	SF3B1	c.2558T>C	p.Ile853Thr	47	5562	not reported	Yes
16	BIRC3	c.1664_ 1665insTT	p.Arg555Ser fsTer2	41	5228	not reported	Yes
16	BIRC3	c.1759A>T	p.Lys587Ter	20	1824	not reported	Yes
17	ATM	c.2572T>C	p.Phe858Leu	48	3682	uncertain	No
17	ATM	c.3161C>G	p.Pro1054Arg	48	2296	benign SNP	No
19	TP53	c.731G>A	p.Gly244Asp	44	4869	reported	Yes
21	NOTCH1	c.7541_ 7542delCT	p.Pro2514Arg fsTer4	51	442	known pathogenic	Yes
21	BIRC3	c.1298_del AAAinsA	p.Arg434Ser fsTer3	28	5926	not reported	Yes
22	ATM	c.1363G>A	p.Val455Met	48	5182	uncertain	No

Figure 3-4. Example of hierarchical gating strategy (patient 6). Lymphocytes were gated based on CD45 expression and side scatter. Singlets were gated based on forward versus side scatter area. CLL cells were gated based on co-expression of CD5 and CD19 (green). T cells were gated based on CD5 expression and absence of CD19 expression (pink). T cells were used as internal CD49d positive controls and CD38 negative controls.

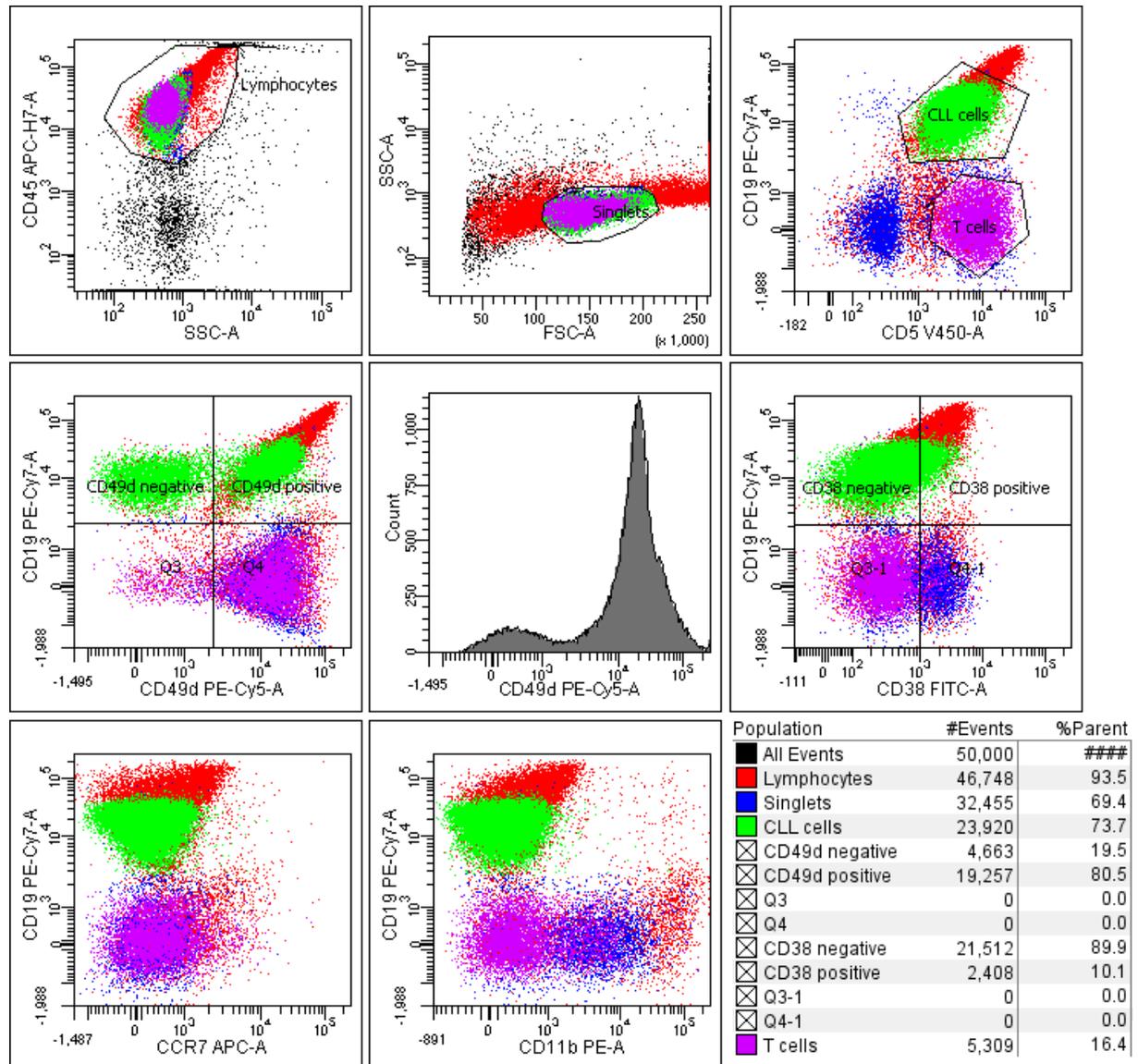


Figure 3-5. CD38, CD11b and CCR7 expression in trisomy 12 and disomy 12 CLL. Individual samples are dots. Mean \pm standard deviation is plotted. MFI = mean fluorescence intensity. MFI values are the difference between the marker's value in the stained CLL cell compartment minus the MFI of unstained lymphocytes. p-values refer to results of unpaired t-tests with Welch's correction.

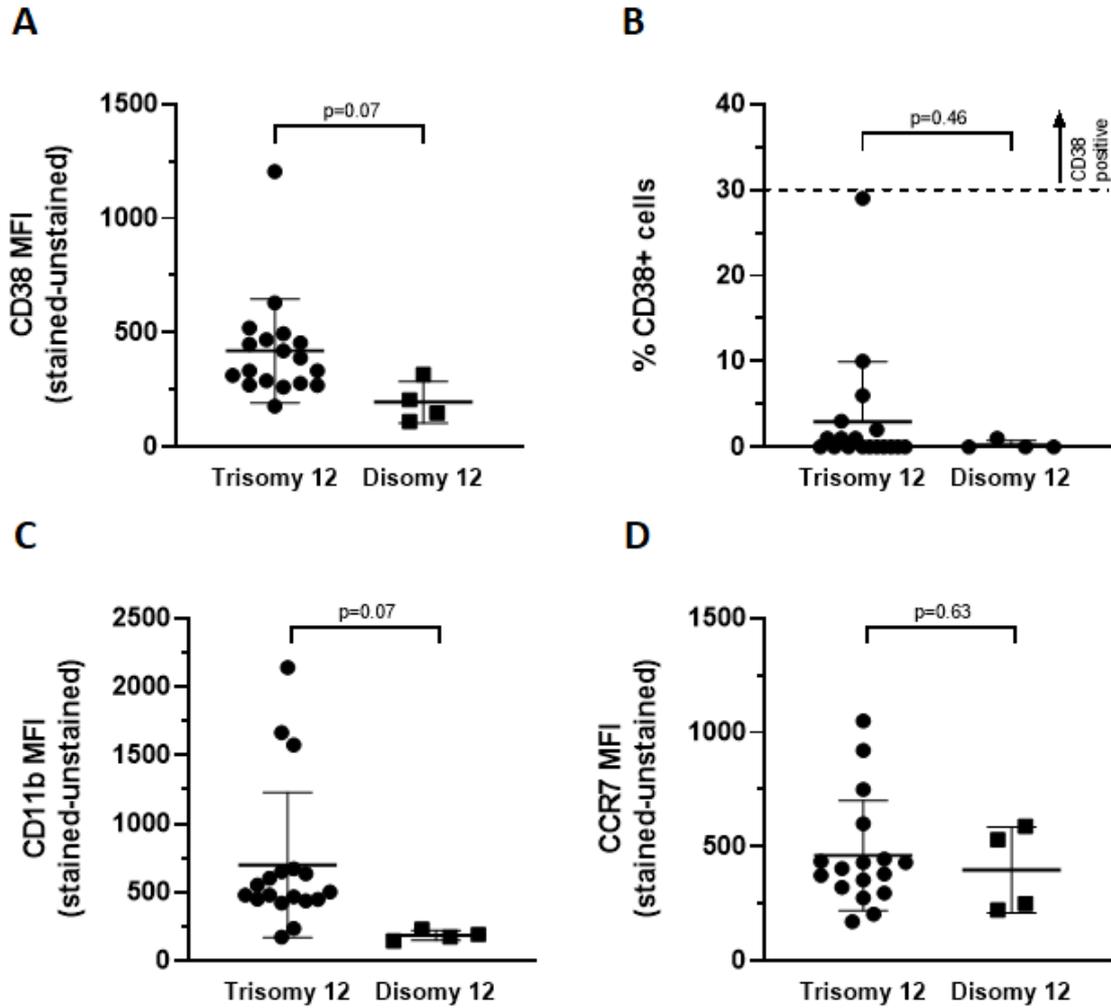


Figure 3-6. CD49d expression in trisomy 12 and disomy 12 CLL. Individual samples are dots. Mean \pm standard deviation is plotted. MFI = mean fluorescence intensity. MFI values are the difference between the marker's value in the stained CLL cell compartment minus the MFI of unstained lymphocytes. (A) CD49d expression (MFI) in trisomy 12 and disomy 12 CLL. (B) CD49d expression (% positive cells) in trisomy 12 and disomy 12 CLL. (C) Number of CD49d+ (>30% of cells expressing CD49d) and CD49d- cases in trisomy 12 and disomy 12 CLL. p-values refer to results of unpaired t-tests with Welch's correction in (A) and (B), p-value in (C) refers to the Fisher's exact test comparing the frequency of CD49d+ cases between the two groups.

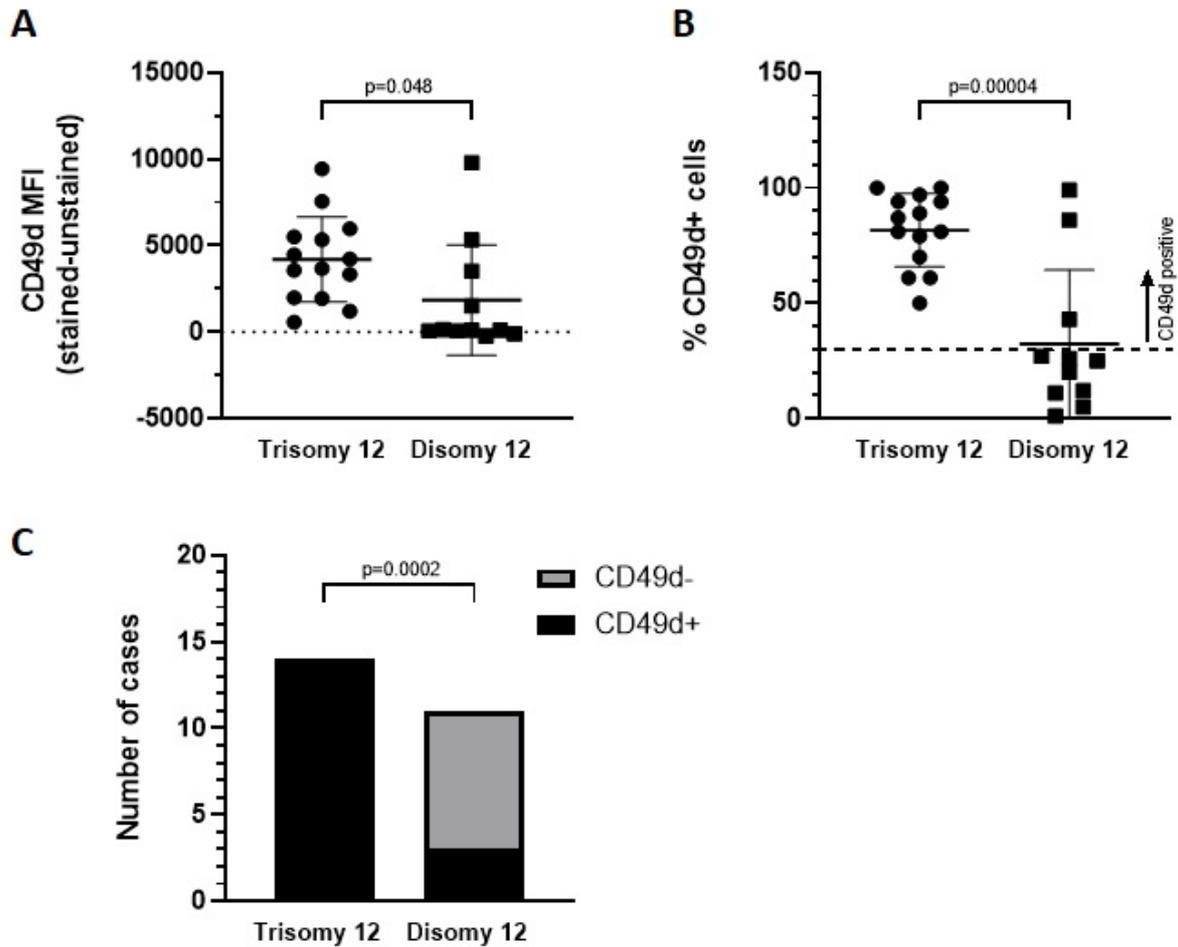
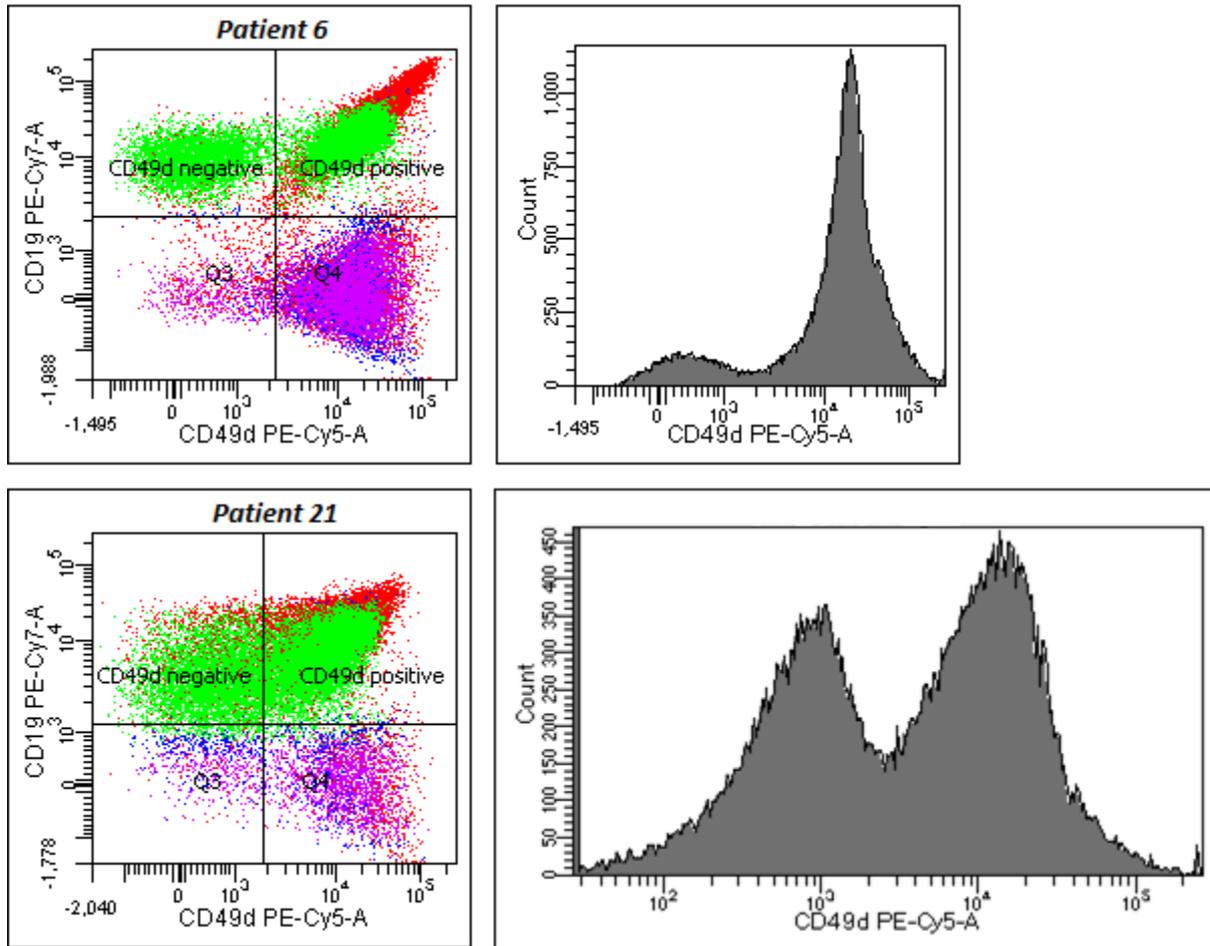


Figure 3-7. Two trisomy 12 CLL cases exhibit bimodal expression of CD49d. Dot plots and histograms showing CD49d expression of CLL cells (green) in patients 6 and 21. CD5+/CD19- T cells (pink) are internal controls.



3.4 DISCUSSION

This chapter describes the characterisation of a local cohort of patients with trisomy 12 CLL. The cohort matches well with larger international series in terms of clinical features, clinical outcome, and CD49d expression, however, the group has a lower-than-expected frequency of *NOTCH1* mutations. The cohort median age of 70 years and male preponderance is in keeping with that expected for a group of patients with CLL. There was missing data for several patients and only a subset of patients had samples available for immunophenotyping, IGHV and mutational profiling, decreasing the power of the study. There is also an inherent selection bias to the cohort in terms of those patients who were selected for cytogenetic testing by their physician that may lead to data skewing. Although cytogenetic analysis is only recommended prior to any treatment decision and not for those patients with asymptomatic early-stage disease, variable approaches by different clinicians resulted in a wide variation in time points for when cytogenetic analysis was performed. For example, there were patients that were found to have trisomy 12 years prior to treatment. On the other hand, a proportion of patients either presented later (when symptomatic) or underwent cytogenetic testing just prior to treatment as per recommendation, and therefore had a shorter time to first treatment. In addition to this, patients not referred for cytogenetic testing will of course not be captured in this analysis, and it is unknown how many patients with trisomy 12 CLL have been treated without cytogenetic testing. Despite this bias, the median OS was identical to a large published international cohort (15.6 years) and the median TTFT was only slightly longer in the presented series (3.3 years versus the published 3.1 years) [130]. Therefore, whilst the cohort is small with only 22 samples, it appears to be representative of the heterogeneous group of trisomy 12 CLL, at least in terms of clinical outcomes.

The type of method used to determine trisomy 12 affected the observed clonal frequency and number and type of additional cytogenetic abnormalities detected. In 9 patients (9/67; 13%), trisomy 12 was detected by chromosomal banding analysis (CBA). This is the least sensitive of the three methods employed and requires active metaphases for detection. Historically, it has been shown that actively dividing cells may be difficult to induce in primary CLL samples with traditional mitogens, meaning some less proliferative or quiescent clones may not be identified [179]. In 15 patients (15/67; 33%), FISH to detect del(13q), del(11q), del(17p) and trisomy 12 was performed with a reported sensitivity of approximately 5%. Low frequency trisomy 12 clones will be identified with this method; however, additional cytogenetic abnormalities will be missed given the small number of probes used in this approach. For example, an additional trisomy 19 as found in one patient using a SNP microarray (patient 52) would not have been detected by FISH. Finally, the majority (43/67; 64%) of patients had trisomy 12 detected by SNP microarray. This technique

provides a whole genome overview (like CBA), however, does not require active metaphases and will identify areas with loss of heterozygosity (LOH) and exact co-ordinates of gains and losses. On the other hand, balanced translocations and low frequency clones will not be identified by microarray (its sensitivity was reported as approximately 20% in all cases). Despite these methodological differences, trisomy 12 was most often present in high clonal frequency. There were cases (identified by FISH or CBA only) that demonstrated low/medium frequency trisomy 12 clones, which is not entirely in keeping with its hypothesised role as a founder lesion CLL [83]. Given the known prognostic impact of additional abnormalities such as trisomy 19, del(14q) or karyotypic complexity (see later), it is possible that FISH alone will not be performed diagnostically and will be complemented with either CBA and/or SNP microarray in the future. This is reflected in the updated 2018 guidelines published by the iwCLL suggesting the inclusion of CBA alongside FISH in clinical trials in recognition of the prognostic impact of a complex karyotype [180]. The advantage of SNP array over CBA is the addition of data including exact chromosomal co-ordinates of loss and/or gain (which makes it easier to identify genes involved) and the identification of loss of heterozygosity (for example, LOH at 17p may suggest the presence of homozygous mutation of the tumour suppressor gene *TP53* which has clear negative prognostic implications). Furthermore, microarray does not require active metaphases, however, it is more expensive, does not identify low frequency clones and outcomes using this technology have not been published or subjected to prospective clinical trials. The sensitivity of cytogenetic analysis and identification of low frequency clones is clearly important information, both prognostically and with regards to the development and evolution of CLL. This will be discussed further in the final chapter (see Chapter 7).

Del(13q) and trisomy 12 have been reported as having a negative co-occurrence in keeping with their hypothesised roles as separate founder lesions in CLL [83]. However, del(13q) was observed in addition to trisomy 12 in 19% of the cohort and in one patient (patient 14), both changes were present at high frequency (90%) and within the same cells, suggesting the lesions are not necessarily mutually exclusive. Del(11q) and del(14q) were the two next most frequent additional abnormalities detected. Del(11q) is a recurrent finding in CLL and an association between trisomy 12 and del(14q) has been previously reported [142, 145]. The reason behind this association is not clear. There may have been further cases with an unidentified del(14q) as this is not routinely tested for in the diagnostic FISH assay. Several (but not all) cases that underwent FISH testing alone did have a region of chromosome 14q interrogated using the IGH locus probe and were reported as “diminished intensity”. Whilst this may reflect somatic VH recombination events, a larger deletion of 14q that contributes to leukaemogenesis cannot be excluded and would fit with the observed

deletions of 14q detected by SNP array in 16% of cases. Further analysis of the size and exact coordinates of these deletions is beyond the scope of this research.

Many other additional abnormalities were observed in combination with trisomy 12 but were different from patient to patient. There were changes observed in the sex chromosomes including Y-, however, this change may indeed be a constitutional change that occurs with advancing age in men rather than a pathogenic event [181]. No germline samples were used to confirm somatic change in any case (all detected changes were presumed somatic) as is the standard practice in the diagnostic laboratory in this setting. Given the advanced age of the population, it is assumed that any oncogenic germline aberrations (such as *TP53* disruption; Li-Fraumeni syndrome) would have been diagnosed at a younger age and previously caused disease.

Additional trisomies (in particular +18 and +19) have been reported in association in trisomy 12 [133, 141]. This was not borne out in the current cohort with only 5% of cases demonstrating an additional trisomy (as opposed to the published 12% [141]). However, as mentioned previously this may be due to the detection technique (FISH analysis which will only detect a trisomy of chromosome 12) or the relatively small cohort size limiting statistical power. Del(17p) was observed infrequently in the cohort (6%) which may reflect the time point chosen for analysis (that is, mainly pre-treatment analyses were chosen if available): del(17p) is known to increase in frequency with relapsed and chemo-refractory disease. Finally, trisomy 12 was observed as part of a complex karyotype in 38% of cases. At first glance, this appears incongruous with the early, pre-treatment time point used for analysis, however, traditionally complex karyotype is defined as 3 or more abnormalities as detected by CBA. This definition was applied in this thesis to patients with ≥ 3 chromosomal aberrations as detected by the more sensitive SNP microarray. There is no published definition of “complex karyotype” using this technology and it is unclear if there is a certain number of aberrations that will correlate with clinical outcome. It seems likely that the number of abnormalities alone will not predict outcome and will be influenced by the specific nature of the abnormalities. This is supported by a recent retrospective study of the clinical outcomes of 5290 CLL patients that challenged the uniformly poor prognosis of patients with complex karyotype (≥ 3 abnormalities by CBA). Patients with ≥ 5 abnormalities had uniformly poor outcomes, however, cases with 3-4 abnormalities had poor prognoses only with concomitant *TP53* mutation [182]. Interestingly, patients with trisomies 12 and 19 and at least one more abnormality (traditional complex karyotype) demonstrated prolonged overall survival.

The definition has been applied in the current study to cytogenetic abnormalities detected by SNP array technology and as such, the clinical implications of a “complex karyotype” in this

setting are unknown. In conclusion, the small cohort size rendered it difficult to identify any novel or to confirm known associations between different chromosomal aberrations with confidence, however, reported associated changes were at least observed in the cohort.

Approximately three quarters of the cohort had an unmutated IGHV, compared to the approximate 50% expected [130]. This again may represent the small cohort size or may reflect the cohort selection bias. That is, unmutated cases are more likely to require treatment and are thus more likely to have undergone cytogenetic analysis as part of the pre-treatment workup. However, this would equally apply to a larger international cohort. Two of the 17 (12%) unmutated cases were assigned to stereotype 8, in line with the known enrichment of stereotype 8 for trisomy 12 cases with unmutated IGHV (they comprise 60% of stereotype 8 cases) [36]. Neither case had developed Richter's transformation, a known complication of subset 8, at the time of analysis.

An interesting observation is that *NOTCH1* mutations were less frequent in this cohort than expected from the published data (9% in this cohort versus 30-45%). The reason for this is unclear, but again may be due to the small sample size. It is possible that technical factors played a role, however, sequencing coverage in exon 34 of *NOTCH1* (the mutational hotspot) was adequate with a minimum coverage of 380, which is sufficiently high to detect even low frequency mutant clones. Mutations in the 3'-untranslated region were not captured by this exonic analysis but represent only a small fraction of the burden of *NOTCH1* mutations in trisomy 12 CLL, so this is not expected to explain the low frequency of *NOTCH1* mutants in the local cohort. *NOTCH1* mutations are also known to be enriched in treatment refractory disease, however, none of the four patients who had undergone previous treatment in the set harboured a mutated *NOTCH1*. Given that only 2 patients in the current cohort had *NOTCH1* mutations, further associations could not be reliably drawn.

In contrast to the low number of *NOTCH1* mutations, a high number of *BIRC3* mutations was observed (27%), in keeping with the known association of these mutations with trisomy 12 [83]; however, 5 of the 6 detected mutations have not been previously reported in the COSMIC database. The reason for this is unclear. Two of these 5 unreported mutations (c.1664_1665insTT and c.1665_1666delAA) resulted in a frameshift indel in a mutational hotspot of the gene (exon 9) and result in a protein truncation similar to the known pathogenic mutation (c.1658_1661delAAGA) at this locus [183]. The potential functional consequences of these mutations have not been described and is not further pursued in this thesis. A low incidence of *SF3B1* and *TP53* mutations was observed as expected given the trisomy 12 cohort has shown almost mutual exclusivity of the former and a low rate of co-occurrence with the latter [83].

Overexpression of CD49d was confirmed in the cohort in terms of both an increased frequency of CD49d-positive cases (>30% of cells expressing CD49d) and increased MFI in the trisomy 12 group. The range of CD49d expression in both trisomy 12 and disomy 12 groups was large, reflecting the biological heterogeneity of the disease, and is not a manifestation of the size of the data set, being observed in larger cohorts [155]. Three disomy 12 cases were CD49d+ but there were no distinguishing clinical or molecular features to account for this discrepancy: one case had a normal karyotype; another had a complex karyotype with del(17p) and the third had del(13q); two cases had an unmutated IGHV, and one had a mutated IGHV. The *NOTCH1* mutational status of the disomy 12 cases are unknown and could potentially account for the high CD49d expression. Interestingly, CD49d expression was significantly lower than in normal B cells from healthy controls (but more so in the disomy 12 group). It appears that trisomy 12 CLL cells more closely resemble normal B lymphocytes (at least in terms of CD49d expression) than their disomic counterparts. This is expanded upon in Chapters 6 and 7. There was also a trend to increased expression of CD11b, however, this appeared to be due to three cases within trisomy 12 that expressed it at high levels and there was again large inter-patient variation. There were no distinguishing features to subcategorise the three high-expressors.

Increased expression of CD38 in the trisomy 12 group was not observed. Overall, only one case demonstrated strong CD38 expression (DIS2) and would be classified as CD38-positive (data not shown), however, this case was later excluded as it met diagnostic criteria for SLL but not CLL. Both biological and technical factors were considered to account for this unexpected finding, although given the fact that strong expression of CD38 was observed in at least one case (and that all samples underwent the same processing), technical factors seem less likely. To understand the unexpectedly low frequency of CD38 positivity, any available diagnostic immunophenotyping reports were first interrogated. CD38 was not always performed at diagnosis, however, in patients 3 and 6 discrepant results were evident. In patient 3, the diagnostic immunophenotype was reported as CD38 positive (with 38% of CLL cells positive for CD38) as compared to the 1% detected in the cryopreserved sample used in the above analysis. Patient 6 also showed a dramatic change with 100% of CLL cells reported as CD38 positive diagnostically, and only 10% positive in the thawed cryopreserved sample. In patient 3 the diagnostic immunophenotype was performed on a fresh peripheral blood sample in 2012 and the sample used for this analysis was taken in 2014. The samples were taken two years apart, so it is possible that CD38 expression was lost during this time as this marker is known to vary during the clinical course of CLL [184]. However, the sample stored (and used in this analysis) for patient 6 was taken contemporaneously to that used for the diagnostic immunophenotyping report and changes this large (10% to 100% positive cells) have not been reported [184]. This raises the

possibility of loss of CD38 expression during cryopreservation and thawing of samples. Changes in expression of CD38 with cryopreservation, however, was refuted in at least one report [185]. A potential change in expression for CD49d, CCR7 or CD11b with processing could not be assessed as none of these markers are routinely performed in a diagnostic setting and no fresh peripheral blood samples were immunophenotyped with these markers at the same time the samples were stored.

One of the most interesting findings from this cohort analysis was that two samples were found to have bimodal expression of CD49d (patients 6 and 21). At the time of performing these experiments, it was thought that the presence of bimodal CD49d expression was unusual and was rare even in larger data sets. For example, in a series of over 1200 CLL cases, 13 of 206 trisomy 12 cases were reported as “mostly” bimodal for CD49d (the exact number of bimodal cases overall was not formally reported but is less than 13 in 206, 6.3%) [155]. Each of the bimodal cases in this study had low/intermediate trisomy 12 clonal frequency (10-27% of nuclei). In the same study, FISH analysis of flow-sorted fractions from three of the bimodal cases showed that the CD49d-positive clone was enriched for trisomy 12 nuclei (present at high frequency) and that the CD49d-negative clone appeared to contain nuclei with disomy 12 only. This finding suggests the presence of two completely different leukaemic clones – one with a high frequency of trisomy 12 that expresses CD49d, and another disomic clone that does not express CD49d. Newer data has shown that bimodal CD49d expression is more common than previously thought and was found in 20% of a cohort of 1630 CLL samples and had negative prognostic implications [186].

As well as bimodal expression of CD49d, patient 6 had two different productive rearranged IGHV sequences and biallelic IGH locus rearrangements indicating the presence of two completely different leukaemic clones within the same tumour. This is an unusual feature in CLL, especially as each clone demonstrated a different mutational status [187]. This case represents a powerful opportunity to study two separate clones with differing mutational statuses of the IGHV (the strongest prognostic marker in CLL) within the same patient, hence removing the large number of other variables encountered when comparing two such clones between two different patients with this inherently highly heterogeneous disease. On the other hand, bimodal expression of CD49d in patient 21 may indicate the presence of two subclones with the same rearranged IGHV sequence. Both bimodal cases afford an opportunity to explore the evolution of trisomy 12 which is currently thought to be a founding lesion in CLL. As such, both patient samples were further analysed, and the results are presented in the following chapter (Chapter 4).

In conclusion, the clinical, genetic and immunophenotypic characteristics of a local cohort of patients where trisomy 12 was determined. The cohort had similar clinical features to larger

international cohorts and demonstrated the expected over-expression of CD49d. A lower-than-expected frequency of *NOTCH1* mutations and CD38 positivity was observed. Two patients with bimodal expression of CD49d and two separate leukaemic clones (patient 6) or subclones (patient 21) were identified. These cases are further analysed in the following chapter with the aim to gain insights into the evolution of trisomy 12 CLL following purification of the separate clones based on differing expression of CD49d. The comprehensive characterisation of this local cohort will now serve as an invaluable tool to further investigate the poorly understood molecular pathogenesis of trisomy 12 CLL.

4 INVESTIGATION OF TWO CASES OF TRISOMY 12 CLL WITH BIMODAL CD49D EXPRESSION

4.1 INTRODUCTION

Characterisation of the local cohort of trisomy 12 CLL identified two patients (patient 6 and patient 21) who demonstrated bimodal expression of the integrin CD49d (see previous chapter). Patient 6 also harboured two different CLL clones based on the targeted IGHV sequencing results. This finding supports the wide recognition of CLL as a clonally complex disease with a high degree of intra-tumoural heterogeneity and raises questions about the origins and evolution of CLL. Whilst approximately 10.5% of CLL cases harbour double IGHV rearrangements, the presence of a double productive rearrangement with discordant mutational status in patient 6 is a rare finding, present in only 0.7% of cases [188, 189].

The genetic profile of the different clones in individual cases with multiple IGHV rearrangements has not been explored in detail before. To undertake such an analysis, it would be necessary to purify the leukaemic clones with the different IGHV rearrangements based on differential expression of a surrogate cell surface marker. This has been performed only once as far as the author is aware, by exploiting differential expression of surface lambda and kappa immunoglobulin light chains in two cases, however, there was no further analysis of the purified clones beyond confirmation of their differing IGHV status [190].

It was envisioned that the separation of two putative unique leukaemic clones or subclones for each patient based on their CD49d expression would allow for a direct comparison between disomy 12 and trisomy 12 CLL, and IGHV-unmutated and IGHV-mutated CLL in the case of patient 6. This was with the aim to identify potential driver events unique to trisomy 12 CLL. The power of these individual cases is that they control for background genetic variability that usually confounds the comparison between disomy and trisomy 12 CLL between large groups of different, often elderly, patients with this inherently heterogeneous and clonally complex disease.

4.2 METHODS

4.2.1 Flow cytometry

Cyropreserved PBMCs from patients 6 and 21 were thawed, washed, and resuspended in RPMI-1640 media (Life Technologies Australia Pty Ltd.) supplemented with 10% FCS (Life Technologies Australia Pty Ltd.) for a final concentration of $0.3-2 \times 10^7$ cells/100 μ L (see also Chapter 2). The samples were

stained with 10-20 μ L of CD5, CD19 and CD49d (the same antibodies listed in Table 3-1) and incubated for between 20 and 60 minutes in the dark. The samples were washed in FACS buffer (with 0.02% sodium azide and 1% FCS) and resuspended in RPMI+10% FCS or FACS buffer. The initial samples for patient 6 were flow sorted using the BD FACS Aria flow cytometer at Flinders University in 2018 (machine now decommissioned). Samples for patient 21 were sorted using the BD FACS Aria Fusion flow cytometer at Flinders University in 2020.

Cell sorting occurred on two separate occasions for patient 6 using two cryopreserved vials each time to ensure adequate material for downstream analyses (the first sort in February 2018 was used for DNA extraction alone, and the second sort in June 2018 for RNA extraction). On the first occasion, three fractions were sorted: CD49d-positive (CD49d+) CLL cells, CD49d-negative (CD49d-) CLL cells and CD5-positive, CD19-negative T cells. On the second occasion, only two fractions were sorted: CD49d+ CLL cells and CD49d- CLL cells.

Two vials were sorted from patient 21 in February 2020 to obtain two fractions: CD49d+ CLL cells and CD49d- CLL cells. The fractions were identified using the following hierarchical gating strategy in both cases: lymphocytes were gated based on forward and side-scatter properties, doublets were then excluded based on forward scatter area and height, CLL cells were then gated based on CD5 and CD19 co-expression and sorted into CD49d+ and CD49d- fractions using the CD49d expression histogram. CD5+ CD19- T cells were also sorted simultaneously in the case of patient 6.

A post-treatment sample from patient 6 (dated 09/03/2016, 6 months following initiation of ibrutinib) was available and subjected to immunophenotyping as per Chapter 3.

4.2.2 DNA extraction & storage for RNA extraction

Sorted cells were used to either extract genomic DNA or total RNA using the methods described in Chapter 2. The samples for RNA extraction were pelleted, resuspended in 1mL of Tri Reagent® or TRIzol™ and stored at -70°C until required (see Chapter 5).

Genomic DNA representing a germline sample was also extracted from a historical skin biopsy of a non-malignant lesion for patient 6 (who was deceased at the time of this study). Material was scraped from 6 unstained slides of the biopsy prepared from the original cell block and placed into an Eppendorf tube with a scalpel blade. The sample was then de-paraffinised with 800 μ L of xylene. The mixture was vortexed, centrifuged at maximum speed in a microcentrifuge for 5 minutes, and washed in 800 μ L of absolute ethanol. The pellet was air dried at room temperature for

one hour and then DNA extracted using the Qiagen™ QIAmp FFPE DNA extraction kit as per the manufacturer's recommendations.

4.2.3 Single Nucleotide Polymorphism array

SNP microarray analysis (Illumina CytoSNP-850Kv) was carried out at the Cytogenetics Laboratory (Division of Genetics & Molecular Pathology, SA Pathology, Adelaide, South Australia) on genomic DNA extracted from the CD49d+, CD49d- of patients 6 and 21, and the T cell fraction of patient 6. Reports were issued to the author.

4.2.4 Whole Exome Sequencing

Genomic DNA from the CD49d+, CD49d- and T cell fractions of patient 6 were sent to the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility (Centre for Cancer Biology, University of South Australia, Adelaide, South Australia) for whole exome sequencing with 150x coverage, using an Illumina™ NextSeq Mid Output kit with 2 x 150bp paired-end reads.

4.2.5 Bioinformatics

The initial bioinformatic filtering and analysis of the WES dataset was performed by the bioinformatics team at the ACRF Cancer Genomics Facility. Copy-number variation (CNV) analysis was performed, and the analysed data supplied to the author. Variants were called using Genome Analysis Toolkit (GATK) software with variants only called if they belonged to the top five tiers (Tiers 1-5), necessitating a minor allele frequency of <1% in the general population and a pass of the caller internal filters. An excel spreadsheet with the GATK variants was supplied to the author for further analysis.

To filter out potential germline variants, any variant common to all three fractions including the non-malignant T cells were excluded. To do this, "CALLED_NUM" was filtered for "3" (genotypes called in all fractions) and then "REF_HOM_NUM_" was filtered for "1" (only one genotype was homozygous to the reference genome), and finally the T cell fraction was filtered for "HOM_REF" (T cells homozygous to the reference). In this way, variants present in both or only one of the leukaemic clones were identified. To filter for mutations found only in the CD49d+ clone, the T cells and CD49d- fraction were filtered for "HOM_REF", "CALLED_NUM" was filtered for "3" and "REF_HOM_NUM" was filtered for "2". A similar method was applied to find variants solely present in the CD49d- clone. To further narrow down the list of potentially causal variants, only variants predicted to alter proteins were considered (synonymous variants, splice site variants, untranslated region variants, and intronic variants were excluded). Variants were then considered only if their variant allele frequency (VAF) was 45% or greater, implying a heterozygous state in the clear majority of the clonal population. Finally, a manual search on PubMed and the COSMIC database

was performed for the remaining variants to select those that would have a plausible role in driving leukaemogenesis.

4.2.6 Sanger sequencing

Variants in 3 genes (tet methyl cytosine dioxygenase [*TET2*], histone-lysine N-methyltransferase 2D [*KMT2D*] and BAF chromatin remodeling complex subunit BCL11B [*BCL11B*]) identified in the WES data of sorted fractions from patient 6 were confirmed by Sanger sequencing. The author performed confirmation of the *TET2* and *KMT2D* variants. Ms Julia Russell (student of Flinders University, Adelaide, Australia) performed the confirmation of the *BCL11B* variant. Primers are listed in Table 9-1. PCR cycling conditions, clean up, and sequencing are described in Chapter 2. The *NOTCH1* and *BIRC3* mutations in patient 21 (see Table 3-4) were also screened for by Sanger sequencing of amplicons generated from the unsorted, CD49d+ and CD49d- cell fractions as described in section 3.2.5 (using the *NOTCH1* and *BIRC3_4* primer set; see Table 9-1).

4.3 RESULTS

4.3.1 Results of IGHV analysis and SNP microarray: Patient 6

Three fractions were sorted from patient 6: CD49d+ CLL cells, CD49- CLL cells and CD5+ T cells (see Figure 4-1). Purity checks following the cell sort confirmed $\geq 98\%$ purity of each fraction based on 2000 or more events. Genomic DNA was extracted from each fraction and subjected to SNP microarray analysis and targeted IGHV analysis by next-generation sequencing. These results are presented in Table 4-1 and Figure 4-2. Full IGHV sequences are included in the Supplementary Appendix deposited in the Flinders University Library.

Firstly, it is evident that there are two different CLL clones in patient 6 that bear completely different IGHV rearrangements: IGHV4-34*02 and IGHV3-21*02 (see also Table 3-3). Sorting by flow cytometry successfully purified each clone as evidenced by the increased frequency (% reads) of one dominant IGHV rearrangement and absence of the alternate rearrangement in each sorted fraction. CD49d+ CLL cells harboured a mutated IGHV4-34*02 and were enriched for trisomy 12 – 90% of CD49d+ cells contained trisomy 12. Note that the array co-ordinates are consistent with an entire duplication of chromosome 12 in this fraction. On the other hand, CD49d- CLL cells harboured an unmutated IGHV3-21*02 and did not contain any detectable trisomy 12. In addition, the array showed that 20% of this clone carried the high-risk 17p deletion that was not detected on routine FISH of the unsorted diagnostic sample. The T cell fraction unexpectedly demonstrated a smaller gain on chromosome 12 (a portion of the long arm) in approximately 13% of the total population. This is at the limit of sensitivity of this assay (which is quoted at 20% in the diagnostic setting)

Figure 4-1. Flow sorting of patient 6. (A) Lymphocytes were gated by forward (FSC-A) and side (SSC-A) scatter. (B) Singlets were gated by forward scatter height (FSC-H) and area (FSC-A). (C) CLL cells were gated by CD5/CD19 co-expression. (D) CD49d+ CLL cells represent 84.8% of population. CD49d- CLL cells represent 15.2% of CLL cells. (E) Purity check of CD49d+ CLL cells. (F) Purity check of CD49d- CLL cells. (G) Purity check of CD5+/CD19- T lymphocytes.

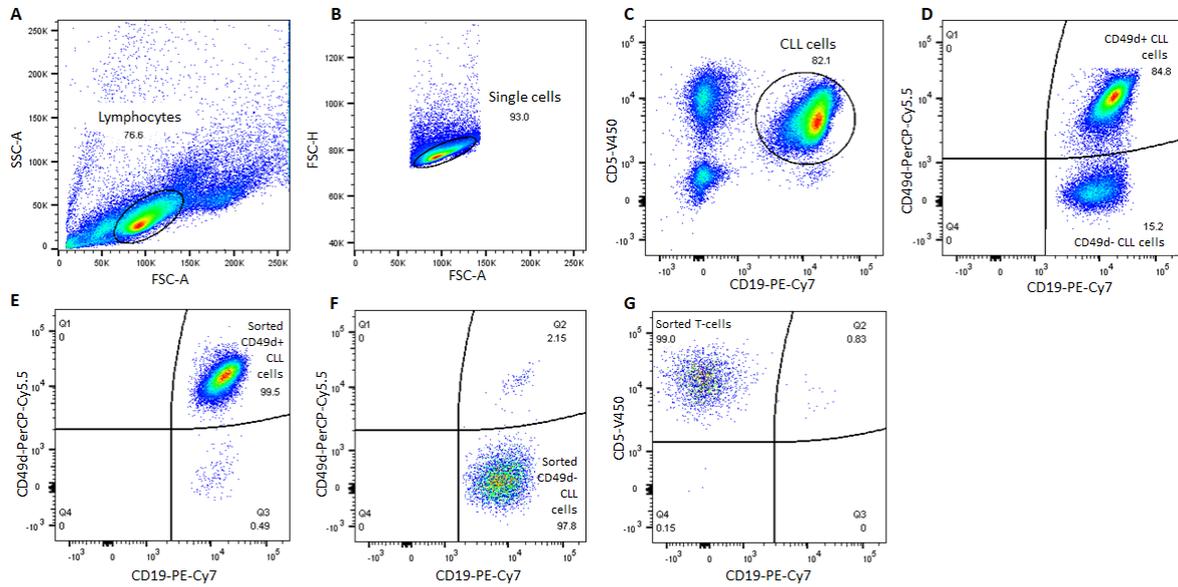
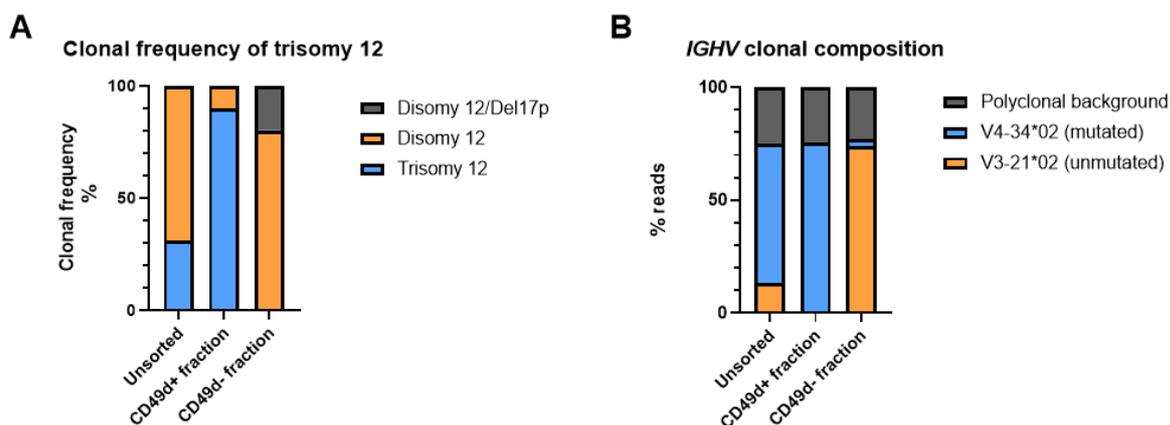


Table 4-1. IGHV and SNP microarray results of the unsorted and three sorted fractions from patient 6. M = mutated IGHV; U = unmutated IGHV. % reads refers to the frequency of the IGHV sequence in the total reads. SNP array summary shows the chromosomal band position altered followed by the exact co-ordinates of the aberration. % refers to the frequency of the alteration within the total cell population.

Patient 6	IGHV gene usage	U/M status	% reads	SNP array summary	Type	%	Size (bp)	
Unsorted sample	IGHV4-34*02	M	62	Not performed (FISH only)				
	IGHV3-21*02	U	13					
Sorted sample	CD49d + CLL clone	IGHV4-34*02	M	75	12p.13.33q24.33(191,619-133,777,645) x 3	Gain	90	133,586,027
	CD49d- CLL clone	IGHV3-21*02	U	72	17p13.3p11.1(8,547-22,242,355) x 1	Loss	20	22,233,809
	T cells		N/A		12q14.2q24.33(64,166,573-133,777,645)x 3	Gain	13	69,611,073

Figure 4-2. SNP microarray and IGHV results of the sorted fractions from patient 6. (A) Clonal frequency (% of cells) with trisomy 12 in the unsorted, CD49d+ and CD49d- CLL cell fractions: trisomy 12 (blue), disomy 12 (orange) and disomy 12 with del(17p) (grey). (B) IGHV clonal composition of sorted fractions. % of top 200 reads with V3-21*02 sequence (orange), V4-34*02 sequence (blue), and polyclonal background (grey).



4.3.2 Results of WES: Patient 6

4.3.2.1 CNV analysis

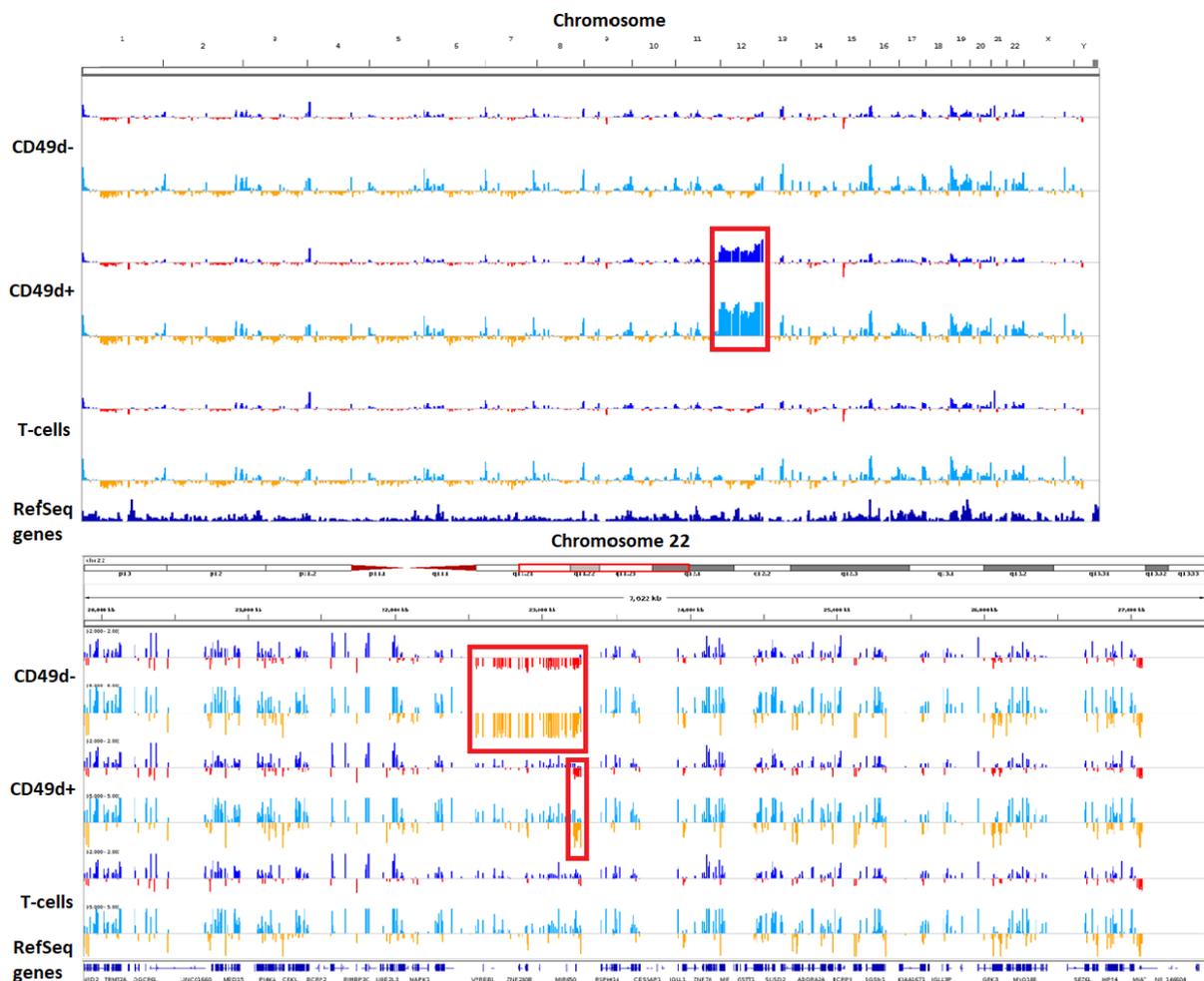
The two separate leukaemic clones along with the T cell fraction were subjected to whole exome analysis. Copy-number variation (CNV) analysis of the WES data confirmed that the trisomy 12 was present in the CD49d+ fraction but not in the CD49d- fraction (see Figure 4-3). There was also an overlapping deletion detected in chromosome 22 in both leukaemic clones, but not the T cells. The deletion breakpoints in the CD49d+ sample are chr22:23,223,572-23,241,802. In the CD49d- sample the 5' breakpoint is not visible and the 3' breakpoint is at chr22:23,247,169. There are two annotated genes in the overlapping region: microRNA 5571 (*MIR5571*) and immunoglobulin lambda like polypeptide 5 (*IGLL5*). The deletion on chromosome 22 was not detected in the SNP arrays of the sorted fractions. No deletion of chromosome 17p was observed in the WES CNV analysis in either fraction, in contrast to the SNP array which detected a sub-clonal deletion of 17p in the CD49d- fraction.

4.3.2.2 Variant analysis

Variant analysis of the WES data identified over 4000 variants in all 3 fractions. The T cells were used as a surrogate germline control and any variant found in all three fractions including the T cells was excluded from further analysis. The number of total variants, those found in common to both CLL clones (but absent in T cells) and those found solely in either the CD49d+ or CD49- CLL clone are presented in Table 4-2. The type of variants detected in each fraction is presented in Figure 4-4. Of

the variants predicted to alter the protein, missense SNVs were the most frequent in each fraction, followed by either in-frame deletions (more common in the CD49d- clone) or frameshift variants (more common in both leukaemic clones together and the CD49d+ clone alone) and finally nonsense SNVs.

Figure 4-3. CNV analysis from WES data of sorted fractions from patient 6. The red boxes highlight the copy number gains (top panel) and losses (bottom panel). The top panel is an overview of the genome demonstrating gains across the entirety of chromosome 12 in the CD49d+ clone alone (highlighted in the red box). The bottom panel shows one area of the long arm of chromosome 22. An overlapping deletion in the CD49d+ and CD49d- clones is demonstrated in the thick red boxes.



Variants were further narrowed down to those predicted to alter proteins: synonymous variants, intron variants, splice site variants and untranslated region variants were all excluded (see Table 4-2 and Figure 4-5). To further refine the list of variants to detect potential driver mutations, those

variants that had a VAF of 45% (i.e., predicted to be present in the heterozygote state in the entire population of cells) or greater were selected for further analysis (see Table 4-2). Variants found in the IGHV and IGLV genes, known to be mutated in CLL, were excluded. This curated list is presented in Table 4-3. A full list of variants found in the CD49d+ clone and CD49- CLL clones are presented in Table 9-10 and Table 9-11 respectively in the Appendix. The list of variants found in common to both leukaemic clones is also presented in Table 9-9 in the Appendix. A full list of all 4216 variants is included in the Supplementary Appendix deposited in the Flinders University library along with this thesis.

Table 4-2. Number of variants detected on whole exome analysis of fractionated samples from patient 6. Protein-altering variants excluded synonymous SNVs, UTR variants, intron variants and splice region variants. Protein-altering variants with a VAF > 45% represent potential drivers. Data was not collected where “-” is listed. Variants of the IGHV and IGLV genes are not included.

Sample	Number of variants	Number of protein-altering variants	Number of protein-altering variants VAF >45*%
<i>All 3 fractions</i>	4216	693	-
<i>Both CLL clones, not T cells</i>	53	9	1
<i>CD49+ CLL clone alone</i>	150	53	15
<i>CD49d- CLL clone alone</i>	136	53	3
<i>T cells</i>	422	25	-

Figure 4-4. Types of variants detected on whole exome analysis of fractionated samples of patient 6. Variants detected in both clones are represented in blue; variants detected in the CD49d+ clone alone are orange; and variants detected in the CD49d- clone alone are grey. Each column represents the absolute number of variants. SNV = single nucleotide polymorphism. UTR = untranslated region.

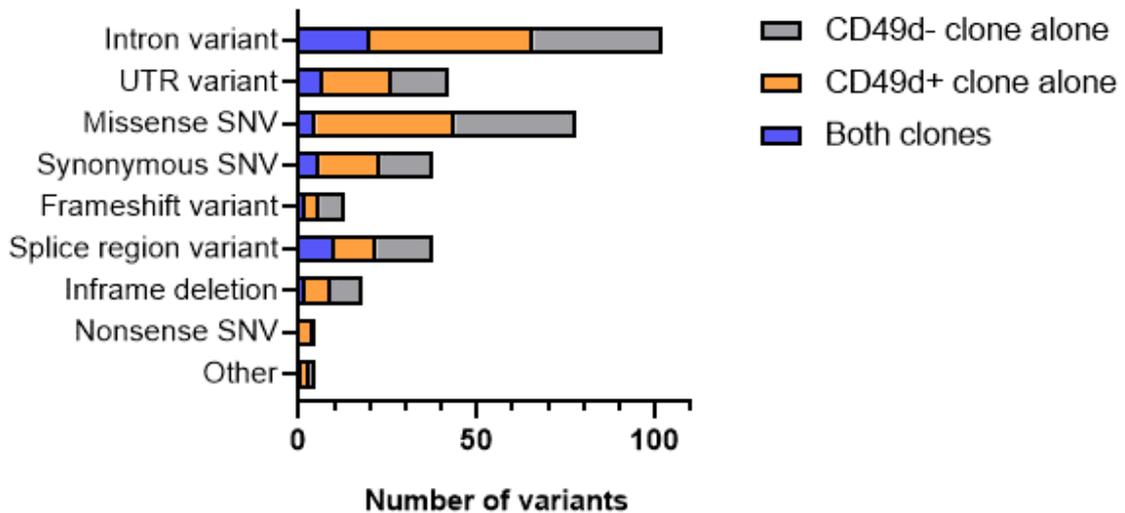


Figure 4-5. Number of protein-altering variants detected on whole exome analysis of fractionated sample from patient 6.

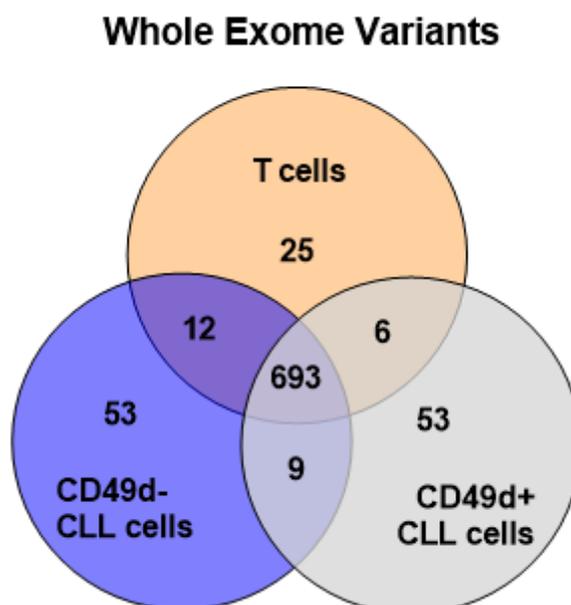


Table 4-3. List of potential driver mutations (with a VAF of greater than 45%) common to both clones (but absent in T cells) and present solely in the CD49d+ or CD49d- CLL clones. aa = amino acid; (+) and (-) refer to the VAF in the CD49d+ and CD49d- clone respectively. Mutations in the immunoglobulin heavy and lambda variable genes, which are known to be mutated in CLL, were excluded. * genes that are pursued further based on their function and possibility of driving leukaemia (see discussion).

Gene	Chromosome band	Type of mutation	Change	Predicted aa change	VAF %	Median depth
Mutations common to both clones						
<i>MDC1</i> *	6p21.33	In-frame deletion	c.4824_4946del	p.Pro1609_Thr1649del	47 (+) 42 (-)	120
Mutations in CD49d+ clone alone						
<i>HOXD12</i>	2q21.1	Missense SNV	c.433G>T	p.Ala145Ser	47 (+) 0 (-)	57
<i>ALDH1L1</i>	3q21.3	Nonsense SNV	c.2143G>T	p.Glu715*	56 (+) 0 (-)	40
<i>IL17F</i>	6p12.2	Missense SNV	c.137G>T	p.Ser46Ile	57 (+) 0 (-)	37
<i>CCDC136</i>	7q32.1	Missense SNV	c.694C>T	p.Arg232Cys	45 (+) 0 (-)	50
<i>RP1L1</i>	8p23.1	Missense SNV	c.6237C>G	p.His2079Gln	50 (+) 0 (-)	156
<i>NRP1</i>	10p11.22	Missense SNV	c.2334G>T	p.Gln778His	55 (+) 0 (-)	44
<i>LZTS2</i>	10q24.31	Missense SNV	c.871G>T	p.Gly291Trp	47 (+) 0 (-)	45
<i>KRTAP5-10</i>	11q13.4	Inframe deletion	c.48_68del	p.Cys17_Gly23del	87 (+) 0 (-)	36
<i>RNF121</i>	11q13.4	Nonsense SNV	c.910C>T	p.Arg304*	52 (+) 0 (-)	49
<i>KMT2D</i> *	12q13.12	Nonsense SNV	c.15256C>T	p.Arg5086*	67 (+) 0 (-)	51
<i>BCL11B</i> *	14q32.2	Missense SNV	c.1387G>A	p.Ala463Thr	53 (+) 0 (-)	242
<i>SALL4</i>	20q13.2	Missense	c.986G>A	p.Arg329His	47 (+)	234

Gene	Chromosome band	Type of mutation	Change	Predicted aa change	VAF %	Median depth
		SNV			0 (-)	
<i>ZNF74</i>	22q11.21	Missense SNV	c.986G>A	p.Arg329His	47 (+) 0 (-)	234
<i>TRIOBP</i>	22q13.1	In-frame deletion	c.2060_2203del	p.Pro687_Ser734 del	66 (+) 0 (-)	69
<i>CCNB3</i>	Xp11.22	Missense SNV	c.2497T>A	p.Leu833Met	61 (+) 0 (-)	38
Mutations in CD49d- clone alone						
<i>SF3B1</i>	2q33.1	Missense SNV	c.1866G>T	p.Glu622Asp	0 (+) 49 (-)	129
<i>TMEM14</i> 5	19q13.2	Frameshift	c.1462_1465del TTTT	p.Phe488fs	0 (+) 100 (-)	7
<i>SHANK3</i>	22q13.33	Missense SNV	c.2854G>A	p.Ala952Thr	0 (+) 52 (-)	48

The type of IGHV rearrangement and mutational status of each clone was confirmed by WES: the CD49d+ clone has 5 missense SNVs detected in the IGHV4-34 allele and the CD49d- clone had 1 missense SNV detected in the IGHV3-21 allele (see Table 9-11 in the Appendix). The IGLV gene on chromosome 22 is also a known area of somatic hypermutation in B cells and both clones contained multiple different missense SNVs at this locus: the CD49d+ clone had 7 missense SNVs in IGLV3-1 allele and the CD49d- clone had 6 missense SNVs in IGLV4-60 allele.

The mutation in *SF3B1*, a known driver of CLL, was detected by WES of the CD49d- clone but it is noteworthy that this mutation was not detected in the unsorted sample (see previous Chapter). Given the evidence that patient 6 had developed 2 independent CLL clones, it was hypothesised that clonal haematopoiesis may have been present. Therefore, the full list of variants was interrogated for mutations in genes known to be associated with clonal haematopoiesis of indeterminate potential (CHIP): the DNA methyltransferase 3 alpha (*DNMT3A*), *TET2*, ASXL transcriptional regulator 1 (*ASXL1*) and Janus kinase 2 (*JAK2*) genes. The incidence of CHIP increases with age and predisposes the carrier to haematological malignancies. A heterozygous missense SNV in *TET2* (c.3609C>G; p.Ser1203Arg) was found common to all three fractions including the non-leukaemic T cells. The VAF

was 46%, 50% and 48% in the CD49d+ clone, CD49d- clone, and T cells respectively. The median read depth was 95. No mutations were detected in *DNMT3A*, *ASXL1* or *JAK2*.

Three mutations of interest alongside the *TET2* variant were selected from the list of potential drivers in Table 4-3 to pursue further. They included the 123bp in-frame deletion in *MDC1* common to both clones and two putative novel drivers in the trisomy 12, CD49d+ clone: the *KMT2D* nonsense SNV located on chromosome 12 and the missense SNV in *BCL11B* on chromosome 14. These potential driver mutations were chosen based on current knowledge of the function of the genes, which is explored further in the discussion. A summary of the immunogenetic features of the two different clones is presented in Table 4-4.

4.3.3 Confirmation of WES variants

To determine if the *TET2* mutation was of germline origin or acquired in a haematopoietic stem cell prior to B or T-lineage commitment, germline DNA was extracted from a historical skin biopsy specimen (the patient was deceased at the time of this research and no other source of germline material was available). The skin biopsy was of a non-malignant lesion and examination by light microscopy of the haematoxylin and eosin-stained slides confirmed no gross infiltration by CLL lymphocytes. Sanger sequencing analysis identified the presence of the *TET2* mutation within the DNA extracted from the skin biopsy, confirming it to be of germline origin (see Figure 4-6). The *KMT2D* and *BCL11B* mutations were confirmed to be present in the CD49d+ CLL clone and the unsorted leukaemic sample, but not in the germline or CD49d- CLL clone (see Figure 4-6).

4.3.4 Post-treatment changes: Patient 6

Patient 6 was treated with two cycles of chlorambucil in 2013 following diagnosis, and then with ibrutinib from 2015. Only one further sample was available for analysis, taken 6 months' post-ibrutinib commencement whilst continuing treatment. The CD49d+ clone had decreased from 84.8% to 37.3% of the total CLL population) and the CD49d- clone had increased in size from 15.2% to 62.7% (see Figure 4-7). The absolute peripheral blood lymphocyte count was $38 \times 10^9/L$ at diagnosis with 81% CD5/19+ cells: the absolute count of the CD49d+ and CD49d- CLL cells was $26.1 \times 10^9/L$ and $4.7 \times 10^9/L$ respectively. The absolute peripheral blood lymphocyte count after six months of ibrutinib therapy measured $13 \times 10^9/L$ with 68% CD5/19+ cells: the absolute count of the CD49d+ and CD49d- CLL cells was $3.3 \times 10^9/L$ and $5.5 \times 10^9/L$ respectively. As such, the CD49d+ clone had diminished in size but the CD49d- clone was higher in the peripheral blood.

Table 4-4. Summary of immunogenetic features of two different CLL clones in patient 6. SNP = single nucleotide polymorphism; arr[GRCh37] refers to microarray and genome build.

	CD49d expression		Karyotype (SNP microarray)		IGHV		Somatic mutations
	% positive cells	CD49d status			Gene usage	Mutational status	
Clone 1	99.5	Positive	arr[GRCh37] 12p13.33q24.33(191,619- 133,777,645)x3	Trisomy 12	IGHV4- 34*02	Mutated	<i>MDC1</i> c.4824_4946del; p.P1609_T1649del
							<i>BCL11B</i> c.1387G>A; p.A463T
							<i>KMT2D</i> c.15256C>T; p.R5086*
Clone 2	2.2	Negative	arr[GRCh37] 17p13.3p11.1(8,547- 22,242,355)x1	Del(17p)	IGHV3- 21*01	Unmutated	<i>MDC1</i> c.4824_4946del; p.P1609_T1649del
							<i>SF3B1</i> c.1866G>T; p.E622D

Figure 4-6. Sanger sequencing confirmation of variants detected by WES on sorted and germline DNA from patient 6. The mutations are listed at the top and highlighted in the red boxes.

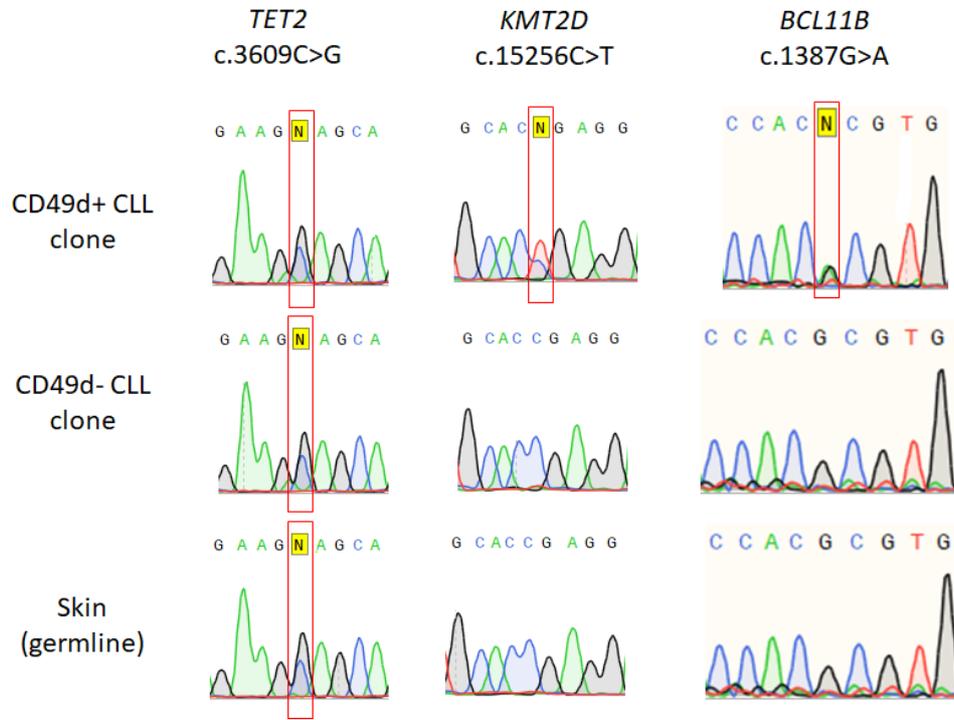
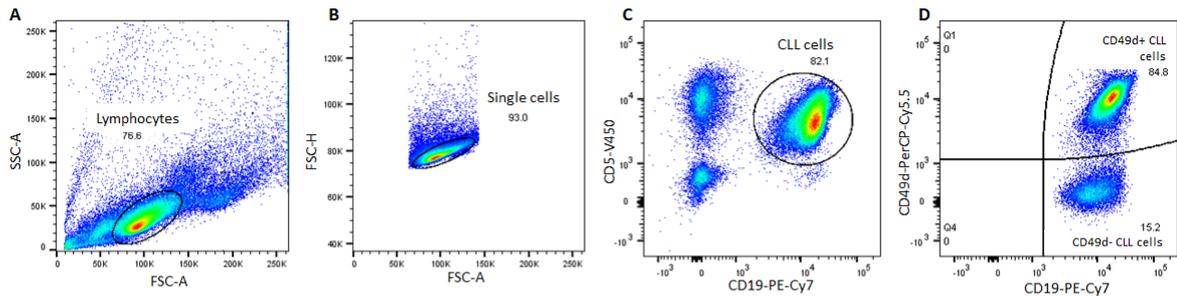
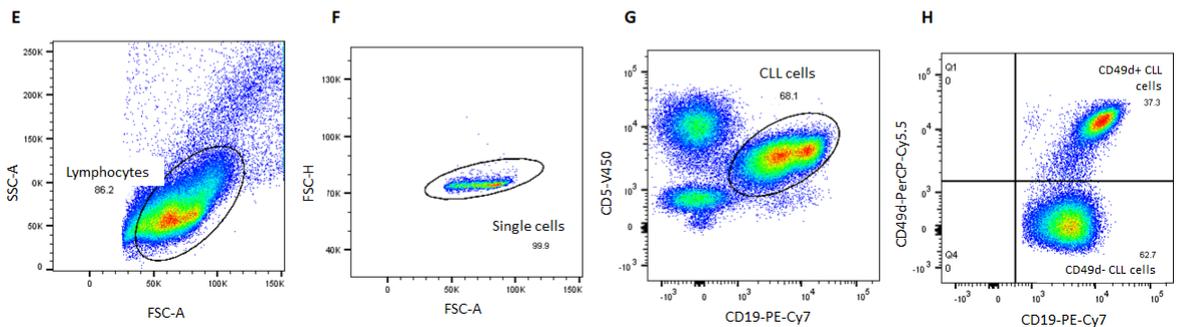


Figure 4-7. Post-treatment change in CD49d+ CLL cells in patient 6. The top panel shows the diagnostic panel and the bottom panel shows the results 6 months' on ibrutinib. (A),(E) Lymphocytes are gated on forward (FSC-A) and side (SSC-A) scatter. (B),(F) Single cells are gated on forward scatter height (FSC-H) and area (FSC-A). (C),(G) CLL cells co-express CD5/CD19. (D) CD49d+ CLL cells are 84.8% at diagnosis. (H) CD49d+ CLL cells are 37.3% post-6 months of ibrutinib therapy.

Patient 6: diagnosticsample



Patient 6: 6 months' post-ibrutinib sample



4.3.5 Results of SNP microarray and targeted mutation analysis: Patient 21

Two fractions were sorted from patient 21: CD49d+ CLL cells (1.4×10^6 cells) and CD49- CLL cells (1.1×10^6 cells) (see

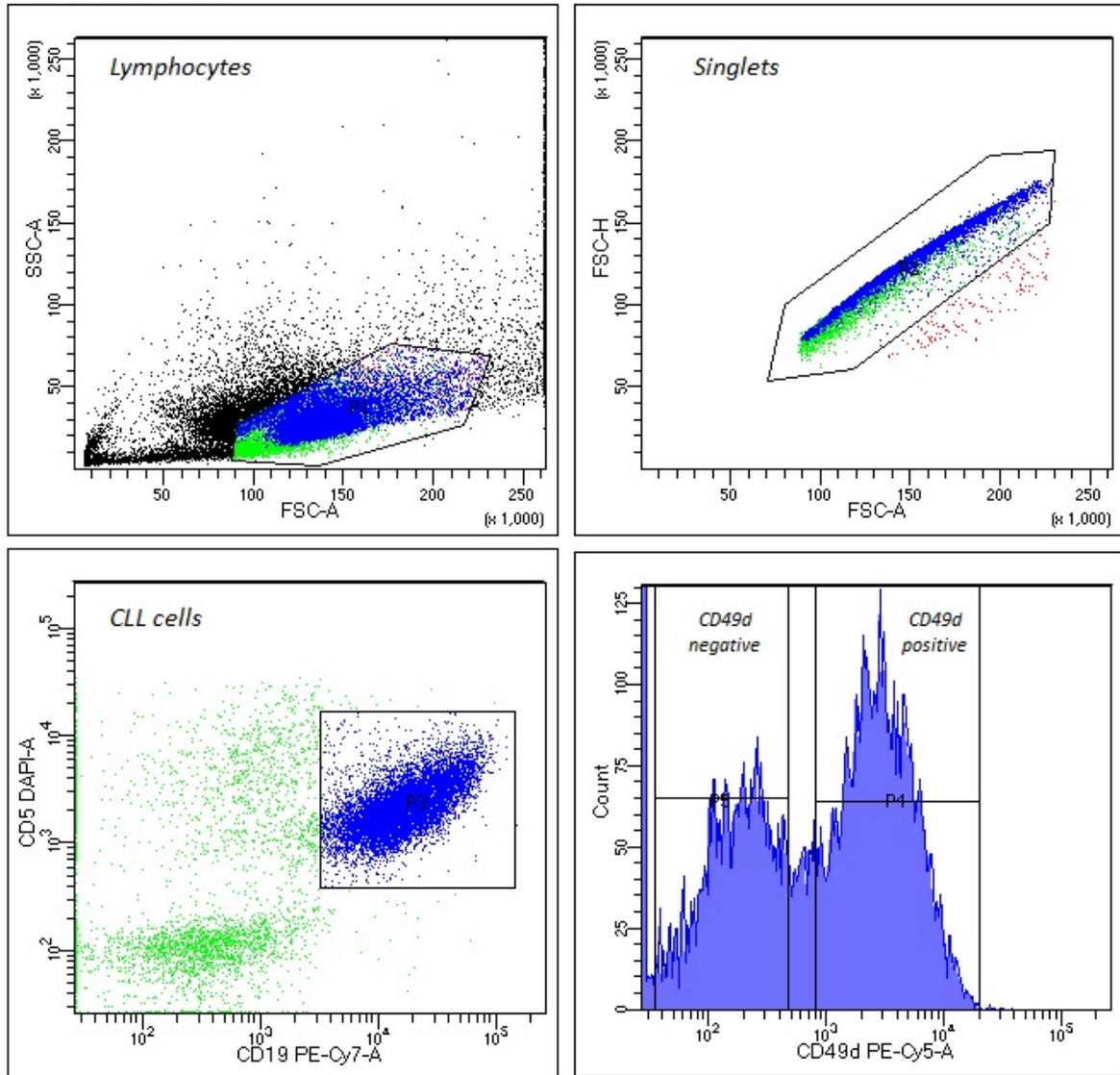
Figure 4-8). Cell viability was not formally determined by flow cytometry on the sorted cells, however, lymphocytes represented 59.6% of all events prior to the sort. Cell viability post-thaw was measured at 82% using a trypan-blue exclusion assay. Purity checks following the cell sort confirmed $\geq 99\%$ purity of each fraction based on 1000 events (see Figure 4-9). Genomic DNA was extracted from each fraction and subjected to SNP microarray analysis. These results are presented in Table 4-5. In brief, both fractions as well as the unsorted sample contained trisomy 12 at a frequency of 70-80%. The unsorted and CD49d+ fractions also contained an additional abnormality in loss-of-heterozygosity at 11q. This region contains the *ATM* and *BIRC3* genes.

Furthermore, as patient 21 was known to have both a *BIRC3* and *NOTCH1* mutation (see Chapter 3), PCR and Sanger sequencing for the known mutations was performed on the CD49d+ and CD49d- sorted fractions to determine which fractions harboured the mutations. The *NOTCH1* c.7541_752delCT variant was detected in all three fractions at a similar allele burden of approximately 50% (see Figure 4-10). The *BIRC3* c.1298_delAAAinsA variant was detected in the CD49d+ fraction as well as unsorted DNA, however, the CD49d- fraction demonstrated the wildtype *BIRC3* sequence (see Figure 4-10). Furthermore, the mutation appears to be enriched in the CD49d+ fraction compared to unsorted DNA based on peak heights of the chromatograms at the start of the frameshift.

Targeted IGHV analysis was not performed on the sorted fractions of patient 21 as only one IGHV rearrangement was detected in the unsorted sample. Whole-exome sequencing and RNAseq was not performed on patient 21.

Figure 4-8. Flow Sort of Patient 21. Lymphocytes were gated based on their forward (FSC) and side scatter (SSC) properties. Doublets were excluded. CLL cells were gated based on their co-expression of CD5 and CD19 (population P3). CD49d positive CLL cells (P4) and CD49d negative CLL cells (P5) were sorted using the CD49d expression histogram. The sorted CD49d positive and CD49d negative cells represented 52.5% and 34.4% of the total CLL population respectively.

Patient 21: Sort



Population	#Events	%Parent	%Total
All Events	22,759	####	100.0
P1	13,561	59.6	59.6
P2	13,401	98.8	58.9
P3	9,832	73.4	43.2
P4	5,160	52.5	22.7
P5	3,384	34.4	14.9

Figure 4-9. Purity check from flow sort of patient 21. An aliquot of the sorted fractions was re-examined on the flow cytometer. The CD49d negative fraction is shown on the left panel (836/837; 99.9% of events fall in the CD49d negative gate). The CD49d positive fraction is shown on the right panel (984/987; 99.7% of events fall in the CD49d positive gate).

Patient 21: Sort Purity Check

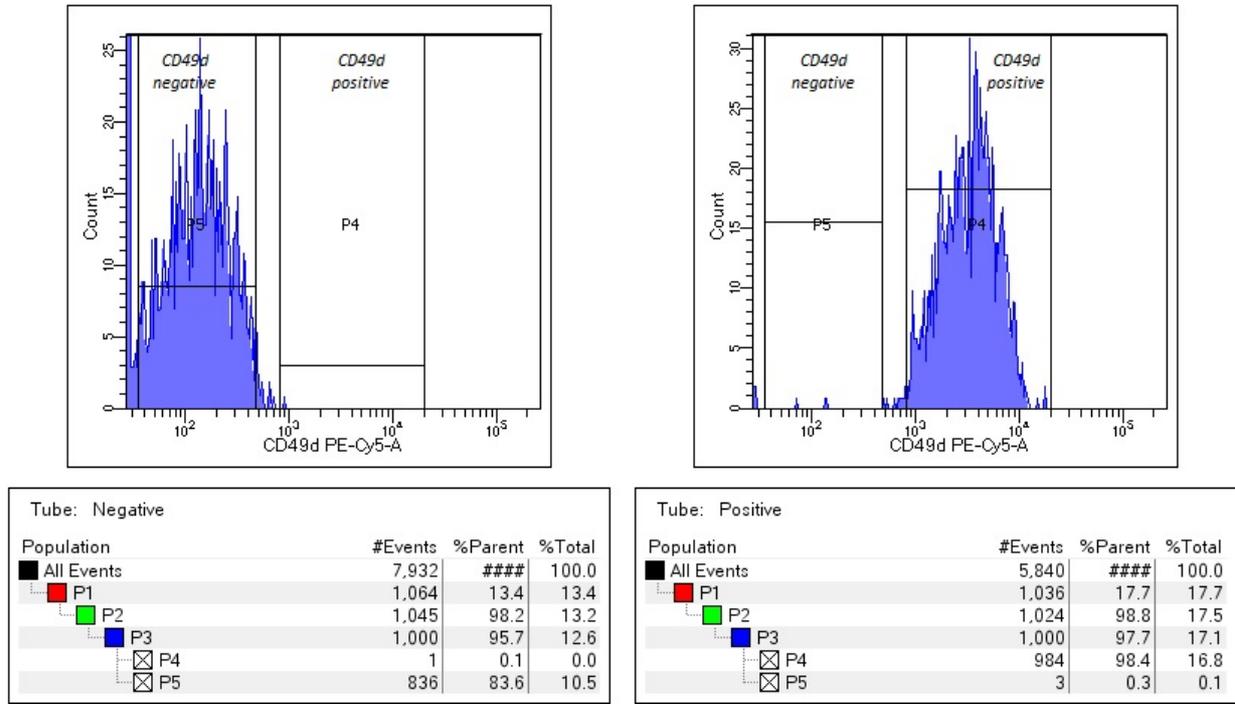
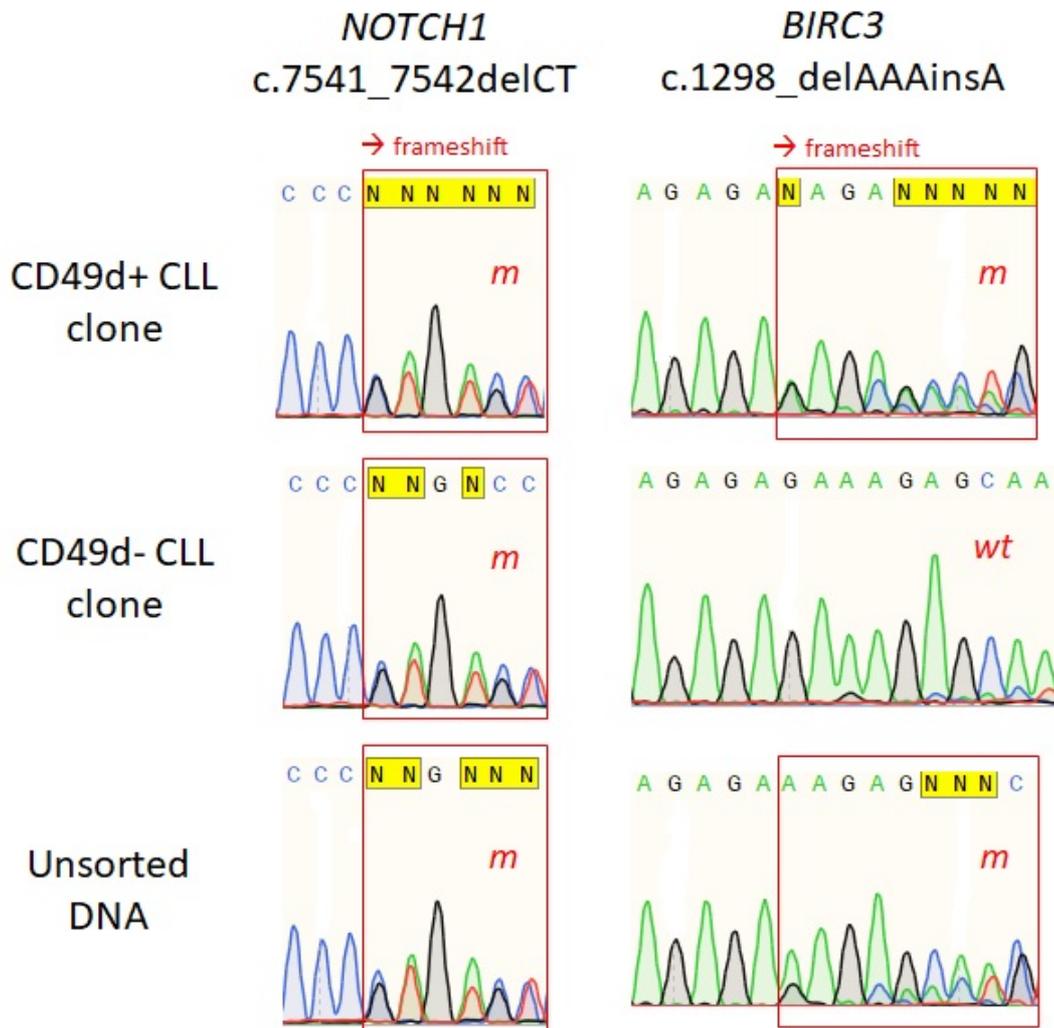


Table 4-5. SNP array results of sorted fractions from patient 21. hmz = homozygous; LOH = loss of heterozygosity.

Patient 21		SNP array summary	Type	%	Size (bp)
Unsorted sample		11q13.2q25(66,766,754-134,934,063) x 2 hmz	LOH	<20	68,167,309
		(12) x 3	Gain	70	-
Sorted sample	CD49d+ CLL clone	11q12.1q25(59,823,561-134,934,063) x 2 hmz	LOH	20	75,110,502
		(12) x 3	Gain	80	-
	CD49d- CLL clone	(12) x 3	Gain	80	-

Figure 4-10. NOTCH1 and BIRC3 Sanger sequencing results on unsorted and sorted samples from patient 21. The NOTCH1 c.7541_7542delCT frameshift mutation is detected in both CD49d+ and CD49d- CLL cells at a variant frequency of approximately 50%. The BIRC3 c.1298_delAAAinsA frameshift mutation is detected in unsorted CLL cells (with an approximate allele burden of 33%) and is enriched in the CD49d+ fraction (allele frequency approximately 50%). The BIRC3 mutation is not detected in the CD49d- clone. m= mutant; wt = wildtype.



4.4 DISCUSSION

CLL is a heterogeneous disease with a complex clonal architecture. Bimodal expression of CD49d was observed in 2 of 22 trisomy 12 CLL patients and flow sorting based on this marker successfully purified two unique leukaemic clones in patient 6 and two subclones in patient 21. This allowed for direct comparisons between (sub)clones within individual cases of CLL. Each case will be discussed individually at first, and then common themes identified from both cases and their implication for the evolution of trisomy 12 will be discussed in the concluding remarks.

4.4.1 Patient 6

Patient 6 carried two unique leukaemic clones with discordant characteristics separated by CD49d expression prior to receiving any treatment. The CD49d- disomy 12 clone was enriched for high-risk genetic features including an unmutated IGHV sequence and the CD49d+ trisomy 12 had low risk features including a mutated IGHV sequence. The CD49d- clone had a heterozygous *SF3B1* mutation as well as an unmutated configuration of the IGHV. Both features are indicative of high-risk CLL; therefore, it is particularly relevant to note that the *SF3B1* mutation was not detected in the unsorted sample. Even at subclonal levels, *SF3B1* mutations are of prognostic importance and correlate with inferior overall survival [191]. The other high-risk lesion of del(17p) was also detected at low frequency (20%) in this clone. It was not detected by FISH in the unsorted diagnostic sample. This is noteworthy as even subclonal *TP53* (located on 17p) aberrations predict for adverse clinical outcomes and are of clinical relevance in the choice of therapy offered patients [183]. Interestingly, no concomitant *TP53* mutation was detected, which is the case in around 80% of del(17p) CLL [192]. Therefore, the overall clinical significance of this deletion is unclear within this patient as he presumably has one functioning copy of *TP53*, however, this does not reduce the importance of detecting these prognostically-relevant alterations that may be present at low levels within minor subclones. Patient 6 would have traditionally been ascribed an “intermediate prognosis”, however, the presence of at least the *SF3B1* mutation would have upgraded him to a more high-risk group and may have warranted more intensive monitoring. The presence of the low level del(17p) (albeit without concomitant *TP53* disruption on the remaining allele) would also be an argument to avoid upfront chemotherapy.

Trisomy 12 was seen at low level (13%) in the T cell fraction of patient 6 by SNP array but not in the CNV analysis from the WES data. It is thought this is a technical issue and that there is no true trisomy 12 in the T cell fraction. The low-level trisomy could be explained by contamination of the T cell sample with CD49d+ CLL cells (although by no more than 2% of the sample based on the flow purity check), however, it would be expected that the array features would be identical (that is, the whole of chromosome 12 was triplicated, rather than just part of the long arm as seen in the T cell fraction). A more widespread mosaicism of trisomy 12 in other non-malignant tissues was thought very unlikely.

The CNV analysis of the whole-exome data confirmed the presence of trisomy 12 solely within the CD49d+ clone. It also revealed an overlapping deletion on chromosome 22 encompassing the *IGLV* gene locus. This deletion was not observed in the SNP microarray despite being of an adequate size (around 1000bp in length) for detection with this technology. It may be that the deletion was present at sub-clonal levels below the level of sensitivity of the SNP array assay,

however, this assay is generally more robust for detecting CNV alterations than WES (at least at the current time), where different computational methods to discover CNV have different biases and are not comprehensive [193]. Deletions at the 22q locus have been reported in CLL using array-based technology [194-196], however, their clinical relevance is uncertain and it is probable that the detected deletions are merely a consequence of the normal process of immunoglobulin lambda variable gene (IGLV) rearrangement and have no consequences for disease biology [197, 198]. Certainly, patient 6 has lambda light-chain restricted disease and several variants in IGLV-3 detected in the WES analysis.

The WES CNV analysis did not detect the clinically relevant sub-clonal deletion 17p of the CD49d- clone observed with the SNP microarray. The bioinformatic WES CNV analysis has several inherent limitations in detecting somatic CNV alterations in mosaic tumour samples [199] and is not as robust as SNP microarray for this purpose as yet. Indeed, it was not the primary focus of the WES analysis in this case and supplemented rather than replaced the microarray data.

The whole-exome analysis of sorted fractions from patient 6 also revealed over 4000 variants. The sheer number detected is noteworthy compared to the average number of variants in a CLL exome, which has been reported as 21.5 [83]. The reason is not clear but may be related to the germline *TET2* mutation. *TET2* is an epigenetic regulator involved in DNA methylation and germline *TET2* variants have been hypothesised to predispose individuals to myeloid malignancies in later life although have not been investigated in CLL [200]. An interesting additional experiment would be to analyse the methylome of this patient and this is discussed in further detail later. Further exploration of predisposing germline variants is beyond the scope of this thesis, although it raises intriguing questions about the underlying reasons for the familial predisposition to CLL and the relationship with clonal haematopoiesis, especially in patients who are diagnosed with CLL in their young adult life.

The large number of variants identified also highlights the intrinsic difficulties in comparing different patients with CLL (a disease of the elderly) and understanding key differences between the different subtypes of CLL. The usual approach has been to compare large numbers of patients with different subtypes – for example, several hundred patients with or without trisomy 12 CLL are interrogated. This approach, however, whilst attempting to control for background genetic heterogeneity with large numbers, means that biologically relevant subtle effects or those confined to a subset of cases are not appreciated. Certainly, the key drivers of trisomy 12 CLL have not been identified in this way to date. Whilst this is an individual case (patient 6), its power lies in being able to control for that variability and compare a trisomy 12 and non-trisomy 12 CLL within the same

patient on the same genetic background. Potential drivers of trisomy 12 CLL, therefore, are identified as residing solely in that clone and not in the disomy 12 clone. In this way, two variants of interest in *KMT2D* and *BCL11B* were identified in the trisomy 12 clone alone and represent putative trisomy 12 driver mutations. These are discussed below, as is the variant in *MDC1* which was common to both clones and may represent an earlier “priming” event in the development of CLL in this individual.

A 123bp in-frame deletion of mediator of DNA damage checkpoint 1, *MDC1*, was found common to both clones (c.4824_4946del; p.Pro1609_1649del). Its presence in both clones and absence in the T cell fraction suggests the presence of a pre-leukaemic progenitor cell prior to IGHV rearrangement but following B-lineage commitment. This adds to the field of evidence about the cell of origin of CLL: it is not necessarily a fully differentiated B cell and CLL may arise much earlier in haematopoiesis [201, 202]. *MDC1* plays an important role in the response to DNA damage, recruiting repair and signalling proteins to the sites of double-strand breaks and is necessary for activation of cell cycle checkpoints [203]. *MDC1* relocates to foci of DNA damage and interacts with phosphorylated histone H2AX, facilitating recruitment of ATM [204]. The deletion observed in patient 6 is predicted to result in loss of part of the PST (proline-serine-threonine repeat motif) domain of the protein whose role has only recently been elucidated [204]. This domain is critical for non-canonical double-stranded DNA repair in the absence of phosphorylated histone H2AX and directly interacts with chromatin. Lack of the PST domain results in increased sensitivity to the effects of ionizing radiation but only in the absence of H2AX. Recurrent mutations in *MDC1* have not been reported in CLL. One patient demonstrated a variant in *MDC1* (c.3560C>G) in the WES analysis of 538 CLL patients [83]. The deletion observed in patient 6 is not in the COSMIC database either, however, point mutations of *MDC1* are reported in 12 of 3517 (0.34%) of “haematopoietic and lymphoid” cancers and do occur within the PST as well as other protein domains. A smaller overlapping deletion of *MDC1* (p.Glu1591_Pro1631) has been reported in one case of lung cancer (COSM6471062).

The role of *MDC1* as a potential tumour suppressor has not been well studied although it has been shown to be aberrantly reduced in subsets of breast and lung carcinomas, and *MDC1* knockout mice display a higher frequency of tumours [205, 206]. It has not been investigated in CLL or other haematopoietic neoplasms. If disruption of *MDC1* were driving CLL in a putative progenitor cell in the case of patient 6, there would necessarily be haploinsufficiency of this presumed loss-of-function mutation resulting absence of cell-cycle arrest despite DNA damage. Alternatively, disruption of *MDC1* may have an additive effect with other disrupted genes in the DNA damage response pathway, for example, *ATM*. *ATM* is known to be a key molecule in CLL pathogenesis. It could be

speculated that if both ATM and MDC1 were affected, the double-stranded DNA repair (DDR) pathway and cell cycle check points would not function correctly. This would presumably lead to the accumulation of DNA damage, an unstable genome and explain the huge number of genetic variants observed in this patient. However, there is no strong evidence that ATM or other members of the DDR pathway are affected in patient 6. The WES data did reveal three variants in *ATM* that were common to all fractions including T cells, suggesting either germline origin or clonal haematopoiesis. The variants were not clearly pathogenic, residing in either intronic or splice site regions. Further functional evidence is clearly needed to support the hypothesis of MDC1 as a cancer driver in CLL, and exploration of the relationship between MDC1 and ATM-deficient CLL is another avenue for future research.

The other two variants of interest were unique to the trisomy 12 clone. The first of these mutations is in *KMT2D* (lysine methyltransferase 2D), also known as *MLL2* (Myeloid/Lymphoid or Mixed-Lineage Leukaemia 2). This gene is located on the long arm of chromosome 12 and the nonsense mutation observed (c.15256C>T; p.Arg5086*) had a VAF of 67%. This suggests it was acquired in one copy of chromosome 12 that was duplicated and potentially offered a survival advantage (although it is acknowledged that it could still represent a passenger mutation). *KMT2D* encodes an epigenetic regulator, specifically a histone methyltransferase which targets histone h3 lysine 4 (H3K4), a mark associated with gene activation [207]. The enzymatic activity of KMT2D SET domain is responsible for this function, and this domain is predicted to be entirely lost with the introduction of an early stop codon in the observed mutation. This is of interest in trisomy 12 CLL which has a unique epigenetic signature [208, 209]. However, a survival advantage conferred to cells with two copies of a *KMT2D* variant would suggest a gain-of-function rather than a loss-of-function mutation if there remains a normal allele on one of the three copies of chromosome 12. Nonetheless, *KMT2D* mutations are recurrent in diffuse large B cell lymphoma and follicular non-Hodgkin lymphoma, two other B cell malignancies, and disruption of *KMT2D* promotes lymphomagenesis in a murine model and leads to diminished global H3K4 methylation in DLBCL cells [210]. It appears that *KMT2D* has diverse roles depending on cell type and the specific transcription factors that recruit it to transcriptional enhancers, and that defective enhancer regulation contributes to tumourigenesis [207]. It would be interesting to ascertain if each clone from patient 6 harboured a different epigenetic signature, however, this was not pursued in the first instance – rather, expression levels of *KMT2D* targets was explored (see Chapter 5).

The second potential driver present solely in the trisomy 12 clone was a mutation in *BCL11B*, B cell lymphoma/leukaemia 11B (c.1387G>a; p.Ala463Thr). *BCL11B* encodes a transcriptional repressor, although its target genes and function has not been fully defined. Loss-of-function of

BCL11B is associated with T-acute lymphoblastic leukaemia [211]. It also has high sequence homology (particularly within the critical zinc-finger domain of the protein) to BCL11A, the dysregulation of which is implicated in B cell lymphoproliferative disorders [212]. Furthermore, of interest in trisomy 12 CLL, it is also known to modulate expression of CCR7 (an integrin known to be overexpressed in trisomy 12 CLL) and is adjacent to the IGH locus at 14q32.3. Del(14q) is also associated with trisomy 12 CLL [142]. The observed mutation is predicted to result in loss-of-function of BCL11B by disrupting DNA-binding through the zinc-finger domain of the protein. This is proposed to have a dominant negative effect (as previously reported in a heterozygous germline mutation of the gene resulting in severe-combined immunodeficiency) with potential for resultant loss of repression of unknown target genes on chromosome 12, contributing to leukaemogenesis [213].

Altogether, the hypothesised origins of CLL in patient 6 are presented in Figure 4-11 and have been published [214]. The CLL clones harbour completely different IGHV rearrangements but contain common mutations that were absent in the T cells. This suggests a common ancestor early in haematopoiesis prior to IGHV rearrangement but following T-lineage commitment. This also implies that the patient was primed for the development of CLL – any number of clones could have developed, but, perhaps, environmental pressures only allowed for the development and expansion of two such clones. Furthermore, trisomy 12 was present in around 90% of CD49d+ clone, suggesting that it was an early clonal feature and is in keeping with the hypothesis of trisomy 12 being a founder lesion in the development of CLL.

There was only one serial sample from patient 6, collected 6 months' after ibrutinib commencement whilst on continuing therapy. The CD49d+ clone was reduced in size in terms of both absolute number and proportion, now only accounting for 37.3% of the total CLL population ($4.7 \times 10^9/L$) as opposed to 84.8% ($26.1 \times 10^9/L$). It is possible that the CD49d+ clone was more sensitive to ibrutinib therapy, and more cells died directly due to release from the protected and proliferative lymph node microenvironment, or the CD49d- clone was more sensitive to treatment and was released more readily from the lymph node microenvironment by ibrutinib leading to a higher percentage and absolute number of CD49d- CLL cells assayed in the peripheral blood. In the latter situation, CD49d+ clone may be more resistant to the microenvironmental disruptive effects of ibrutinib due to the strong expression of the integrin, but no samples were available from lymph node specimens to assess this possibility. Alternatively, it is also possible that *in vivo* interruption of BCR signalling directly led to decreased CD49d expression as has been previously reported after 30 days of ibrutinib treatment [161]. The differential *in vivo* effects of ibrutinib on the two clones is of clinical importance, however, *in vitro* studies on the cell fractions with ibrutinib was not pursued due to limited patient samples.

4.4.2 Patient 21

Whilst patient 6 harboured two completely different CLL clones, patient 21 had two clones that were more closely related with the CD49d+ clone appearing to be a daughter subclone that subsequently expanded to over half the entire population. The hypothesised clonal evolution of patient 21 is presented in Figure 4-12. Both the CD49d+ and CD49d- clones harboured the same IGHV rearrangement and unlike patient 6, CD49d expression did not separate a disomic and trisomic 12 clone. Both populations contained trisomy 12, demonstrating that despite the known association, CD49d expression is not invariably linked to the presence of trisomy 12 and other factors are contributing to its relative upregulation. This is in keeping with the variable prognosis of trisomy 12 CLL despite the poor prognostic impact of CD49d expression. It would be of interest to examine the methylation patterns of the *ITGA4* promoter in each of the clones, as hypomethylation of the promoter has previously been associated with increased CD49d expression [155]. Again, the high clonal frequency of +12 suggested its acquisition early during CLL development. Both clones contained the *NOTCH1* mutation, however, only the CD49d+ clone acquired loss of heterozygosity at 11q and a *BIRC3* mutation. *BIRC3* is located within the area of LOH, meaning there is a homozygous mutation and complete disruption of *BIRC3* within the subclone.

4.4.3 Conclusion

In summary, two different cases of CLL with bimodal CD49d expression were interrogated and highlighted the inherent clonal complexity and heterogeneity of the disease, particularly trisomy 12 CLL. Both cases demonstrated a high clonal frequency of trisomy 12 in keeping with its acquisition early in CLL development, or at least suggesting chromosomal duplication was so advantageous, that clones harbouring it were positively selected. Also, the presence of CLL lesions in B cell progenitors was evident and raises questions about the potential contribution of clonal haematopoiesis and germline variants to the development of CLL. In the case of patient 21, it was also demonstrated that CD49d expression is not invariably linked to trisomy 12 and other mechanisms are contributing to its relative upregulation. Patient 6 was shown to have a truly biclonal case of CLL and allowed for the identification of three potential driver mutations of CLL. Two of the mutations were unique to the trisomy 12 CLL clone. If the identified mutations are truly drivers of CLL, it is unclear why they have not previously been identified in large WES studies of CLL, but it may be due to technical reasons such as lower sequencing coverage of these loci in the larger studies or more stringent bioinformatic filtering. Furthermore, the identification of these variants in a solitary case of CLL is of limited importance if not observed in a larger sample size and is not proof of pathogenicity. Initially, a targeted look at these mutations in a larger cohort of CLL patients would be necessary. Alternatively, functional studies (including investigation of changes to the epigenome potentially induced by *MDC1*

and *KMT2D* mutations) are required. This leads on to the following chapter in which a whole transcriptome analysis of the two clones in patient 6 was performed with the aim to identify changes in gene expression in trisomy 12 CLL, and more specifically changes that may be induced by the identified variants. Finally, tracking the changes in clonal composition of these cases over time with therapeutic pressure would be of value and may expose differential sensitivities of the two clones to treatment.

Figure 4-11. Proposed evolution of CLL in patient 6 (see text). There is a germline *TET2* variant in haematopoietic stem cells (HSCs). Further genetic lesions (red arrow) arise in progenitor cells, predisposing to leukaemia. *MDC1*, mediator of damage cell cycle checkpoint 1, is mutated in both clones but not T cells. Pre-leukaemic cells undergo *IGHV* rearrangement and utilise different V segments of the gene (V3 or V4). Further genetic lesions (red arrows) give rise to CLL. In the orange clone, no somatic hypermutation (SHM) of the *IGHV* is evident (U-CLL), an *SF3B1* mutation is present at high frequency and a small 17p deleted subclone is identified (dark orange). In the grey clone, novel drivers *KMT2D* and *BCL11B* are mutated, the *IGHV* is hypermutated (M-CLL), trisomy 12 is present at high frequency and CD49d is expressed. LN = lymph node.

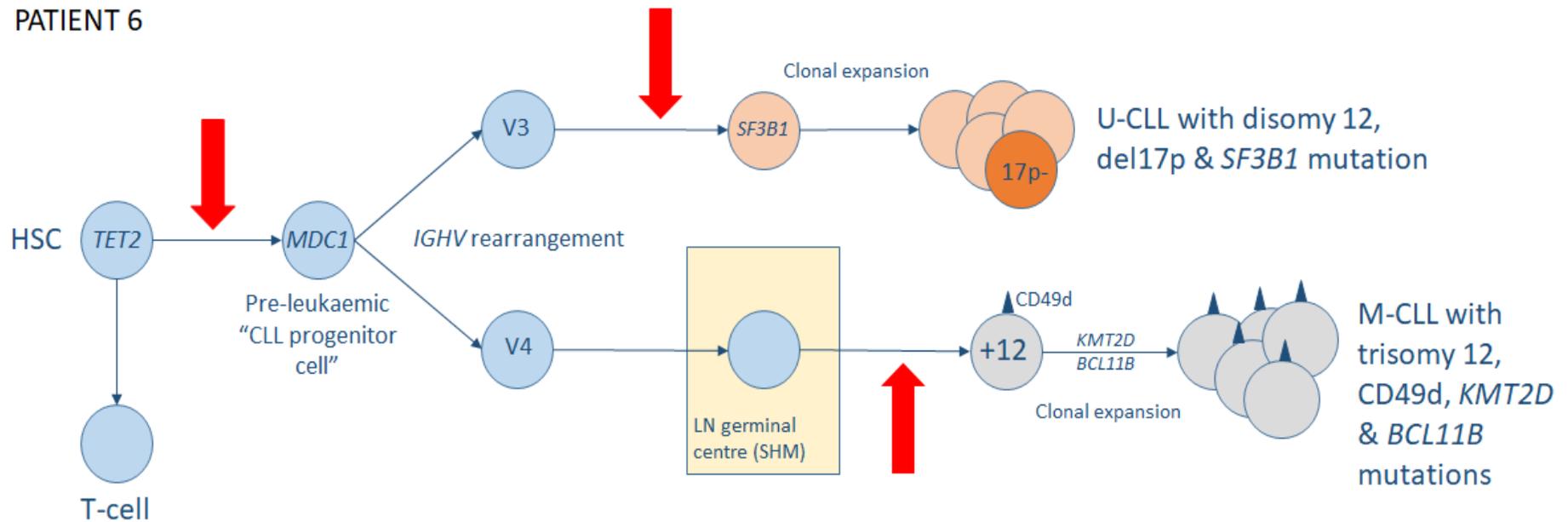
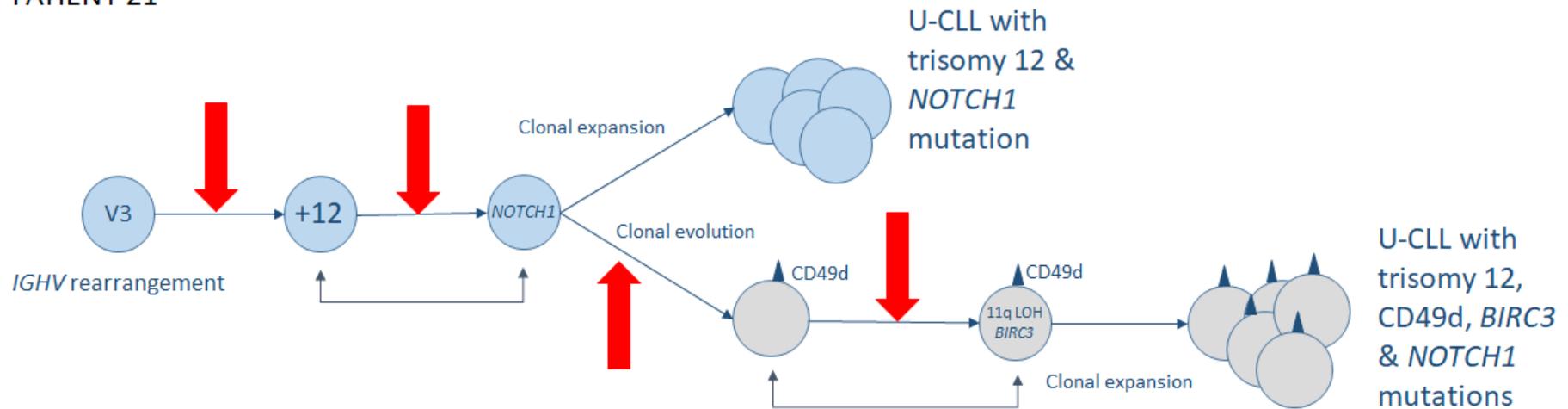


Figure 4-12. Proposed evolution of CLL in patient 21 (see text). Genetic lesions (red arrows) arise following IGHV rearrangement. Trisomy 12 and/or NOTCH1 mutation arise early in the evolution of CLL in the blue parent clone. The parent clone gives rise to a daughter clone (grey) that subsequently acquires CD49d expression via unknown mechanisms and loss of heterozygosity (LOH) at 11q and a BIRC3 mutation. This clone subsequently expands to outcompete the parent clone. U-CLL refers to an unmutated IGHV.

PATIENT 21



5 ANALYSIS OF GENE EXPRESSION IN TRISOMY 12 CLL

5.1 INTRODUCTION

Trisomy 12 CLL has a unique gene expression signature compared to other subtypes of CLL with enrichment of genes involved in the PI3K-AKT-mTOR, chemokine and BCR signalling pathways [138, 150, 151]. This unique expression pattern may explain its enhanced sensitivity, at least *in vitro*, to kinase inhibitors targeting these pathways [138]. Despite this finding, the critical genes that lead to and/or propagate leukaemogenesis in trisomy 12 CLL remain unknown. Furthermore, whilst there is an observed gene-dosage effect with many overexpressed genes in trisomy 12 CLL mapping to chromosome 12, there are clearly other mechanisms at play as several overexpressed genes do not map to chromosome 12. For example, overexpression of insulin growth factor 1 receptor (*IGF1R*) on chromosome 15 is associated with trisomy 12 CLL [152, 215]. Interestingly *IGF1R* expression is associated with resistance (rather than enhanced sensitivity) to the PI3K inhibitor, idelalisib, in a murine model [215].

One particular gene of interest in trisomy 12 CLL is *ITGA4*, which is responsible for encoding the cell surface marker CD49d. CD49d is expressed highly on normal B cells, arises early in B cell maturation on common lymphoid progenitors [5, 154] and plays a role in leukocyte adhesion and migration. Despite its high expression in normal B cells, CD49d has been found to be an independent poor prognostic marker in CLL in numerous series [162, 216-219]. The reason why it confers high-risk disease is not clear. In addition to this, overexpression of CD49d is associated with trisomy 12 CLL which does not have a uniformly poor prognosis.

Despite its well documented differential expression in CLL, the regulation of *ITGA4* and CD49d has yet to be elucidated in the published literature. One publication [155] has addressed regulation of *ITGA4* and proposed a methylation-dependent mechanism whereby CD49d-negative cases have higher levels of methylation in two CpG islands of the *ITGA4* gene - one within 1kb upstream of the translation start site and one within the first exon and intron of *ITGA4*. There are several limitations to this publication: methylation levels of these islands were highly variable, correlation between methylation percentage and CD49d expression (as measured by percentage positive cells) was weak with a r^2 value of 0.6, and the highest level of methylation observed in the CD49d- cases was approximately 20%. More recently, methylation of only three key CpG loci in the *ITGA4* gene has been shown to associate with CLL compared to healthy controls [220], however, this is of little clinical relevance and no robust connection between methylation levels of *ITGA4* and

CD49d protein expression was made in either group. Other mechanisms of CD49d regulation have not been explored in CLL despite the prognostic relevance of the marker and only a weak relationship between *ITGA4* mRNA levels and CD49d expression has been demonstrated [155]. Finally, it is unclear whether trisomy 12 CLL actively upregulates CD49d or if the converse is true and disomy CLL actively downregulates its expression. There has been no systematic comparison of CD49d expression in CLL cells to normal B cells. Chapter 3 presented a comparison of trisomy 12 and disomy 12 CLL CD49d expression to only 3 healthy controls (see section 3.3.2.3) and showed downregulation of CD49d on all CLL cells, albeit to a lesser extent on trisomy 12 CLL.

To explore gene expression changes in trisomy 12 and CD49d+ CLL, a direct comparison of the transcriptome between the purified and separated trisomy 12, CD49d+ and disomy 12, CD49- CLL clones in patient 6 was performed. As both these clones arose in the same individual, it was thought that the confounding effects encountered when comparing large groups of patients with either disomy or trisomy 12 CLL would be minimised. The two clones were subjected to whole transcriptome analysis via RNAseq with the following aims:

1. To determine the expression of the *MDC1*, *KMTD2* and *BCL11B* variants detected in the exome sequencing of the trisomy 12 clone (see Chapter 4);
2. To identify transcription factors that may be involved in *ITGA4* gene regulation; and
3. To identify differentially expressed pathways in trisomy 12 CLL and to confirm findings in a wider trisomy 12 CLL cohort (see also Chapter 6).

5.2 METHODS

5.2.1 RNA extraction

Following flow cytometry experiments on thawed cryopreserved PBMCs, 1mL of TRI Reagent® was added to cell preparations, mixed, and incubated at room temperature for 5 minutes before storing at -80°C until required. RNA was extracted using the method described in Chapter 2 for the patient 6 RNAseq sample. RNA extractions for all other patient samples included the addition of 20µg of RNA-grade glycogen (ThermoFisher) to the aqueous phase following the chloroform phase separation and heating at 55°C for 5 minutes during the final resuspension. RNA quality was assessed using the Agilent 2100 Bioanalyzer instrument according to the manufacturer's protocol and only RNA with an RNA Integrity Number (RIN) of ≥8.0 was used in downstream experiments.

5.2.2 RNAseq

Total RNA was sent to the ACRF Cancer Genomics Facility for RNAseq with polyA selection (to enrich for mRNA). Sequencing was performed using an Illumina™ NextSeq Mid Output kit with 2 x 150bp

read lengths. The initial bioinformatics was performed at this facility, with reads mapped to the human genome (hg19) after removing adapter sequences using Spliced Transcripts Alignment to a Reference (STAR). The genomic facility provided a list of genes with normalised read counts and estimated log₂fold changes to the author. Aligned reads were interrogated by importing bam files into Integrative Genomics Viewer (IGV) [221].

5.2.3 Bioinformatics

To identify potential targets, genes were considered if they were either up or downregulated in the CD49d+ clone by 10 times or greater (log₂fold >3.32 or <-3.32), and if either clone had greater than 100 reads. This curated list is presented in Table 9-12 and Table 9-13 of the Appendix. Gene-set enrichment analysis was then performed using Enrichr (<http://amp.pharm.mssm.edu/Enrichr>) and the Gene Ontology Resource (<http://geneontology.org/>). The Bonferroni correction was used with the latter analysis. Both analyses utilised Protein Analysis Through Evolutionary Relationships (PANTHER) software.

5.2.4 Quantitative RT-PCR confirmations

Quantitative real-time PCR (qRT-PCR) was performed on a cohort of trisomy 12 and disomy 12 CLL, and on sorted samples from patients 6 and 21 (CD4d+ and CD49d- CLL clones) for the following genes: *ITGA4*, *KMT2D*, tumour necrosis factor alpha induced protein 3 (*TNFAIP3*), *IRF4* and enhancer of zeste homolog 2 (*EZH2*). Primers are listed in Table 9-3 and were designed to span exons. Details of qRT-PCR methods are outlined in Chapter 2.

5.2.5 Bisulfite treatment and sequencing of *ITGA4*

DNA samples from unsorted, CD49d+ and CD49d- CLL cell fractions (patients 6 and 21) were bisulfite treated with Qiagen™ Epiect Bisulfite kit according to the manufacturer's instructions and sequenced with previously published amplicons [155]. The PCR conditions were as follows: 94°C for 4 minutes, 45 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 90 seconds, followed by a final 7 minutes at 72°C. Sanger sequencing of the amplicons was performed using both forward and reverse primers for each product. An additional primer (*ITGA4* primer 5; see Table 9-4) was used for sequencing purposes only (to bridge a homopolymer region) and was nested within the first PCR product. Any cytosine that failed to be converted to thymidine (from uracil) on the sequencing chromatogram (forward read) was considered to be methylated.

5.2.6 Statistics

The software GraphPad Prism 9 was used to perform statistical analysis. Relative mRNA expression of targets was normalised to the mean of the trisomy 12 group (or CD49d+ group, depending on the

comparison) such that the mean of that group was equal to 1.0. Expression levels were graphed on individual scatter plots; means and standard deviations were also plotted. There were at least three different samples ($n \geq 3$) per group. Unpaired t-tests were used to compare groups; equal standard deviations were not assumed (Welch's correction) and a p-value of <0.05 was assumed to be statistically significant. Correlation between variables was investigated using linear regressions and assigned an r^2 value.

5.3 RESULTS

RNA was extracted from the CD49d+ and CD49d- sorted fractions of patient 6 on two separate occasions (RNA from sorted fractions on the 29/06/2018 was used for RNAseq and RNA from sorted fractions on the 10/09/2020 was used for qRT-PCR confirmations). This was necessary due to low yield in the initial sort; all the RNA was utilised in the RNAseq experiment. RNA was also extracted from thawed cryopreserved PBMCs in a cohort of trisomy 12 and disomy 12 CLL patients. Due to variable RNA quality following thawing of samples, RNA was also extracted from PBMCs that had been directly stored in TRI Reagent® or TRIzol™ after blood collection (if available). An arbitrary decision was made to use samples with RNA integrity numbers of ≥ 8.0 after initial qRT-PCR experiments showed variable amplification of targets (including failure to amplify the housekeeping gene *GUSB*) and/or low expression of *GUSB* (data not shown).

5.3.1 Gene expression changes in trisomy 12 CLL

5.3.1.1 RNAseq of CD49d+ and CD49d- CLL clones of patient 6

There were 172 million read pairs in total for the RNAseq on the two CLL clones isolated from patient 6: 83 and 89 million read pairs for the CD49d+ and CD49d- clones respectively. 78% of genes had similar expression between the two samples (\log_2 fold change between -1.0 and 1.0; that is, up to two-fold change higher or lower between clones). This was expected given they both are CLL samples from the same individual (see Figure 5-1). Changes in expression levels between the two clones followed an approximate normal distribution (see Figure 5-2). The number of genes up- or down-regulated by at least 2-fold in the CD49d+ (trisomy 12) clone with respect to the CD49d- clone is presented Table 5-1. As expected, expression of *ITGA4* was higher in the CD49d+ clone with a \log_2 fold change of 6.57 (absolute fold change of 95x) with respect to the CD49d- clone and read counts of 6691 and 70 in the CD49d+ and CD49d- clones respectively. The complete a list of genes with normalised read counts and estimated \log_2 fold changes is provided in in the Supplementary Appendix deposited in the Flinders University library along with this thesis.

Figure 5-1. Normalised read count plot of RNAseq of two CLL clones from patient 6. Each blue dot represents a single gene transcript. The number of reads of each transcript for the CD49d+ clone (y axis) is plotted against the number of reads of the same transcript for the CD49d- clone (x axis).

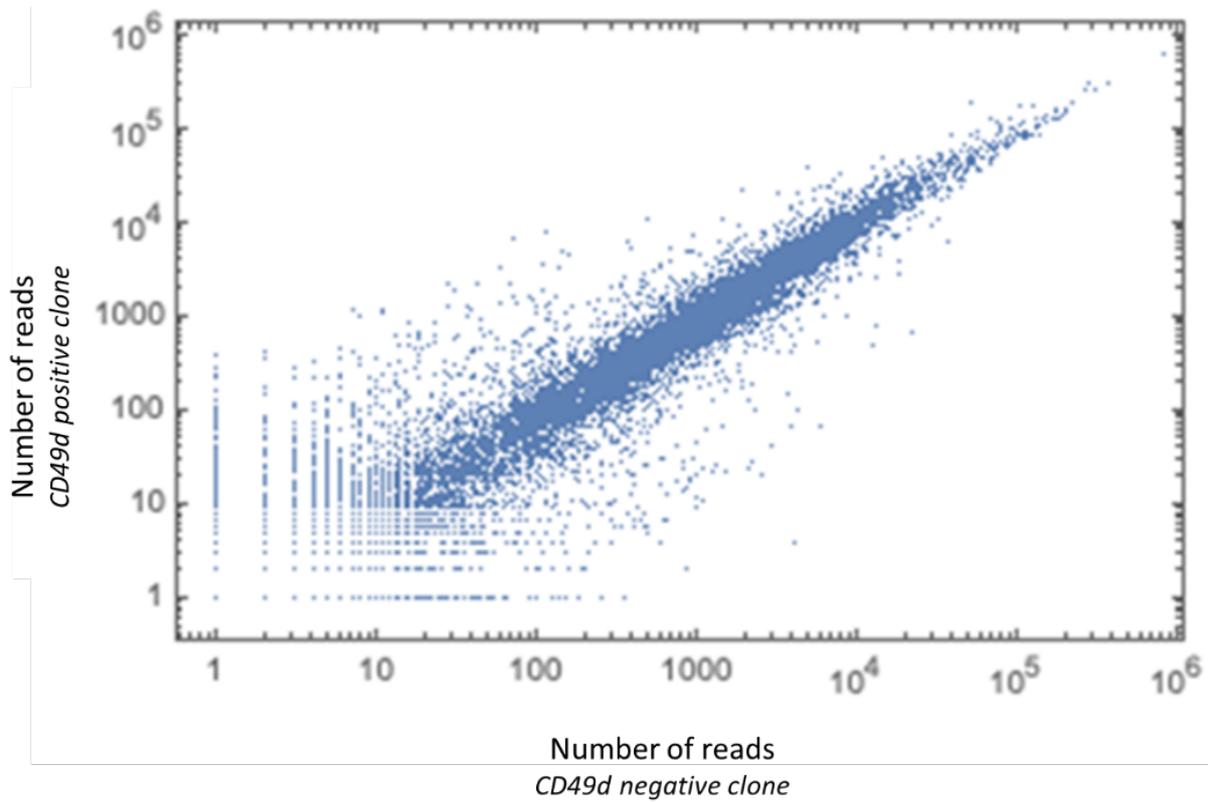


Figure 5-2. Histogram of changes in gene expression levels between the two CLL clones from patient 6.

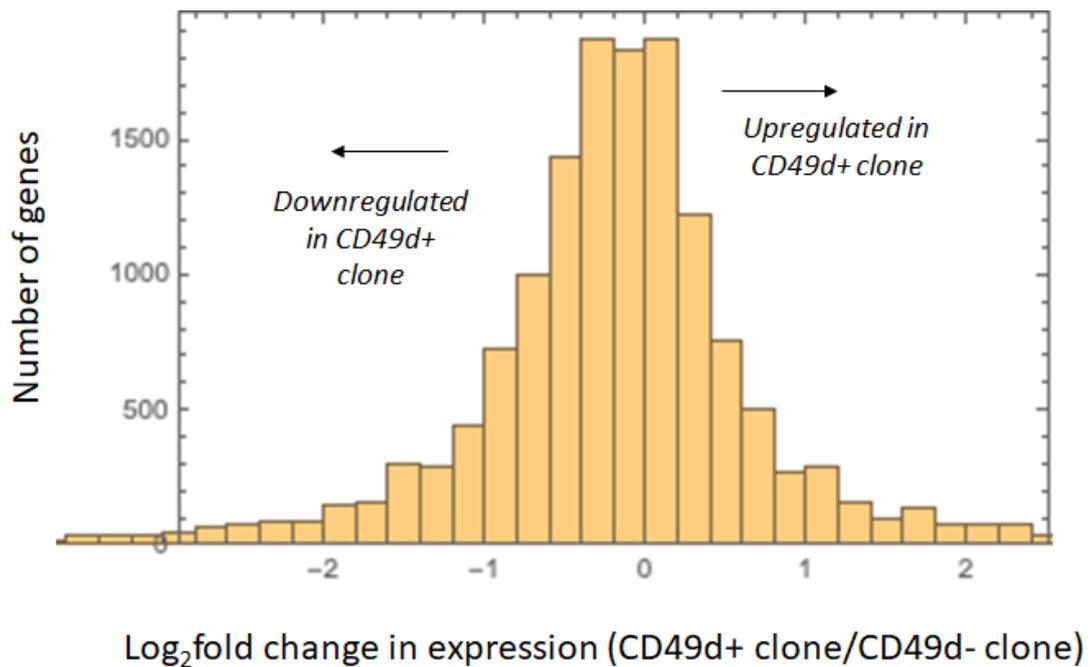


Table 5-1. Number of genes up- or down-regulated by at least 2-fold in the CD49d+ clone with respect to the CD49d- clone. * these genes were used for gene-set enrichment analysis. Negative numbers refer to downregulation (i.e. -2 means there is downregulation by 2x in the gene product in the CD49d+ clone).

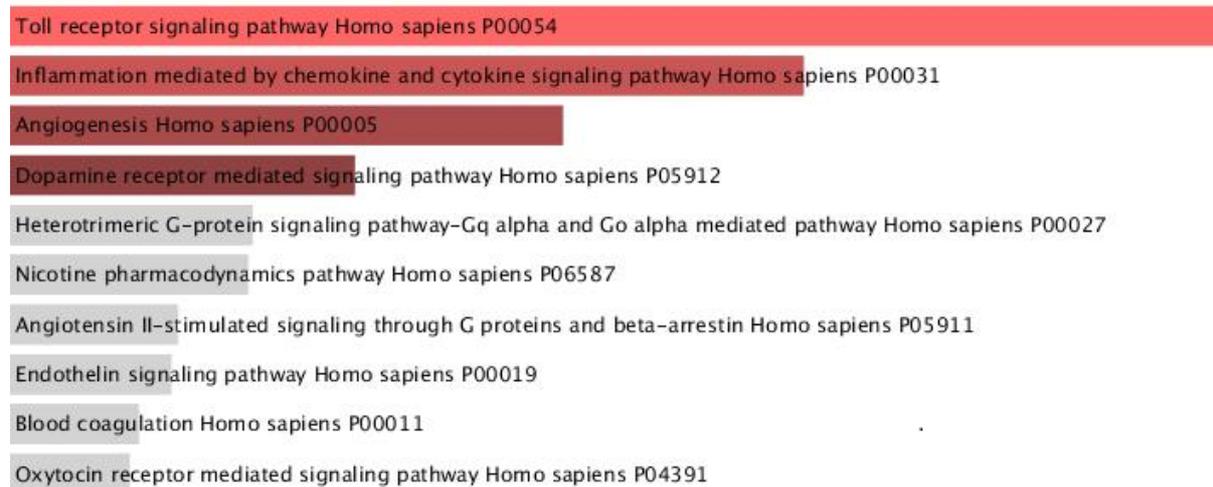
Log ₂ fold change	Absolute fold change (with respect to the CD49d+ clone)	Number of genes (% of total)	Number of genes with read count >10	Number of genes with read count >100
Upregulated in the CD49d+ (trisomy 12) clone				
1	2	2352 (10.1)	1135	567
2	4	1108 (4.7)	627	251
3.32	10	469 (2.0)	369	138*
6.64	100	31 (0.1)	30	13
Downregulated in the CD49d+ (trisomy 12) clone				
-1	-2	2739 (11.7)	1671	659
-2	-4	1029 (4.4)	566	271
-3.32	-10	333 (1.4)	241	123*
-6.64	-100	12 (0.05)	12	10

5.3.1.2 Gene-set enrichment analysis of RNAseq

Genes that had a minimum of 100 absolute reads and that were up or down-regulated at least 10-fold (138 and 123 genes respectively) were subjected to gene set enrichment analysis using two different platforms, the Enrichr webserver and the Gene Ontology (GO) Resource. The most enriched pathway was Toll receptor signalling in both analyses (p=0.000435 and p=0.00927 for the Enrichr and GO analyses respectively), with the following genes represented in the pathway: toll-like receptor 8 (*TLR8*), interleukin 1 receptor associated kinase 3 (*IRAK3*), *CD14*, prostaglandin-endoperoxide synthase 2 (*PTGS2*), toll-like receptor 4 (*TLR4*) (Enrichr analysis), and *CD14*, disable homolog 2-interacting protein (*DAB2IP*), *IRAK3*, *LILRA2*, leukocyte immunoglobulin like receptor A4 (*LILRA4*), *TLR4* and protein tyrosine phosphatase receptor type S (*PTPRS*) (GO analysis). The Enrichr analysis is presented in Figure 5-3.

Investigation of the toll-like receptor pathway is the main focus of Chapter 6 where validation of the above RNAseq findings and toll-stimulating assays are pursued to address the potential role of toll-signalling in trisomy 12 CLL. As such, it is not further discussed in this Chapter.

Figure 5-3. Gene-set enrichment analysis of RNAseq data (patient 6) using PANTHER via the Enrichr platform. Toll receptor signalling is the most enriched pathway for genes differentially expressed between the trisomy and disomy 12 CLL clones.



5.3.1.3 Expression of *MDC1*, *KMT2D*, *BCL11B* and selected targets (RNAseq)

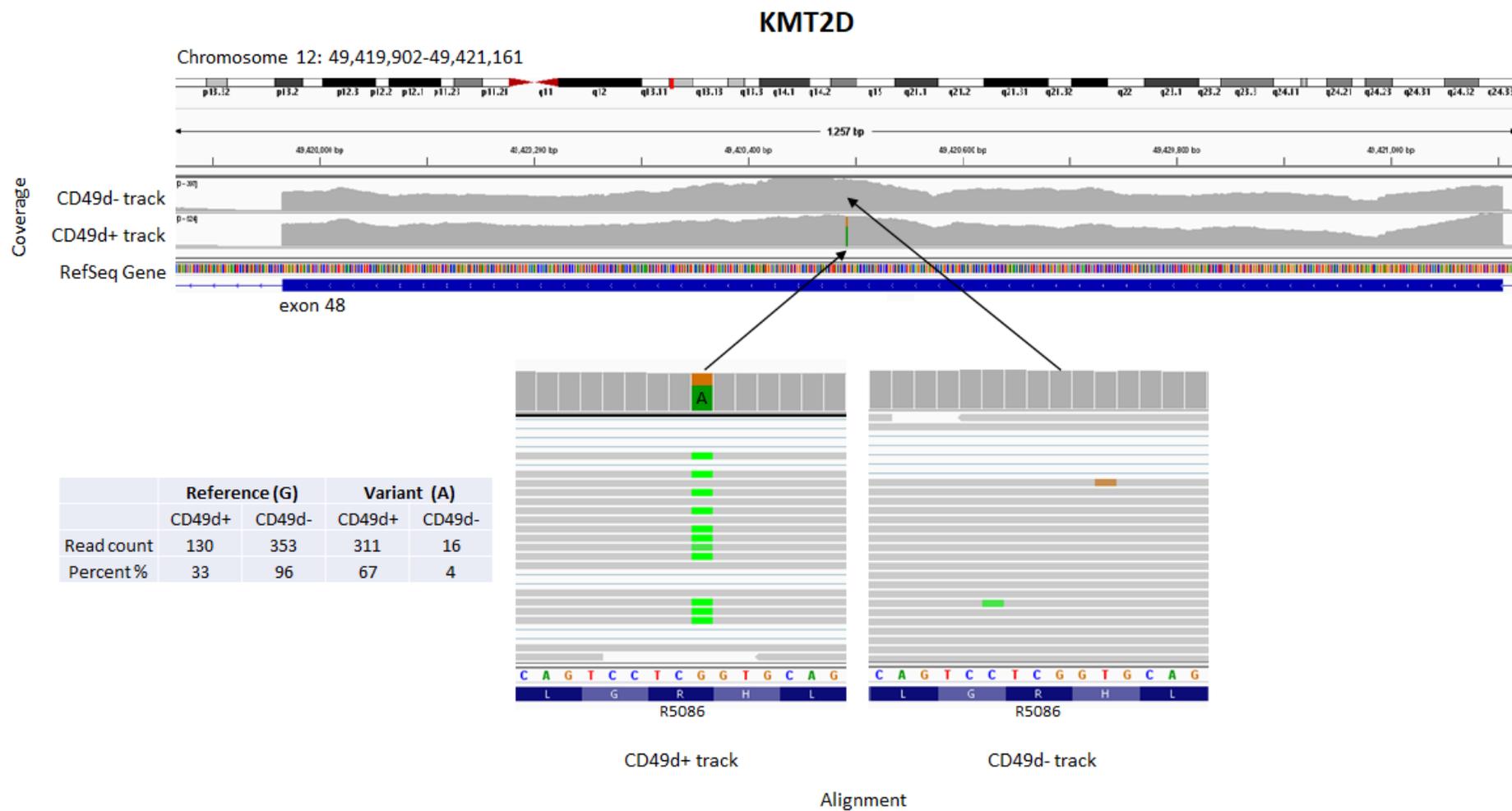
The RNAseq data from both clones was able to be interrogated for expression of the potentially pathogenic mutations of *MDC1*, *KMT2D* and *BCL11B* identified from the whole exome sequencing (described in Chapter 4). Expression of these targets in the CD49d+ and CD49d- clones of patient 6 is presented in Table 5-2. *MDC1* c.4824_4946del was identified as a heterozygous in-frame deletion in both the CD49d+ and CD49d- clones in the WES. Interestingly, the RNAseq analysis demonstrated similar levels of expression of *MDC1* in both clones (with read counts of 3386 and 2766 in the CD49d+ and CD49d- clones respectively), but the mutated allele was not expressed. *BCL11B* was not expressed with read counts of 31 and 1 in the CD49d+ and CD49d- clones respectively. The *BCL11B* missense variant c.1387G>A observed in the exome sequencing in the CD49d+ clone was not detected in the RNAseq.

KMT2D expression was 1.8x greater in the CD49d+ trisomy 12 clone compared to the CD49d-disomy 12 clone in line with expectation (it lies on chromosome 12). It was expressed at high levels in the CD49d+ clone (20354 reads). Furthermore, the *KMT2D* nonsense variant detected in the CD49d+ clone in the WES (*KMT2D* c.15256C>T; p.R5086*) is expressed (see Figure 5-4) and is present in 67% of reads of the CD49d+ trisomy 12 clone, implying that one copy of chromosome 12 with the variant was duplicated. Targets of *KMT2D* were variably expressed and are not uniformly up- or down-regulated in the CD49d+ clone (see Table 5-2).

Table 5-2. Expression levels and read counts of MDC1, KMT2D, BCL11B and selected KMT2D targets in the RNAseq of CD49d+ and CD49d- clones of patient 6. SOCS3 = suppressor of cytokine signaling 3; SGK1 = serum/glucocorticoid regulated kinase 1; MAP3K8 = mitogen-activated protein kinase kinase kinase 8; TNFRSF14 = tumour necrosis factor receptor superfamily member 14; TRAF3 = tumour necrosis factor receptor associated factor 3; IKBKB = inhibitor of nuclear factor kappa B kinase subunit beta. See text for details on other genes.

Gene	Normalised read count (CD49d+ clone)	Normalised read count (CD49d- clone)	Relative expression in CD49d+ clone (x)	Chromosome
MDC1	3386	3766	0.9	6p
BCL11B	31.4	1.0	31.4	14q
KMT2D & targets				
KMT2D	20354	11297	1.8	12q
SOCS3	457	1108	0.4	17q
SGK1	1445	238	6.1	6q
MAP3K8	4977	876	5.7	10p
TNFAIP3	4109	4261	1.0	6q
TNFRSF14	4496	5491	0.8	1p
TRAF3	5562	4170	1.4	14q
IKBKB	3861	3957	1.0	8p

Figure 5-4. RNAseq analysis of KMT2D gene transcript in patient 6. The top panel shows the chromosomal location, RefSeq Gene map and coverage (read counts) of KMT2D at exon 48 in both the CD49d- and CD49+ tracks. The CD49d+ and CD49d- coverage tracks have been zoomed in at the location of the variant c.15256C>T; p.R5086* (note that the reverse complement is shown in the RNAseq data). The CD49d+ track shows 311 (67%) of reads at this location are the variant A (highlighted in green) rather than the reference A.



5.3.1.4 Expression of the transcription factors EZH2 and IRF4

Given one of the aims of this work is to understand the regulation of *ITGA4* expression in the two clones, a manual search of transcription factors that potentially regulate *ITGA4* was performed. Seventy transcription factors are predicted to bind to the CpG islands of the *ITGA4* gene on the University of California Santa Cruz (UCSC) genome browser (transcription factor ChIP-seq clusters from the ENCODE 3 track at genome.ucsc.edu; data not shown). Only two transcription factors, *IRF4* and *EZH2*, were differentially expressed in the CD49d+ CLL clone of patient 6 (see Table 5-3). *IRF4* was 3.9x fold higher in the CD49d+ clone, and *EZH2* was 10x fold higher in the CD49d- clone.

Table 5-3. Expression of EZH2 and IRF4 in RNAseq of CD49d+ and CD49d- clone of patient 6.

Gene	Normalised read count (CD49d+ clone)	Normalised read count (CD49d- clone)	Absolute fold change in CD49d+ clone (x)	Chromosome
<i>IRF4</i>	6870	1822	3.9	6p
<i>EZH2</i>	263	1823	0.1	7q

5.3.1.5 Quantitative RT-PCR confirmation of differentially regulated genes

Confirmation of the RNAseq findings by qRT-PCR was not possible in the exact same RNA samples from patient 6 due to low yield following the sort, resulting in all the RNA being used in the RNAseq experiment. As such, a second sort on another sample from patient 6 was performed on 10/09/2020 to collect RNA for the confirmation qRT-PCRs (data not shown; purity of fractions was >98% as per the initial sort). The quality of the RNA, however, was not the same as the initial sort and RNA from the second sort did not have RIN > 8.0 (RIN of 6.4 in CD49d+ clone and an unrecordable RIN in the CD49d- clone due to degraded RNA). Nonetheless, this RNA was used in the confirmation qRT-PCRs due to the precious nature of the patient sample and the absence of an excess of stored material. Confirmation qRT-PCRs were performed in the sorted samples of patient 6 and 21 for the following genes: *ITGA4*, *KMT2D* and its target *TNFAIP3*, and the transcription factors *IRF4* and *EZH2*. qRT-PCRs were then performed on a wider cohort of trisomy 12 and disomy 12 CLL to determine if any of the changes observed in patient 6 were applicable to a larger cohort of CLL patients. The results are presented below.

5.3.1.5.1 Expression analysis of sorted samples from patient 6 and 21

RNA from the second sorted sample of patient 6 was used to confirm the RNAseq findings initially. *ITGA4* and *IRF4* amplified in the CD49d+ clone but failed to amplify in the CD49d- clone (which was of lesser RNA quality) and comparisons could not be made. *KMT2D*, *TNFAIP3* and *EZH2* all amplified in both clones, and *KMT2D* showed a similar fold change in the CD49d+ clone in both the RNAseq and qRT-PCRs (1.80x and 1.74x upregulated respectively). *TNFAIP3* and *EZH2* expression were confirmed to be higher in the CD49d- clone by qRT-PCR concordant with the RNAseq findings (see Table 5-4).

Table 5-4. Comparison of RNAseq and qRT-PCR data for patient 6.

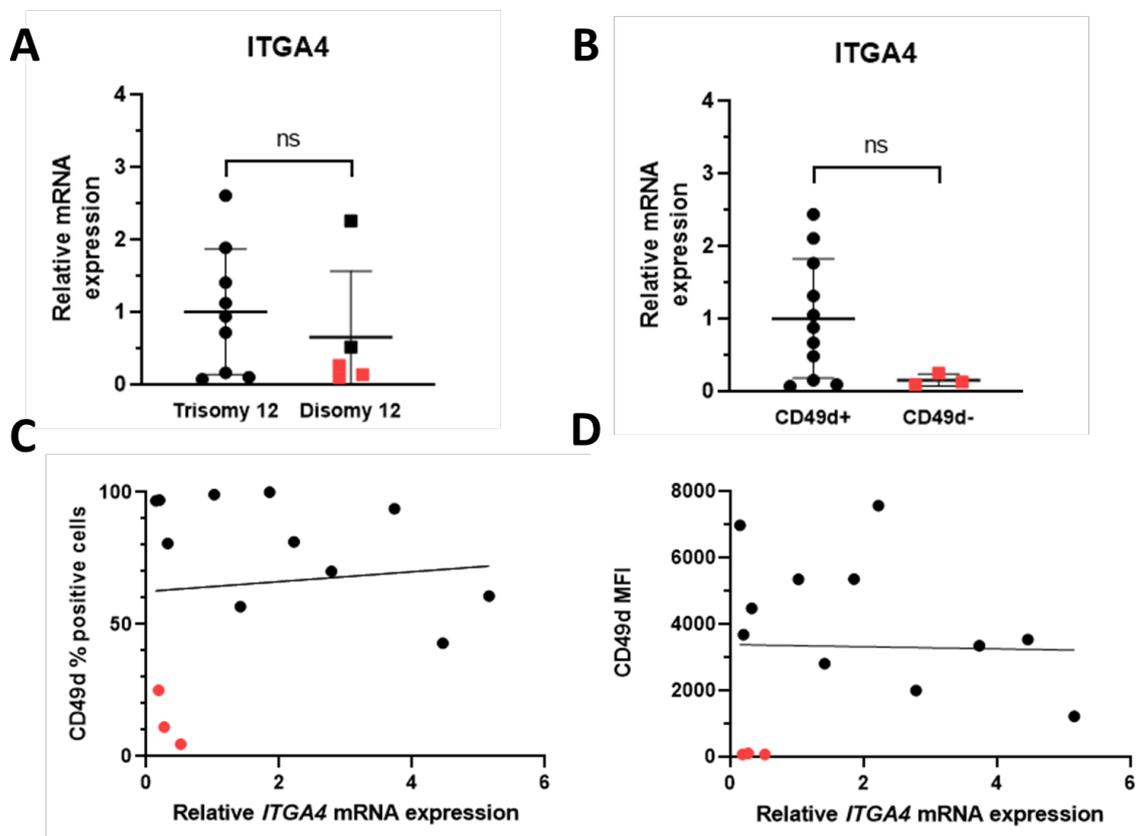
Gene	Expression (CD49d+ clone)	Expression (CD49d- clone)	Relative expression in CD49d+ clone
<i>KMT2D</i>			
RNAseq read count	20354	11297	1.80
qRT-PCR expression (relative to <i>GUSB</i>)	7.57	4.35	1.74
<i>TNFAIP3</i>			
RNAseq read count	4109	4261	0.96
qRT-PCR expression (relative to <i>GUSB</i>)	5.37	12.37	0.43
<i>EZH2</i>			
RNAseq read count	263	1823	0.14
qRT-PCR expression (relative to <i>GUSB</i>)	0.36	72.35	0.005

Whilst *ITGA4* failed to amplify in the CD49d- clone of patient 6, it amplified in both CD49d+ and CD49d- clones in a sorted sample from patient 21. *ITGA4* mRNA expression was 54-fold higher in the CD49d+ clone in line with expectation.

5.3.1.5.2 Expression of *ITGA4* in the wider cohort

CD49d expression at the protein level (as measured by flow cytometry) was shown to be increased by both MFI and percentage positive cells in the trisomy 12 compared in the disomy 12 cohort in Chapter 3 (see Figure 3-6). To determine if this was also the case at the mRNA level, expression analysis of *ITGA4* by qRT-PCR was performed. Unexpectedly, there was no statistically significant difference between *ITGA4* mRNA expression between trisomy 12 and disomy 12 CLL, nor between CD49d+ and CD49d- CLL (see Figure 5-5A,B). However, there is a clear trend to lower *ITGA4* expression in the three CD49d- samples (DIS1, DIS11, DIS13) that are highlighted in red in Figure 5-5. A robust correlation between *ITGA4* mRNA expression and CD49d protein expression (by MFI or percentage positive cells) could not be found using linear regression (data shown below) or log-linear scales as previously published (data not shown) [155] (see Figure 5-5C,D).

Figure 5-5. *ITGA4* mRNA expression (relative to the housekeeping gene *GUSB*) in trisomy 12 and disomy 12 CLL (A) and in CD49d+ and CD49d- CLL (B). Expression levels have been normalised to the mean of the trisomy 12 and CD49d+ cases respectively. Mean +/- standard deviation is plotted. Correlation between percentage positive CD49d cells or CD49d MFI and relative *ITGA4* mRNA expression is shown in (C) and (D) respectively. The line of best fit (linear regression) is plotted. The three CD49d- samples are shown as red dots (DIS1, DIS11, DIS13).



5.3.1.5.3 Expression of *KMT2D* and *TNFAIP3* in the wider cohort

KMT2D was upregulated in the CD49d+ trisomy 12 clone (by 1.7x) in patient 6 (see Table 5-4). One of its targets *TNFAIP3* (encoding an inhibitor of an intermediary in TLR4 and NF- κ B signalling) was downregulated in the CD49d+ clone in patient 6 (by 2.3-fold; see Table 5-4). To determine if *KMT2D* and *TNFAIP3* are relevant to trisomy 12 CLL, their expression was measured in a larger cohort of trisomy 12 and disomy 12 CLL (with at least 4 samples in each group). No statistically significant difference in expression of either *KMT2D* or *TNFAIP3* was observed between the groups (Figure 5-6A,B). *KMT2D* and *TNFAIP3* expression did, however, have a weak correlation with a r^2 value of 0.45 (Figure 5-6C) in keeping with *KMT2D*'s proposed function of upregulation of *TNFAIP3* expression [222]. Furthermore, there was a positive correlation in *KMT2D* and *TNFAIP3* expression for all disomy 12 samples (n=3) and for the majority (6/8) of trisomy 12 samples (see Figure 5-6D). Two of 8 trisomy 12 samples had a negative relationship between *KMT2D* and *TNFAIP3* expression (patients 15 and 70). The mutational status of *KMT2D* in these patients is not known. *KMT2D* expression did not correlate with *ITGA4* expression (data not shown, $r^2=0.06$).

5.3.1.5.4 Expression of *IRF4* and *EZH2* in the wider cohort

Since both *IRF4* and *EZH2* were differentially regulated in the CD49d+ clone in patient 6 and both are predicted to bind to the *ITGA4* promoter, it was postulated that they may have a regulatory role in CD49d expression. However, neither *IRF4* nor *EZH2* were differentially expressed in a cohort of CD49d+ CLL samples compared to CD49d- samples (see Figure 5-7A,B), and neither *IRF4* nor *EZH2* expression correlated with *ITGA4* expression at the mRNA level (see Figure 5-7C,D). There was no correlation between *IRF4* expression and CD49d MFI (see Figure 5-7E), however, there was a weak negative correlation between *EZH2* expression and CD49d MFI (excluding one outlier, patient 5; $r^2=0.5$, $p=0.03$) (see Figure 5-7F).

Figure 5-6. Expression of *KMT2D* and *TNFAIP3* (relative to the housekeeping gene *GUSB*) in a cohort of trisomy 12 and disomy 12 CLL (A and B respectively). Expression levels have been normalised to the mean of the trisomy 12 cases. Mean +/- standard deviation is plotted. Correlation between *KMT2D* and *TNFAIP3* expression is shown in (C). The line of best fit is plotted in the solid line with an r^2 value of 0.4; the line of identity (representing a perfect correlation) is the dotted line. The three orange dots are disomy 12 samples; the remainder are trisomy 12. (D) shows the paired expression of *KMT2D* and *TNFAIP3* in individual samples joined by solid lines.

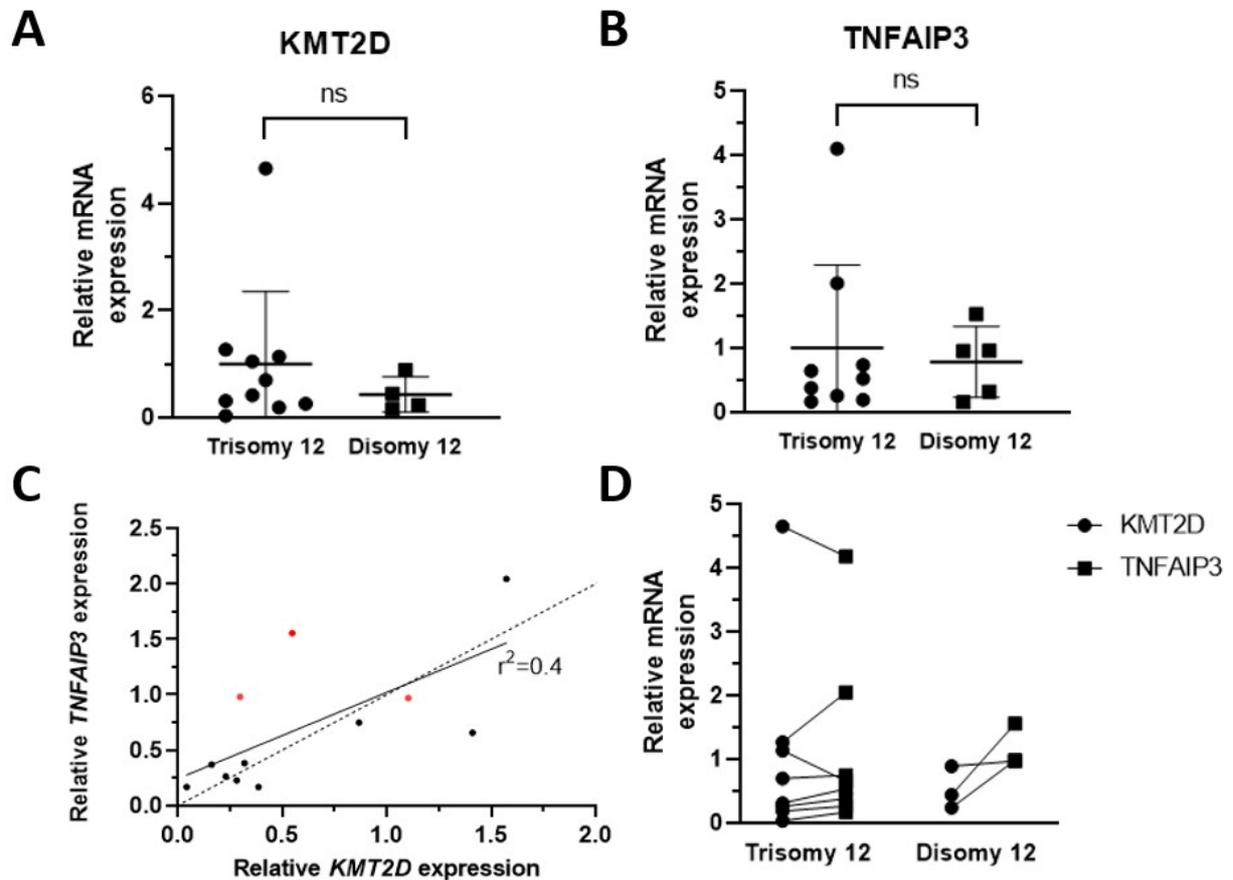
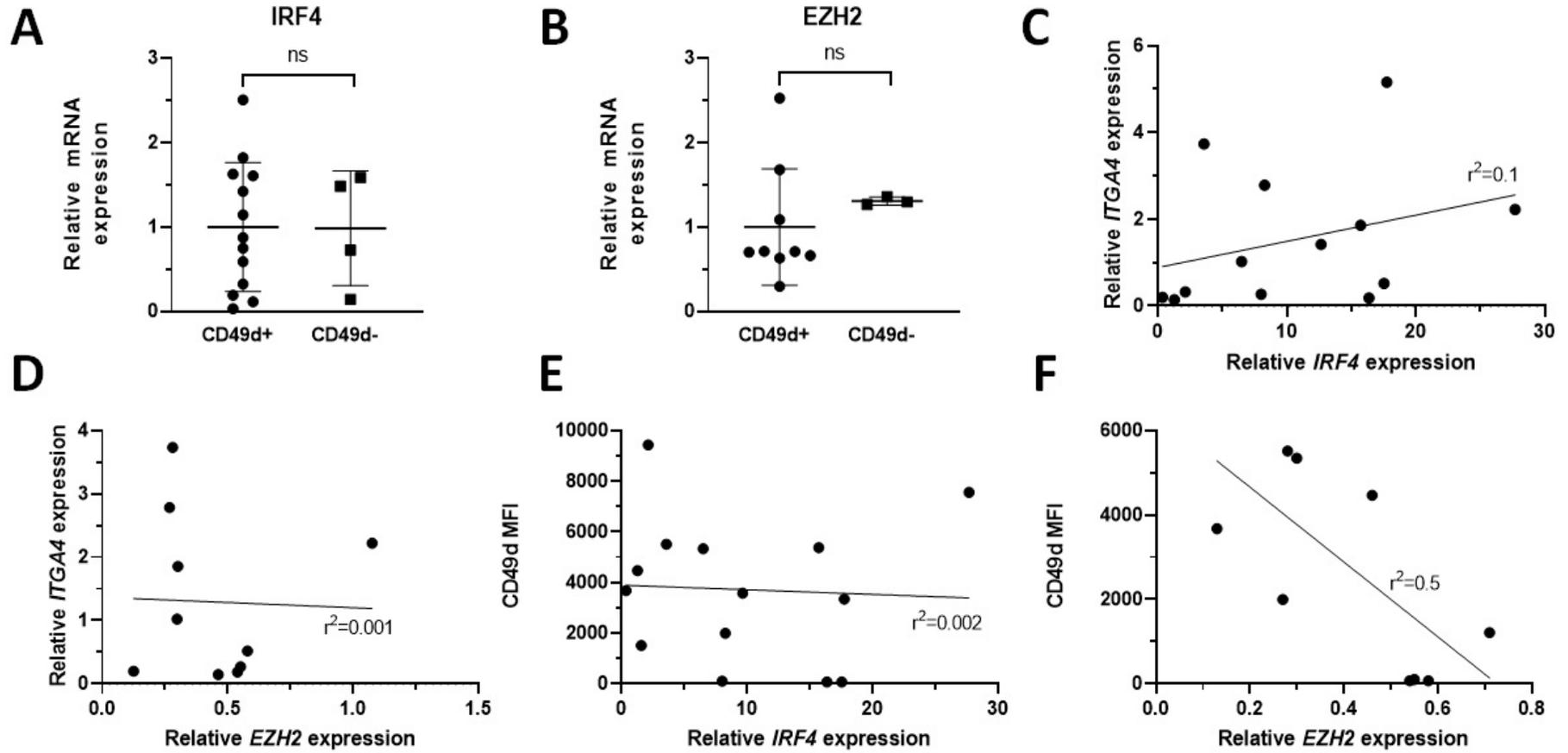


Figure 5-7. Expression of IRF4 and EZH2 (relative to the housekeeping gene GUSB) in a cohort of CD49d+ and CD49d- CLL (A and B respectively). Expression levels have been normalised to the mean of the CD49d+ cases. Mean +/- standard deviation is plotted. Correlation between IRF4 or EZH2 and ITGA4 mRNA or CD49d MFI are shown in (C), (D), (E) and (F). The line of best fit is plotted.



5.3.1.5.5 RNA quality

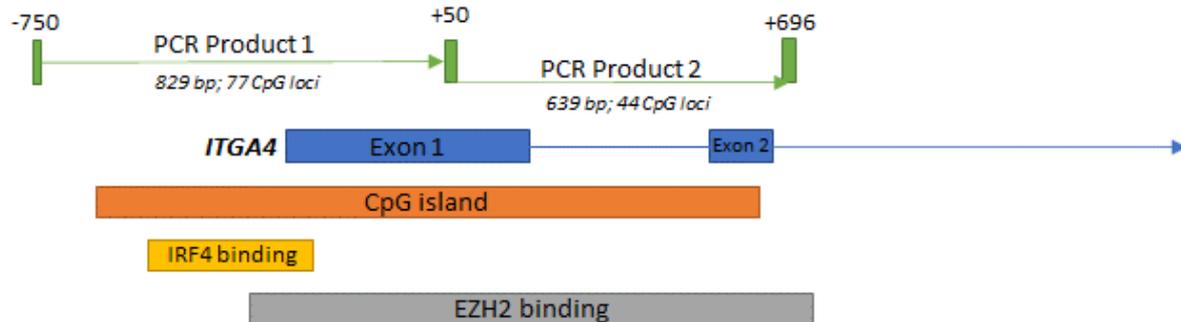
Not all samples yielded RNA of sufficient quality to perform all the qRT-PCRs. This was further investigated given the difficulty in confirming the RNAseq data in a wider cohort. Technical reasons for this discrepancy were first investigated before concluding that the index case findings could not be extrapolated to a larger cohort. There were 30 samples available for RNA extraction: 20 had only cryopreserved vials of PMBCs available and 10 had PMBCs which were stored directly in TRIzol™ and snap-frozen at -80°C . RNA integrity as measured by the RIN on the Agilent Bioanalyzer 2100 instrument was highly variable with 3 samples being completely degraded (and assigned the lowest possible RIN of 1.0 for statistical purposes) and 20 samples passing the quality criteria for downstream experimentation with a $\text{RIN} \geq 8.0$ (see Table 5-5). Of the sorted samples for patients 6 and 21, only RNA from the CD49d+ clone in patient 6 had a measurable RIN at 6.4 and the remaining samples were degraded (see Table 5-5). Note that the RNA used for the RNAseq experiment for patient 6 was of superior quality with a $\text{RIN} > 8.0$ (and hence passing quality control prior to RNAseq).

The method of storing the samples impacted the quality of the RNA (see Figure 5-8A). RNA extracted from thawed PBMC preparations was of significantly inferior quality (as determined by the RIN) as opposed to RNA extracted from PMBCs that had been stored in TRIzol™ directly following sample collection ($p=0.01$; Figure 5-8A). The time from blood collection to processing did not affect RNA quality obtained from cryopreserved vials (see Figure 5-8B) - there was no significant correlation with an r^2 value of 0.1 ($p=0.15$). Two samples (patients 22 and 73) produced good quality RNA (RIN 8.5 and 8.1 respectively) despite processing times of 24-hours, and one sample (patient 17) had completely degraded RNA despite a processing time of 4 hours. This suggests that RNA quality is affected by the method of sample processing but not the time to process (at least within 30 hours).

Table 5-5. RNA quality. The RNA integrity number (RIN) for RNA samples is presented alongside the sample storage/processing method (THAW = RNA was extracted after thawing cryopreserved vials of PBMCs; THAW/SORT = RNA was extracted after thawed cells were flow sorted; DIRECT = RNA was extracted from PMBCs that were directly stored in TRIzol™ after sample collection). Time to processing refers to the time from blood collection to PBMC separation and storage and is recorded in the nearest hour (N/A = data not available). RIN numbers of 1.0 were assigned to degraded samples.

Patient ID	Storage/processing method	Time to process	RIN	Patient ID	Storage/processing method	Time to process	RIN
2	THAW	3	8.7	21	THAW/SORT	4	1
5	THAW	N/A	9	21	THAW/SORT	4	1
6	THAW/SORT	21	6.4	22	THAW	24	8.5
6	THAW/SORT	21	1	68	THAW	1	8.9
7	DIRECT	21	8.2	70	THAW	21	9
8	THAW	N/A	8.3	71	THAW	N/A	6.9
9	DIRECT	N/A	9	73	THAW	24	8.6
10	DIRECT	N/A	8.4	DIS1	THAW	7	8.1
11	DIRECT	N/A	8.3	DIS3	THAW	28	7
12	DIRECT	N/A	7.5	DIS4	THAW	46	1
14	DIRECT	N/A	9	DIS5	THAW	2	8
14 (repeat)	DIRECT	N/A	7	DIS7	THAW	2	6.4
15	THAW	N/A	8.3	DIS9	THAW	23	1
17	THAW	4	1	DIS10	DIRECT	20	9.5
18	THAW	4	4.9	DIS11	DIRECT	N/A	8.6
19	THAW	3	6	DIS12	DIRECT	4	9.1
20	THAW	5	6.9	DIS13	DIRECT	3	8.7
21	THAW	4	8				

Figure 5-9. Map of *ITGA4* amplicons. Two amplicons of bisulfite-treated DNA were generated (product 1 and 2) covering an *ITGA4* CpG island identified on the UCSC genome browser including part of the 5'-untranslated region and first intron. Numbers refer to genomic location. Predicted IRF4 and EZH2 binding sites (Encode data from UCSC) are included. UCSC = University of California Santa Cruz at <https://genome.ucsc.edu/>.

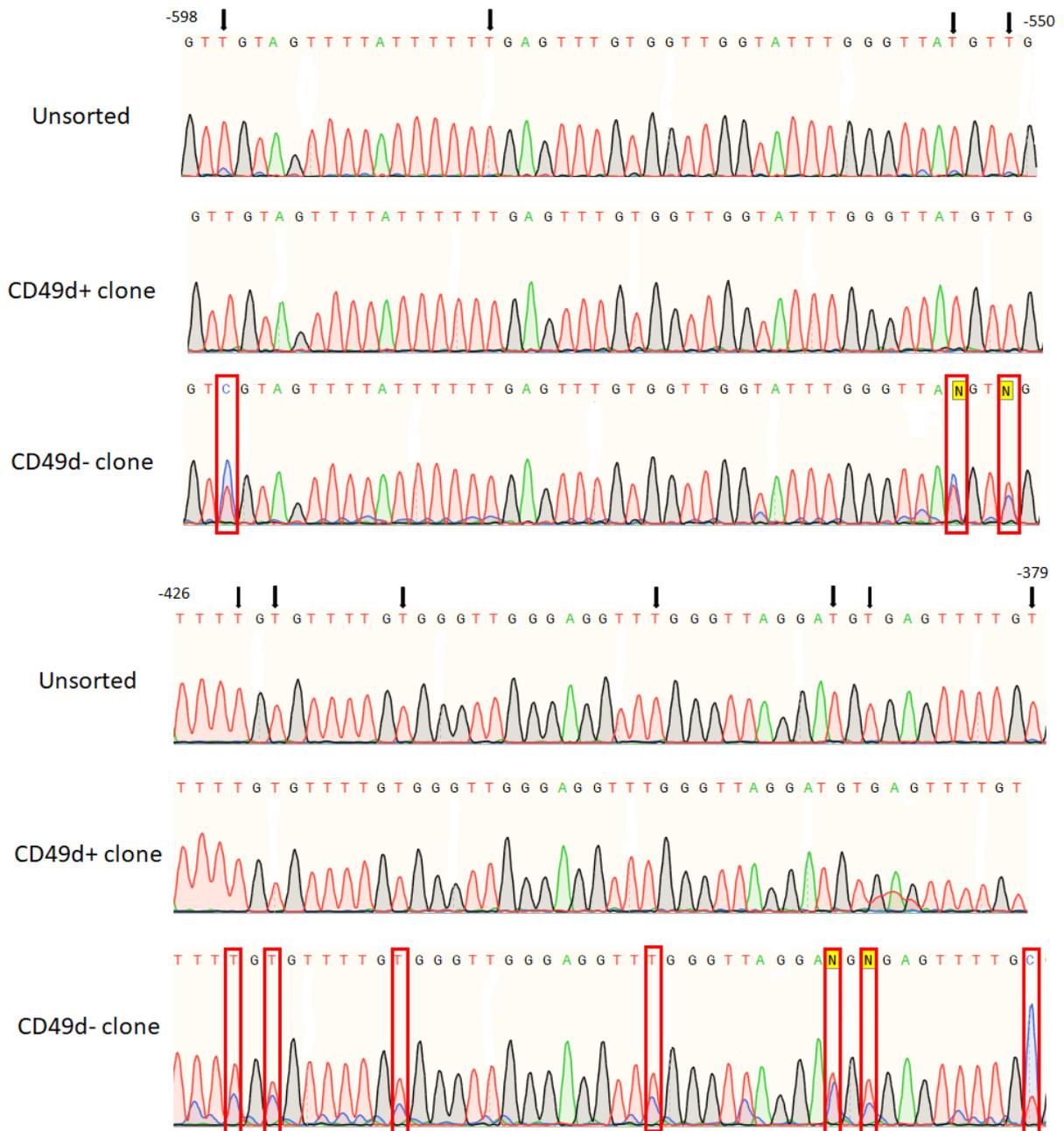


The results are presented in Table 5-6 and Figure 5-11. The regions sequenced are outlined in the gene map in Figure 5-9 and represent CpG islands in the *ITGA4* promoter, first exon and first intron. Of 121 CpG loci in the amplicons combined, between 68-99 CpG loci were successfully sequenced depending on the sample (see Table 5-6). Interestingly, the results were not concordant between the two patients. In patient 6, the CD49d- fraction demonstrated 34% methylation compared to 0% in the CD49d+ clone. That is 34% of CpG loci demonstrated two peaks in the sequencing. It is unclear where a proportion of cells within the CD49d- fraction were fully methylated at these loci, or that a larger proportion of cells within the sample showed methylation differences between the two *ITGA4* alleles. In contrast, there was no methylation observed in either the CD49d+ or CD49d- clone in patient 21, suggesting an alternate mechanism of gene regulation at least in this sample. *ITGA4* bisulfite sequencing of more patients was not undertaken.

Table 5-6. Number of methylated CpGs in *ITGA4* amplicons of sorted samples from patient 6 and 21.

Patient	Fraction	Number of CpG loci sequenced	Number of CpG methylated	% methylation
6	Unsorted	70	0	0
	CD49d+	68	0	0
	CD49d-	87	30	34
21	Unsorted	98	0	0
	CD49d+	99	0	0
	CD49d-	76	0	0

Figure 5-10. Bisulfite sequencing of ITGA4 from sorted samples (patient 6). The CD49d- leukaemic clone demonstrates methylation at multiple CpG loci. The black arrows denote CpG loci in the genomic sequence. Methylated CpGs remain as cytosines during bisulfite treatment (red boxes). Numbers are the gene co-ordinates.



5.4 DISCUSSION

This chapter investigated gene-expression changes in trisomy 12 CLL with respect to disomy 12 CLL, and potential mechanisms of *ITGA4* gene regulation. Patient 6 provided an opportunity to investigate the differing transcriptome between trisomy 12 and disomy 12 CLL within the same patient, overcoming confounding variables that are encountered when comparing large groups of patients with trisomy and disomy 12 CLL. As such, RNAseq analysis of both clones from patient 6 was performed with the aim to identify key differentially regulated pathways. A strict bioinformatic filtering strategy was employed to identify only those pathways and transcripts with large changes between the clones. The key limitation of the strategy is that the patient represents only a single case of CLL and that the findings would need to be validated in a larger cohort of patients to confirm or refute their biological and potential clinical relevance in this heterogeneous disease. Another key limitation is that the clones do not only differ in their ploidy of chromosome 12, but also in their mutational status of the IGHV and CD49d expression. Nonetheless, the case was used as a starting point in the exploration of the trisomy 12 CLL transcriptome. RNAseq of the whole cohort of trisomy and disomy 12 CLL was not performed due to budgetary constraints and the fact that this has been previously performed on even larger numbers of CLL patients. However, it should be noted that a trisomy 12 CLL subset analysis of these large RNAseq datasets was not the focus of these publications, nor did they reveal critical pathways explaining the pathogenesis of trisomy 12 CLL [223-225].

The bioinformatic filtering strategy used arbitrary cut-offs and was designed to identify the most deregulated pathways and ignore genes that had large fold changes but that may not be reproducible or biologically relevant given low absolute read counts/expression. A more permissive strategy and/or different gene-set enrichment analysis software may have identified different pathways. Nonetheless, validation in a wider cohort and functional experiments to determine the significance of identified pathways would still be required. The data was not reanalysed using a different bioinformatic strategy and the most deregulated pathway identified (toll-like signalling) will be explored in detail in the following chapter.

The RNAseq demonstrated upregulation in the gene encoding CD49d, *ITGA4*, in the CD49d+ clone as expected. This finding could not be confirmed, however, by qRT-PCR in patient 6. This is likely due to the differing qualities of RNA used in both experiments – traditionally, the same RNA would have been used in validation experiments, however, there was limited sample, especially in the CD49d- clone which represented only 18% of the total. Cells were lost during flow-sorting due to

the inherent nature of the procedure itself, but also due to strict gating to ensure purity of samples. Because of these limitations, a second sorting experiment was performed to yield the necessary amount of RNA, but this was of lower quality (with a RIN of 1.0) for unclear reasons as the protocols were identical apart from the addition of glycogen and a heating step. Further sorting experiments were not performed, again due to the limited number of cryopreserved vials for the patient in question.

RNA quality was a broader issue throughout this study. Sample numbers were limited due to several samples failing quality assessment on the Agilent 2100 Bioanalyzer instrument. A stringent RIN cut-off was applied after optimisation experiments demonstrated that some samples failed to reliably amplify the housekeeping gene *GUSB* (which has previously been validated in the author's laboratory; unpublished data) or amplified it at low levels (CT>35 cycles) with unsatisfactory variation in technical replicates (>0.5 cycles) (data not shown). The quality of the RNA samples did depend on the method of storage of the PBMCs following blood collection: a method in which they were stored immediately in TRIzol™ to inactivate RNases following separation from whole blood led to better quality RNA compared RNA extracted from PBMCs that were cryopreserved in RPMI media, fetal calf serum and DMSO. Samples that were stored in the former manner were chosen preferentially, however, only a third of samples were stored in this way. The time to processing of the blood sample was also considered in exploring RNA quality differences between samples, however, it did not appear to make a difference at least within the first 30 hours. There were no samples that were processed after 48 hours. It is also acknowledged that there may be intrinsic biological differences to the samples affecting the quality of the RNA, however, this was difficult to assess given the small sample size. For example, it is possible that different sub-clones within each sample respond differently to processing and sample storage and that some are more "fragile" than others and quickly lose viability. Finally, the issue of variable and poor RNA quality will again be addressed in the following chapter which presents an experiment culturing cells for 48-hours to determine if RNA quality can be improved following initial thawing of samples.

Despite the concerns over RNA quality, the qRT-PCR did confirm upregulation of *ITGA4* by 54x in the CD49d+ clone in patient 21 as expected. It was also expected that CD49d surface expression would correlate with expression of *ITGA4* at the mRNA level in the wider cohort of patients. Surprisingly, this was not the case. Zucchetto *et al.* [155] concluded that *ITGA4* mRNA expression did correlate with CD49d expression, however, the correlation was weak (r^2 0.6) and was made between percentage positive CD49d cells (and not on CD49d MFI) and used *ITGA4* mRNA expression on a log rather than linear scale for unclear reasons. It is therefore possible that there is

not a direct relationship between expression at the mRNA level and at the protein level. This could be explained by post-translational modifications, sub-clonal heterogeneity, storage of preformed CD49d or the time of sampling and life-cycle stage of the CLL cell. For example, it is possible that CD49d acts similarly to CD38: CD38 expression appears to change throughout the life cycle of a CLL [226]. Expression of CD49d could also be altered by the freeze-thaw process as has been previously reported [227]. Finally, the CD49d antibody used in this study recognises CD49d in both its active and inactive states and thus is unlikely the reason for the observed lack of correlation in RNA and protein expression.

Both RNAseq and qRT-PCR were used to determine gene expression differences in patient 6. The fold-change in gene expression between the two methods varied, as expected, given the inherent differences between the assays including their differing sensitivities and amplification efficiencies. Furthermore, read counts are normalised across the entire gene in RNAseq, whereas qRT-PCR will only amplify part of the gene of interest and may not amplify certain splice variants. Nevertheless, both methods showed concordant results in terms of up- or down-regulation of gene expression.

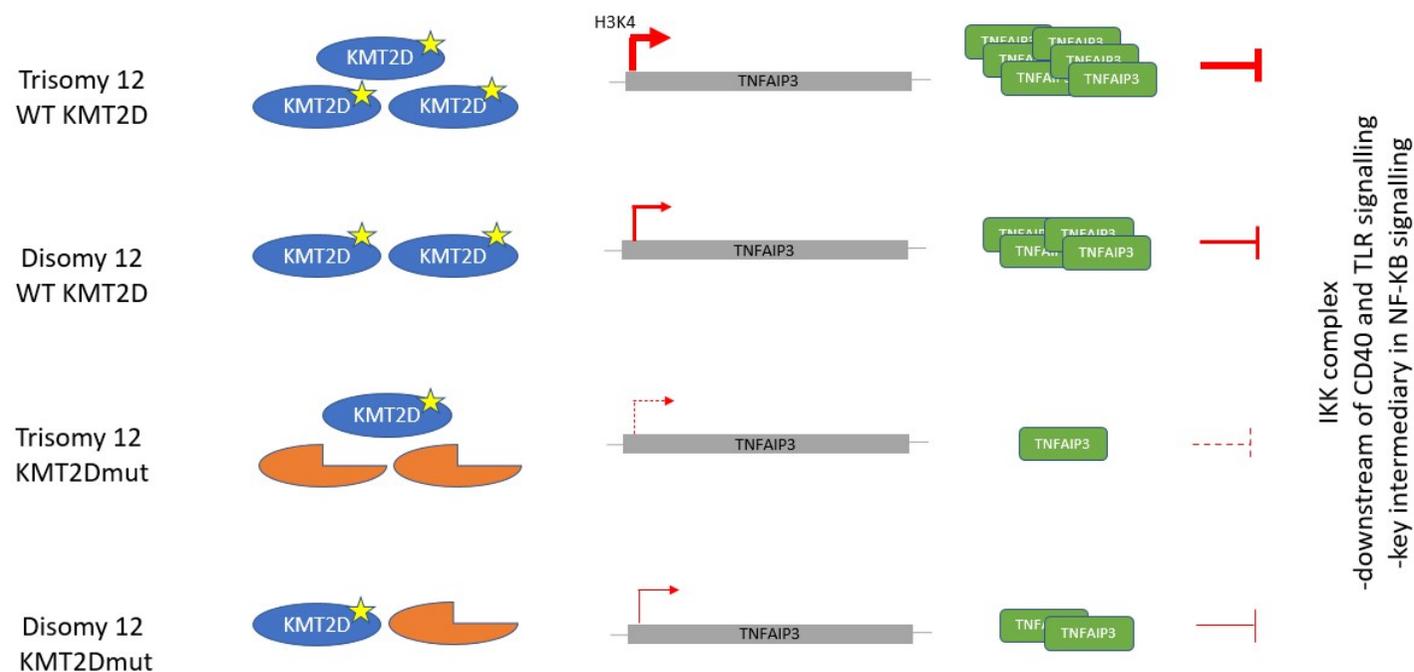
The *MDC1* and *BCL11B* variants detected in the whole exome sequencing of patient 6 were not expressed in the RNAseq data. It is possible that these genes are expressed at a different stage during the development of this patient's CLL and are usually downregulated during the process of lymphocyte maturation. For example, expression of the mutant *MDC1* in early lymphoid progenitors may have primed the development of further lesions that led to the malignant state, and then the wild-type and mutant *MDC1* alleles were silenced in mature lymphocytes once overt CLL developed. This is purely hypothetical, however, and would require a functional model to assess this possibility. Gene-editing techniques, such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology, could be used to introduce the mutation and assess its function.

The *KMT2D* variant on chromosome 12 identified in the WES was expressed in the CD49d+ trisomy 12 clone alone and was expressed at a level higher than *ITGA4*. It was expressed in 67% of the reads suggesting that it was present on 2 of the 3 copies of chromosome 12. This implies that the chromosome 12 with the variant was duplicated and the mutation arose prior to the duplication event. As hypothesised in the previous Chapter, this represents a novel driver mutation in trisomy 12 CLL and it is possible that some "fitness" advantage was conferred on cells then harbouring an extra copy of the variant. Alternatively, it may still represent a passenger mutation. The functional effects of the variant were not pursued in a cell-line model in the first instance: the potential importance of *KMT2D* was first investigated in the wider cohort to determine if it was applicable outside of the

singular case in patient 6. Sequencing of *KMT2D* in the whole cohort was not possible due to budgetary constraints: a next-generation sequencing panel would necessarily be utilised as the gene is large at 41.9 kb and there are no mutational hotspots (mutations are spread throughout the gene in non-Hodgkin lymphoma [228]). It is also acknowledged that over 500 CLL whole exomes have been sequenced and recurrent *KMT2D* mutations have not yet been reported. This could be explained by the fact that of these exomes, there are still relatively few trisomy 12 CLL exomes (<100 in total) and that the sequencing depth at the *KMT2D* gene was not adequate to detect variants (depth at this locus is not routinely reported).

Instead of sequencing *KMT2D* in the whole cohort, mRNA expression of *KMT2D* and one of its bona fide targets *TNFAIP3* [222] was first evaluated. It was hypothesised that any *KMT2D* expression would correlate positively with expression of *TNFAIP3*, and that samples with a protein truncating *KMT2D* mutation similar to that seen in patient 6 would not demonstrate a positive relationship between *KMT2D* and *TNFAIP3* expression (see Figure 5-11). Furthermore, it was hypothesised that any sample without the positive correlation would contain trisomy 12. The data did not confirm the hypothesis but neither did it reject it. The following four points are clear from the data: firstly, *KMT2D* expression was not higher in the trisomy 12 group compared to the disomy 12 group. It may be expected to be approximately 1.5x higher in the trisomy 12 group given its location on chromosome 12, however, not all genes on chromosome 12 appear to be uniformly over-expressed in this manner in trisomy 12 CLL [138]. This is reminiscent of the non-uniform upregulation of genes on chromosome 21 in trisomy 21 [229], however, it is noted that this is in the congenital rather than an acquired setting. Secondly, *TNFAIP3* expression was not significantly different between the trisomy 12 and disomy 12 cohorts. Thirdly, in keeping with the previously reported role of *KMT2D* in upregulation of *TNFAIP3* [222], a positive correlation was demonstrated between their expression. Fourthly, all 3 disomy 12 samples and 6 of 8 trisomy 12 samples demonstrated a positive relationship between *KMT2D* and *TNFAIP3* expression. Two trisomy 12 samples, however, had a negative relationship between *KMT2D* and *TNFAIP3* expression. It is possible that these samples bear a mutated *KMT2D* and would benefit from targeted sequencing of the gene. Confirming a mutant *KMT2D* by a different method, for example, by detecting a mutant truncated protein is not straightforward. For example, the *KMT2D* mutation in patient 6 is predicted to result in a truncated protein, however, due to the sheer size of the protein (593389 Daltons) it would not be easily visualised on a Western blot. An epigenetic approach to detect difference in the epigenome from a dysregulated *KMT2D*, a histone methyltransferase, would also be possible.

Figure 5-11. Hypothesised role of KMT2D in trisomy 12 CLL. In trisomy 12 CLL with wild-type (WT) KMT2D (top panel), there are three copies of an intact KMT2D protein with catalytic SET domain (yellow star). KMT2D functions to maintain the gene activation H3K4 mark and drives expression of target genes such as TNFAIP3. The TNFAIP3 protein in turn inhibits the IKK complex downstream of TLR and CD40 signalling, which leads to inhibition of NF- κ B signalling. This effect is more pronounced in trisomy 12 CLL than disomy 12 CLL (second panel). In disomy 12 CLL with a mutant KMT2D lacking the enzymatic domain (in orange; bottom panel), TNFAIP3 expression is not upregulated. In trisomy 12 CLL with two mutant copies of KMT2D, TNFAIP3 expression is downregulated to an even greater extent as the mutant proteins bind the usual KMT2D partners but do not have enzymatic activity. The result is an even smaller amount of TNFAIP3 available to inhibit the IKK complex and NF- κ B signalling. H3K4 = trimethylated histone H3 lysine 4; IKK = inhibitor of nuclear factor kappa B kinase; TLR = toll-like receptor.



Regulation of *ITGAA4* was also investigated. CD49d is a poor prognostic marker in CLL, however, it is highly expressed in normal B cells. It is possible that CD49d expression in trisomy 12 CLL is a marker of CLL cells that more closely aligned to healthy B cells and CD49d itself does not drive aggressive clinical behaviour. Certainly trisomy 12 has an epigenome that is more closely aligned to normal B cells [208, 230] than the other subtypes of CLL. As such, CD49d may be downregulated and “switched off” in CD49d- CLL rather than being upregulated and overexpressed in more aggressive CD49d+ CLL. The mechanism by which this occurs is not clear. Methylation of the gene promoter does not appear to be the sole mechanism: in the CD49d- clone of patient 21, no methylation was observed in the *ITGA4* promoter, and in the CD49d- clone of patient 6, only 34% methylation was detected. Single cell sequencing or cloning of PCR products following bisulfite sequencing was not performed (as per the initial report by Zucchetto *et al.* [155]). It is therefore possible that there were sub-clones within both patient samples that had a different degree of methylation undetected below the sensitivity of Sanger sequencing, however, >98% of cells in the CD49d- fractions showed no CD49d expression by flow cytometry so it would be assumed if methylation was a major mechanism of *ITGA4* downregulation, it would be detected in this manner. Other mechanisms of regulation were considered. Transcription factors that were differentially regulated between the CD49d+ and CD49d- clones in patient 6, and that were predicted to bind to the *ITGA4* gene promoter, were investigated. Two such transcription factors were identified: IRF4 and EZH2. *IRF4* was upregulated in the CD49d+ clone with respect to the CD49d- clone, and the reverse was true of *EZH2*.

IRF4 (previously *MUM1*) is a critical regulator of B cell maturation and a susceptibility locus for CLL: *IRF4* SNPs have been identified in genome-wide association studies in CLL [25, 231]. Furthermore, in a knock-out mouse model of *IRF4*, absent expression of *IRF4* was shown to lead to the spontaneous development of CLL with 100% penetrance [231]. Conversely, *IRF4* has oncogenic effects in plasma cell myeloma [232] and overexpression of *IRF4* is the result of a recurrent chromosomal translocation in this disease, t(6;14) [233]. Despite its relevance to haematological malignancy, *IRF4* expression was not shown to correlate with *ITGA4* expression in this thesis and was not differentially regulated between CD49d+ and CD49d- CLL. *EZH2* is the catalytic component of the polycomb repressive complex 2 (PRC2) and functions as a lysine methyltransferase, establishing the histone H3 lysine 27 (H3K27) trimethylation repressive epigenetic signature [234]. Furthermore, *EZH2* has been shown to be overexpressed in high-risk U-CLL and increase with disease progression [235]. It is also an important factor in other B cell malignancies: gain-of-function *EZH2* mutations are reported at high frequency in follicular non-Hodgkin lymphoma and germinal-centre type diffuse

large B cell lymphoma [236, 237]. *EZH2* expression was increased in the CD49d- clone in patient 6 and it was hypothesised that it could be repressing *ITGA4* expression. This could not be confirmed, however, in patient 21 (in which *EZH2* expression was similar between CD49d+ and CD49d- clones) or the wider cohort: *EZH2* expression did not correlate with *ITGA4* expression nor show increased expression in the CD49d- cases. Nonetheless, there was a weak negative correlation between *EZH2* expression and CD49d MFI that may indicate a trend for downregulation of CD49d with increased *EZH2* expression; this requires exploration in a larger data set.

One of the key limitations of this body of work is again the small sample size and heterogeneity of the samples, both in terms of their underlying biology and sample quality. There was marked variability in expression of all the targets and the analysis was hampered by sample availability. Despite the absence of any clear differences detected in the trisomy 12 and disomy 12 CLL groups in this chapter, it was postulated that changes relevant to trisomy 12 CLL may not be detected in CLL cells circulating in their “steady state” and the critical pathways may be revealed following cellular activation, mimicking microenvironmental signals that support CLL cell survival and proliferation *in vivo*. One such potentially important pathway identified in the RNAseq gene-set enrichment analysis was toll-like receptor signalling, which is investigated in the following chapter at baseline and following pathway activation.

6 INVESTIGATION OF TOLL-LIKE SIGNALLING IN TRISOMY 12 CLL

6.1 INTRODUCTION

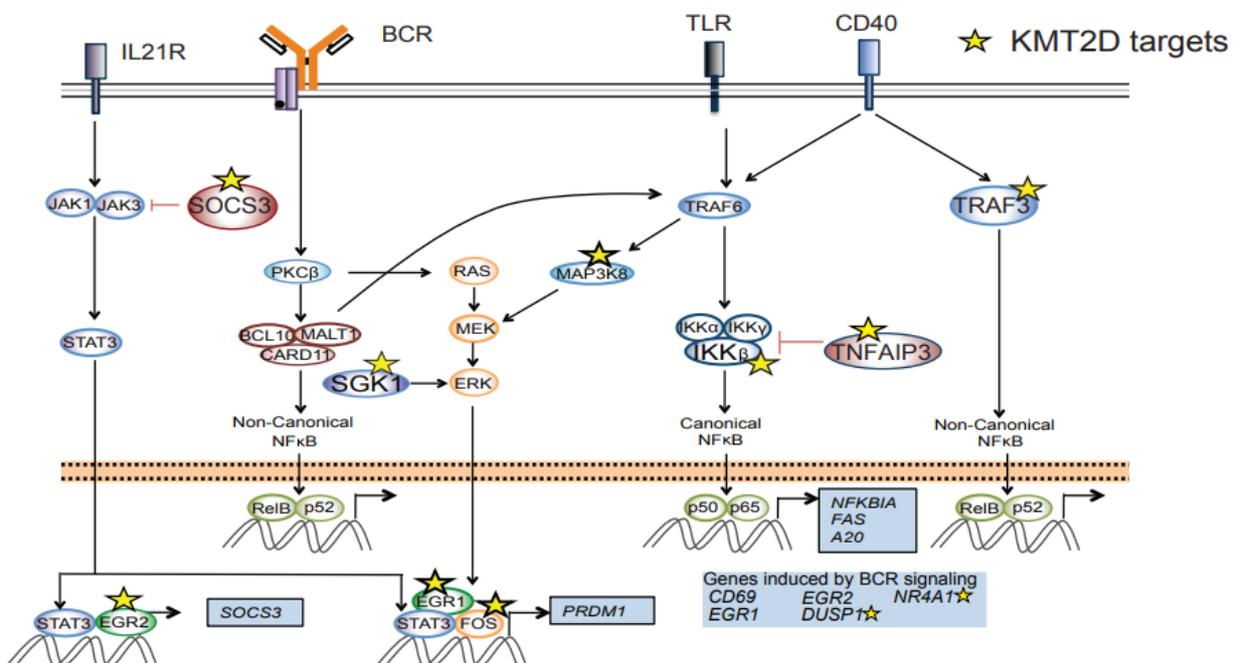
Toll-like receptor signalling was the most differentially regulated pathway between the CD49d+ trisomy 12 M-CLL and CD49d- disomy 12 U-CLL clones in patient 6. Toll-like receptors (TLRs) are integral components of the innate immune system which act immediately to initiate host-defence mechanisms upon encountering a pathogen [5]. They have limited specificity compared to the adaptive immune system and recognise Pathogen Associated Molecular Patterns (PAMPs) that are pathogen-specific and conserved throughout evolution. There are 10 TLR genes in humans that express 10 different TLRs that each respond to different molecular patterns and reside either on the cell surface or intra-cellularly on endosomal membranes. For example, TLR9 on the endosomal membrane recognises microbial unmethylated CpG dinucleotides. This has been harnessed in the laboratory by the utilisation of CpG oligonucleotides to induce CLL cell metaphases for chromosomal banding analysis [238, 239]. TLR4, along with its accessory proteins MD2 and CD14, recognise lipopolysaccharides (LPS) derived from the outer membranes of gram-negative bacteria and lipoteichoic acids of gram-positive bacteria. TLR4 is expressed on B lymphocytes, macrophages, and dendritic cells, and *TLR4* mRNA was the most upregulated in the trisomy 12 clone of patient 6. Upon LPS binding to TLR4/MD2, homodimerisation of the complex ensues and leads to activation of downstream pathways including NF- κ B signalling and the production of inflammatory cytokines and Type 1 interferons as part of the anti-microbial response.

TLR expression in CLL has been previously investigated. Muzio et al. [240] demonstrated the expression of TLR1, TLR2, TLR6 and TLR10 in 8 cases of CLL and showed activation of NF- κ B signalling and protection from spontaneous apoptosis after stimulation of TLR1, TLR2 and TLR6 with bacterial lipopeptides. *TLR4* mRNA was undetectable, however, it is unclear if any of the 8 cases investigated expressed CD49d or contained trisomy 12. Arvaniti et al. [241] investigated TLRs in a larger series of 192 CLL cases and confirmed expression of TLRs in CLL along with high expression of members of the NF- κ B signalling and IRF pathways. Endogenous inhibitors of TLR signalling were low to undetectable, and it was concluded that TLR signalling was competent in CLL. There was variability in expression of all TLRs, with TLR7 most highly expressed. TLR4 was low to undetectable in the majority of the cohort, however, there was significant variation in its expression in cases positive for TLR4. Again, it is unclear if the positive cases expressed CD49d or harboured trisomy 12. There were differences between U-CLL and M-CLL cases and the different BCR stereotype subsets, suggesting that “CLL clones with distinctive antigen reactivity are able to respond in a distinct fashion [sic] to different members of the TLR family, alluding to subset-biases recognition of and selection by the respective ligands”. Notably, there was a significant increase in *TLR4*

mRNA expression in M-CLL compared to U-CLL, which mirrors the relationship observed in the RNAseq data of patient 6. Rybka et al. [242] showed that TLR4 expression was lower on CLL cells than on B cells from healthy controls and was lower in CLL with advanced Rai stage.

There is also a connection between TLR4 and NOTCH signalling, which is of interest in the context of trisomy 12 CLL given the known association between mutations of *NOTCH1* and trisomy 12. Zhang et al. [243] demonstrated that TLR4 signalling enhanced NOTCH signalling, which in turn negatively regulated TLR4 triggered inflammation, but was dependent upon the NICD PEST domain (which is truncated in the prototypical frameshift *NOTCH1* mutation in CLL). It should be noted, however, that this finding may not be applicable to CLL as experiments were performed in macrophage cell lines and not B lymphocytes nor CLL cells. Finally, a connection between KMT2D and TLR signalling has also been proposed. Ortega-Molina et al. [222] showed KMT2D controlled expression of multiple key regulators of the TLR pathway including TNFAIP3 (see Figure 6-1).

Figure 6-1. KMT2D targets of key B cell regulatory pathways. KMT2D targets TNFAIP3, an inhibitor of the IKK complex which is involved in NF- κ B signalling and is downstream of TLR. Reprinted by permission from Springer Nature: Nature Publishing Group. *Nature Medicine* vol. 21, pp. 1199-1208 (The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development, Ortega-Molina et al.). © Nature American 2015 [222]



Therefore, TLR signalling is competent albeit variable in CLL and can induce a variety of outcomes, including proliferation, apoptosis or anergy reversal [244]. Furthermore, TLR signalling is an alternative mechanism of activating B cells other than via the BCR but the expression and function of members of the TLR family appears to relate to the structure of the BCR and specific antigen reactivity of the CLL clone [241]. The relationship between TLR signalling and trisomy 12 CLL has not been previously investigated. The RNAseq data from patient 6 showed that TLR signalling was upregulated in the CD49d+, trisomy 12, IGHV and *KMT2D*-mutated clone. As such, expression of specific members of TLR signalling was investigated at baseline and following receptor stimulation in a cohort of trisomy 12 and disomy 12 CLL.

6.2 METHODS

6.2.1 TLR4, IL8 and CD14 expression analysis

Flow cytometry was performed for the surface markers TLR4 and CD14 (part of the TLR4 complex) in a cohort of 17 trisomy 12 and 11 disomy 12 CLL patient samples. Three healthy controls were also utilised (see Chapter 2). Following thawing and washing of cryo-preserved PBMC preparations (as outlined in Chapter 2), 5×10^5 cells were incubated with 15 μ L of the cocktail listed in Table 6-1 for 20 minutes in the dark. Mean MFI and percentage positive cells for both TLR4 and CD14 were recorded for the CD5/CD19+ CLL cell population. The mean MFI from unstained cells was subtracted from the stained MFI reading to account for auto-fluorescence: as such, some readings are negative values as the stained CLL population has an MFI less than the MFI of the unstained total lymphoid population (gated on forward and side scatter only). Flow cytometry experiments were performed on the FACS Aria Fusion flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA) and data analysis performed using FlowJo™ software (BD Biosciences) with the assistance of Dr Giles Best (College of Medicine & Public Health, Flinders University).

Gene expression analysis of *TLR4*, *CD14*, and *IL8* was performed by qRT-PCR as outlined in Chapter 2 with the primers listed in the Appendix (see Table 9-3). RNA was extracted from left-over cells following thaw that were not used in the concurrent flow cytometry experiments.

Table 6-1. Antibody cocktail for TLR4 and CD14 expression analysis.

Antibody	Fluorophore	Titre	Volume added to cocktail (μL)	Brand (clone)
CD5	V450	1:4	10	BD (L17F12)
CD14	FITC	1:2	20	Pharmingen
CD49d	PE-Cy5	1:2	20	Biolegend
TLR4	PE	neat	40	ThermoFisher (HTA125)
CD19	PE-Cy7	1:2	20	BD (SJ25C1)
CD45	APC-H7	1:8	5	BD (2D1)
<i>Cocktail made up to 120μL with 5μL of FACS buffer</i>				

Following 48 hours of culture, cells were harvested for RNA or used in flow cytometry experiments. Cell viability was assessed using the mitochondrial membrane potential dye, 1,1',3,3',3',3'-hexamethylindodicarbo-cyanine iodide (DiIC₁(5); Cayman Chemicals, Ann Arbor, MI, USA) and propidium iodide (PI; Sigma Aldrich). 3 μL of 1.6 μM DiIC₁(5) was added to 100 μL of cells and incubated in the dark for 10-15 minutes. 3 μL of 1mM PI was then added, and the cells were analysed on the CytoFLEX flow cytometer (Beckman Coulter). Cells that were positive for DiIC₁(5) but negative for PI were assigned as early apoptotic; cells that were negative for both markers were viable; and cells that were negative for DiIC₁(5) and positive for PI were non-viable, dead cells. Cells not used in the viability assay were centrifuged (300 x g for 5 min), resuspended in 100 μL of media and stained with 15 μL of the TLR4 antibody cocktail (see Table 6-1). Following 20 minutes of incubation in the dark, samples were read on the CytoFLEX flow cytometer (Beckman Coulter) and CD14 and TLR4 MFI values were determined for the CD5/CD19+ CLL cell population. The following gating strategy was employed: lymphocytes were gated based on their forward (FSC-A) and side scatter (SSC-A) properties, CLL cells were gated based on CD5 and CD19 co-expression, and single cells were gated based on FSC-H versus FSC-A.

RNA was extracted following 48-hours of culture, cDNA was generated, and qRT-PCRs were performed using the methods outlined in Chapter 2. The following genes of interest were investigated: *TLR4*, *IL8*, *KMT2D*, *TNFAIP3* and *ITGA4*. Primers are listed in the Appendix in Table 9-3 RNA extracted from 8 samples cultured under the control conditions (patients 2, 12, 14, 15, 18, 73, DIS4, and DIS5) was also analysed on the Agilent Bioanalyzer to assess quality.

6.2.2 Statistics

Relative mRNA expression was normalised to the mean of the trisomy 12, CD49d+ or U-CLL groups depending on the comparison. For example, comparisons of gene expression at baseline between trisomy 12 and disomy 12 CLL, were conducted by normalising the mean of the trisomy 12 group to 1.0. For baseline expression of mRNA or protein (as determined by MFI), the two groups were compared using an unpaired Student's t-test with Welch's correction (not assuming equal standard deviations). For the comparison of TLR4 MFI between the three groups (trisomy 12 CLL, disomy 12 CLL and healthy controls), a one-way ANOVA was performed. Simple linear regressions were used to determine correlation between two continuous variables. For the stimulation experiments, cell viability, mRNA and protein expression were compared using a 2-way ANOVA to determine how the response differed between biological groups (for example, trisomy 12 and disomy 12 CLL) and between cellular conditions (control and LPS-stimulated). A p value of <0.05 was considered statistically significant and all statistical analyses were performed using GraphPad Prism software (versions 8 and 9).

6.3 RESULTS

6.3.1 Expression of TLR signalling components in RNAseq from patient 6

The gene set enrichment analysis performed on the RNAseq data showed differential regulation of the toll-like signalling between the CD49d+ and CD49d- clones of patient 6 (see Figure 5-3). Nine genes were identified in this analysis and are presented in Table 6-2. *TLR4* mRNA was increased in the CD49d+ clone by 34-fold (see Table 6-2 and Figure 6-2).

The Kegg (Kyoto Encyclopedia of Genes and Genomes) toll-like receptor signalling pathway [245] is shown in Figure 6-3. Expression levels of *CD14*, *TLR4* and *IL8* were investigated in further samples: *CD14* and *TLR4* were chosen as they were among the most dysregulated and because they are cell-surface markers amenable to expression analysis by flow cytometry. *IL8* was also chosen as a downstream "end-result" of pathway activation as it was both highly expressed (1173 reads) and highly upregulated in the CD49d+ clone in the RNAseq data (fold change of 338-times).

Table 6-2. Expression levels and read counts of genes in toll signalling pathway in the RNAseq of patient 6. *denotes genes that are negative regulators of toll signalling pathways. Bold genes were selected for qRT-PCR confirmation.

Gene	Normalised read count (CD49d+ clone)	Normalised read count (CD49d- clone)	Relative expression in CD49d+ clone	Chromosome
Toll receptor signalling pathway				
TLR8	231	9	25	Xp
<i>IRAK3*</i>	109	3	32	12q
CD14	125	6	19	5q
<i>PTGS2</i>	388	1	261	1q
TLR4	348	10	34	9q
<i>DAB2IP*</i>	10	196	0.05	9q
<i>LILRA2*</i>	111	1	75	19q
<i>LILRAR4*</i>	30	944	0.03	19q
<i>PTPRS</i>	25	807	31	19p

Figure 6-2. RNAseq coverage of TLR4 for the CD49d+ and CD49d- clones of patient 6. The whole gene is visualised. The top track shows piled reads (coverage) for CD49d- clone, the second track shows piled reads (coverage) for the CD49d+ clone and the bottom tracks shows the reference gene with exons represented in blue boxes.

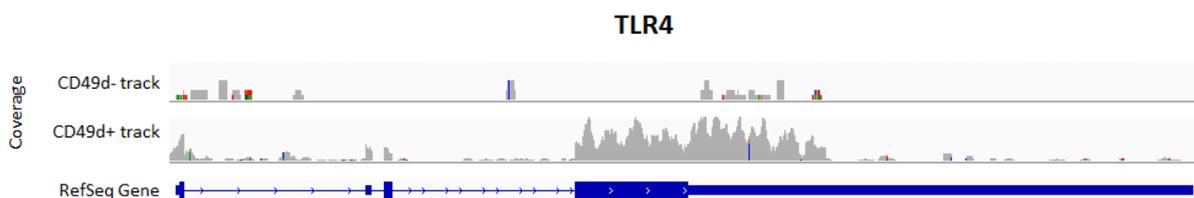
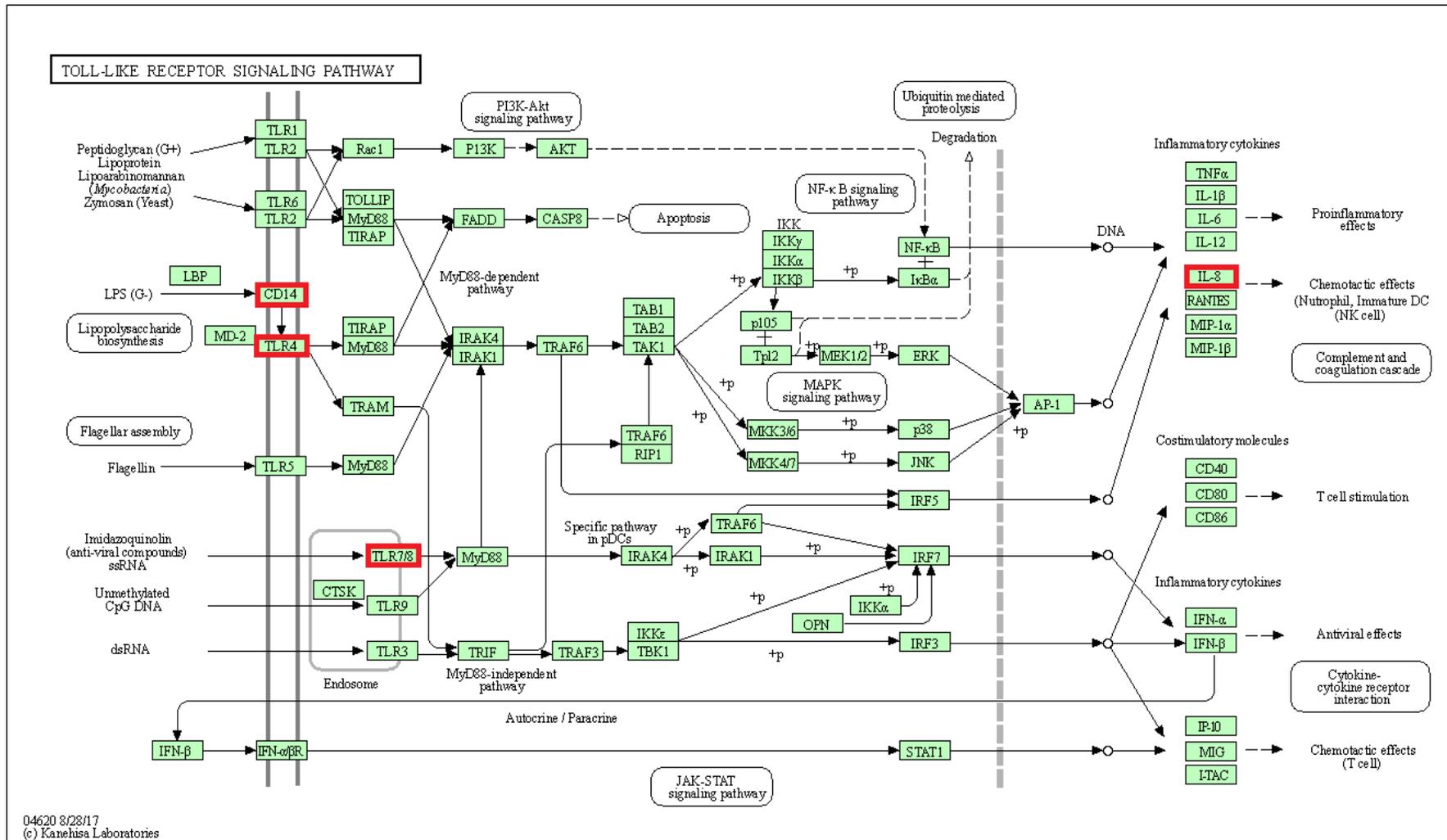


Figure 6-3. The KEGG toll-like receptor signalling pathway [245]. Targets differentially expressed between the two clones in the RNAseq data of patient 6 are boxed in red: CD14, TLR4, TLR8 and IL8. Both CD14 and TLR4 are involved in response to LPS (lipopolysaccharide) and result in pro-inflammatory cytokine release.

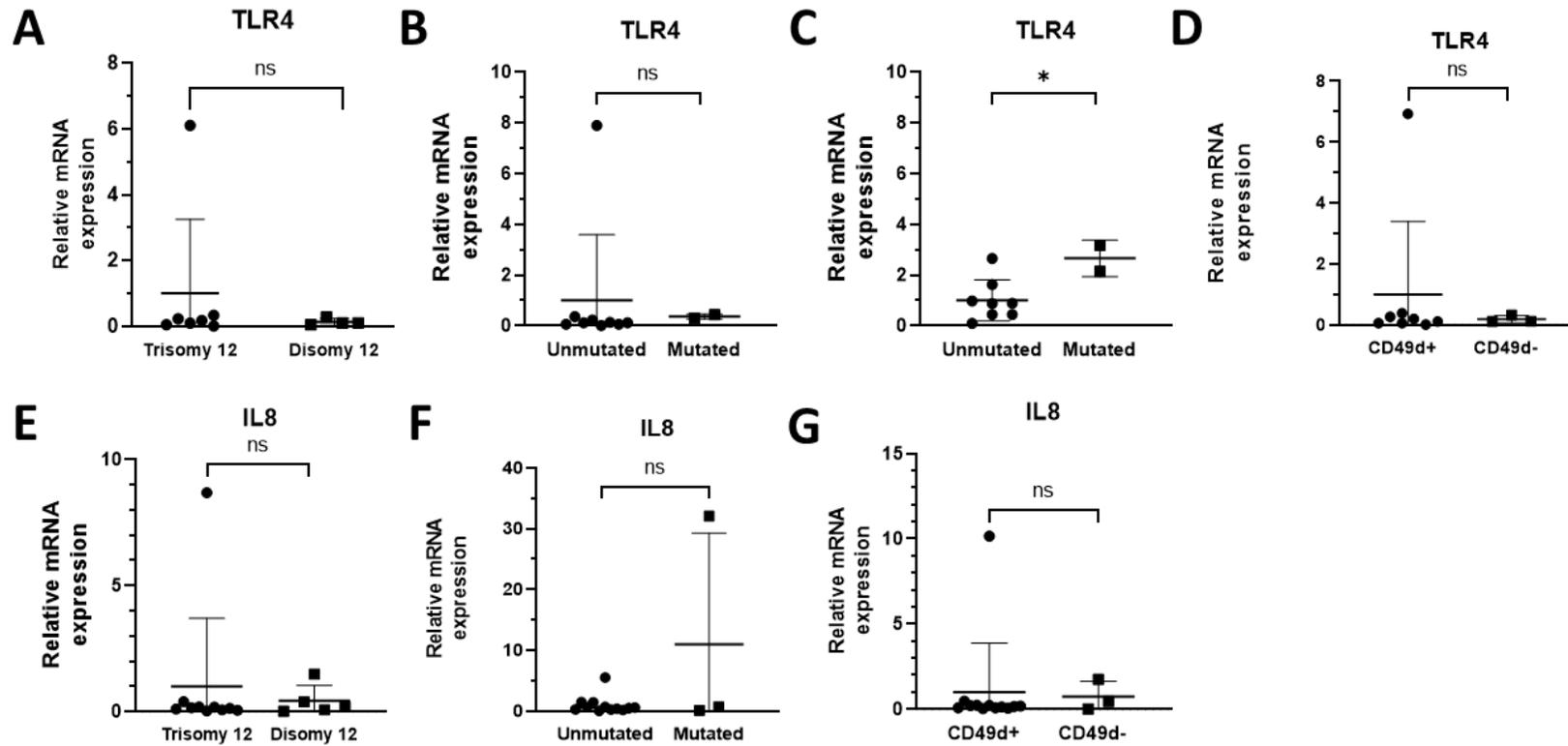


6.3.2 Expression of TLR4 pathway targets at baseline

6.3.2.1 Gene expression of TLR4 and IL8

The expression of *TLR4* and *IL8* mRNA was investigated in a cohort of trisomy 12 and disomy 12 CLL with varying expression of CD49d and IGHV-mutational status at baseline (prior to culturing). *CD14* and *TLR8* expression analysis was not performed due to difficulties in primer optimisation with multiple amplified products using three different sets of primers (data not shown). Raw gene expression results are presented in the Appendix (see Table 9-16). There was no statistically significant difference in relative mRNA expression of either *TLR4* or *IL8* between a cohort of trisomy 12 and disomy 12 CLL (see Figure 6-4A,E). The same was true when comparing CD49d+ and CD49d- cases (see Figure 6-4D,G) and IGHV-unmutated and IGHV-mutated cases (see Figure 6-4B,F). There did appear to be a trend towards an upregulation of *TLR4* in IGHV-mutated CLL, and even though there were only two samples in the latter subgroup, statistical significance ($p=0.03$) was achieved if one large outlier was removed from the IGHV-unmutated cohort (see Figure 6-4C). This outlier was patient 21 who had a relative *TLR4* expression of 42.34x higher than the housekeeping gene (7.9x higher than the mean) as opposed to all other samples having relative expression of 2x or less. Other than being bimodal for CD49d, no other obvious clinical or molecular characteristics could account for the large change (note that patient 21 did have a *NOTCH1* mutation, but so did patient 11 who had a relative *TLR4* expression of only 0.07x). Expansion of case numbers was attempted, however, due to variable RNA quality (see 5.3.1.5.5), few samples from IGHV-mutated patients in the tissue bank, varying degrees of *TLR4* amplification and time constraints, this was not achieved.

Figure 6-4. Relative mRNA expression of TLR4 and IL8 at baseline. Individual samples are represented as dots/squares. Means and standard deviations are plotted. The means of the left-hand group of each graph have been normalised to 1.0. ns= not significant; *=*p* value <0.05. (A) TLR4 expression in trisomy 12 and disomy 12 CLL. (B) TLR4 expression in IGHV-unmutated and IGHV-mutated CLL. (C) The same analysis as B, removing one outlier (patient 21) from the unmutated-IGHV group. (D) TLR expression in CD49d+ and CD49d- CLL. (E) IL8 expression in trisomy 12 and disomy 12 CLL. (F) IL8 expression in IGHV-unmutated and IGHV-mutated CLL. (G) IL8 expression in CD49d+ and CD49d- CLL.



6.3.2.2 Protein expression of TLR4 and CD14

TLR4 cell surface expression was examined in the CD49d+ and CD49d- CLL cell populations in patient 6 (see Figure 6-5). The mean TLR4 MFI was not significantly different between the two populations (226 in the CD49d+ clone vs. 195 in the CD49d- clone) and cells positive for TLR4 did not distribute unevenly into the CD49d+ population. Despite this, TLR4 expression was significantly higher in an expanded trisomy 12 CLL cohort (n = 18) compared to a disomy 12 CLL cohort (n = 11) as measured by MFI (p=0.007, see Figure 6-6A) or percentage positive cells (mean trisomy 12 = 76%, mean disomy 12 57%, p=0.005, see Figure 6-6C). Raw data is presented in Table 9-14 of the Appendix. There was a wide variability in expression of TLR4 with a range of 1878 (mean±sd=1380±556) in the trisomy 12 group and a range of 1322 (mean±sd=701±510) in the disomy 12 group. TLR4 MFI was also significantly higher in CD49d+ CLL compared to CD49d- CLL (p=0.02, see Figure 6-6D), but there was no difference between IGHV-unmutated and IGHV-mutated CLL (see Figure 6-6E).

TLR expression was similar between the trisomy 12 cohort (n = 18) and healthy controls (n = 3) (see Figure 6-6A). However, TLR4 expression was significantly lower in the disomy 12 CLL cohort compared to the healthy controls (p=0.004; see Figure 6-6A).

CD14 expression (as measured on the CLL cell population, excluding any contaminating monocytes) was significantly increased in trisomy 12 CLL compared to disomy 12 CLL (p=0.003, see Figure 6-6D) but again showed large variability with ranges of 132 (mean±sd=33±28) and 131 (mean±sd=-13±48) respectively. CD49d expression (as determined by mean MFI) did not correlate with TLR4 or CD14 expression ($r^2=0.02$ and 0.01 respectively; see Figure 6-7A,B), even when considering the trisomy 12 subgroup alone ($r^2=0.02$; see Figure 6-7D). There was a positive correlation between TLR4 expression and its accessory molecule CD14 ($r^2=0.4$, p=0.0002; see Figure 6-7C).

Figure 6-5. TLR4 expression on the CD49d+ and CD49d- CLL clones in patient 6. Lymphocytes were gated based on forward (FSC-A) and side (SSC-A) scatter. Single cells were identified based on forward scatter height (FSC-H) and area (FSC-A). CLL cells were gated based on co-expression of CD5 and CD19. CD49d+ CLL cells had a mean TLR4 MFI of 226 and CD49d- CLL cells had a mean TLR4 MFI of 195.

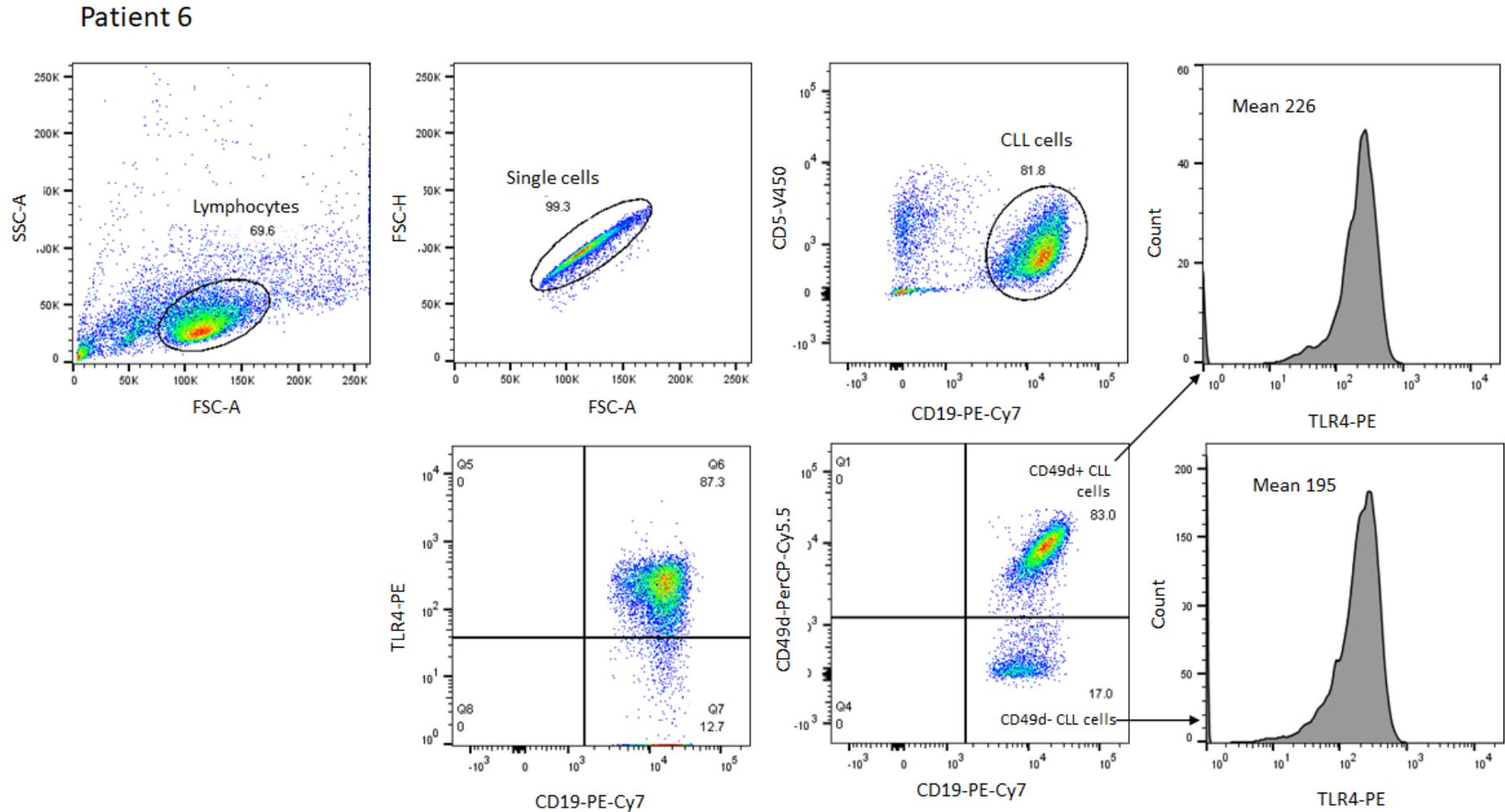


Figure 6-6. TLR4 and CD14 cell surface expression. Individual samples are plotted as dots, squares, and triangles. Means and standard deviations are plotted. ns=not significant; * and ** are statistically significant with p -values of <0.05 . MFI = mean fluorescence intensity (A) TLR4 MFI comparison between trisomy 12, disomy 12, and healthy controls. (B) CD14 MFI comparison between trisomy 12 and disomy 12 CLL. (C) TLR4 percentage positive cells in trisomy 12 and disomy 12 CLL. (D) TLR4 MFI in CD49d+ and CD49d- CLL. (E) TLR4 MFI between IGHV-unmutated and IGHV-mutated CLL.

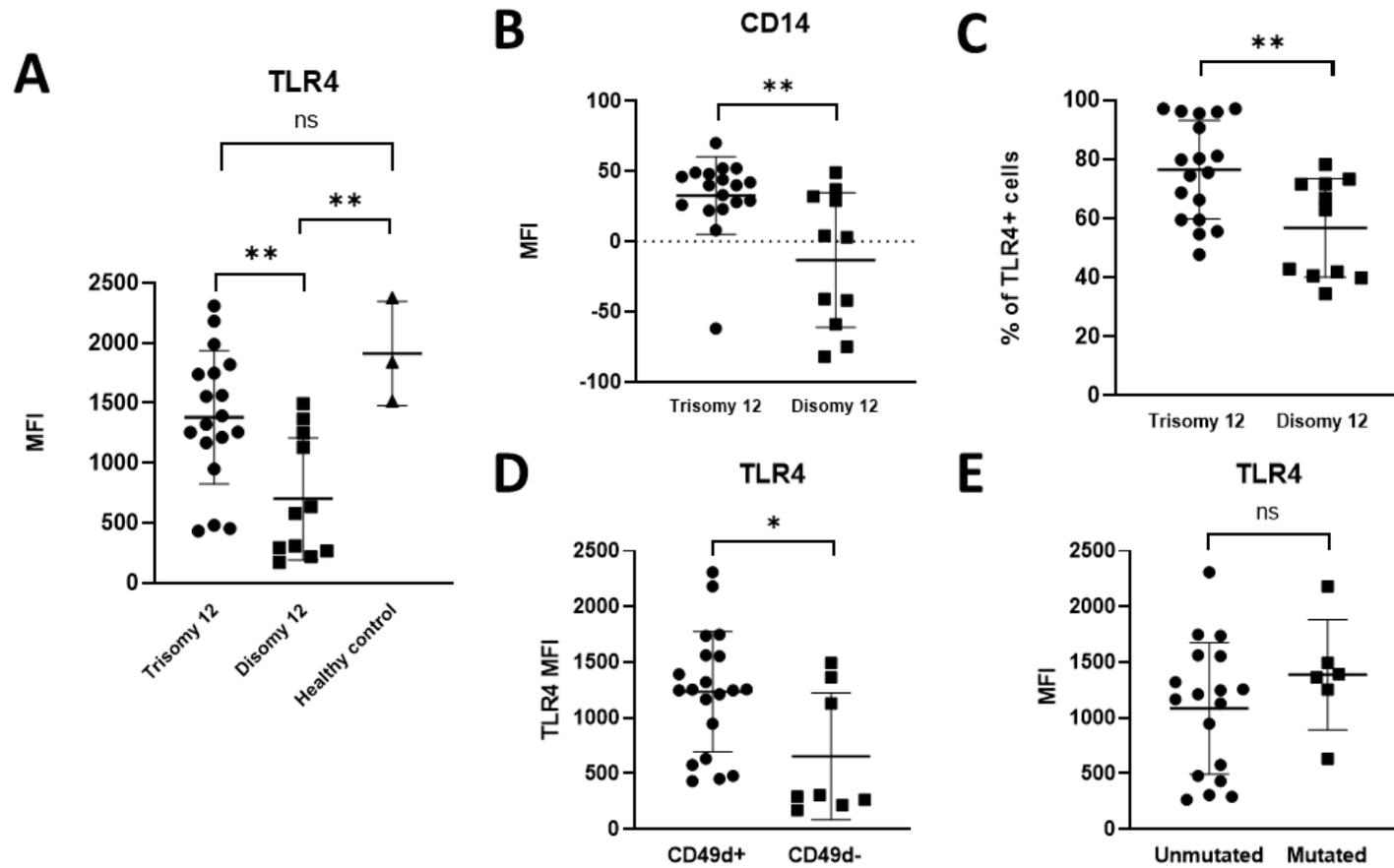
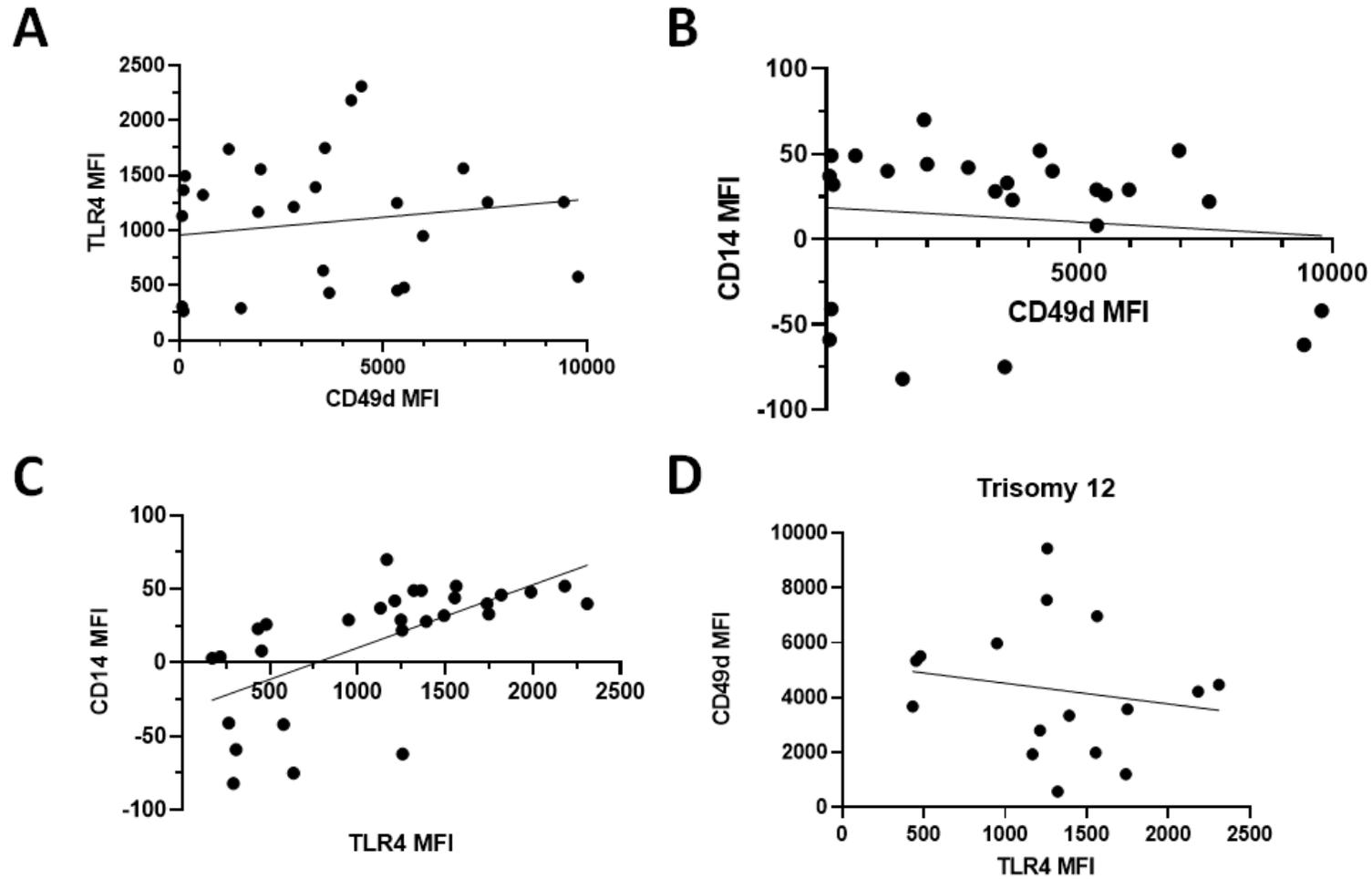


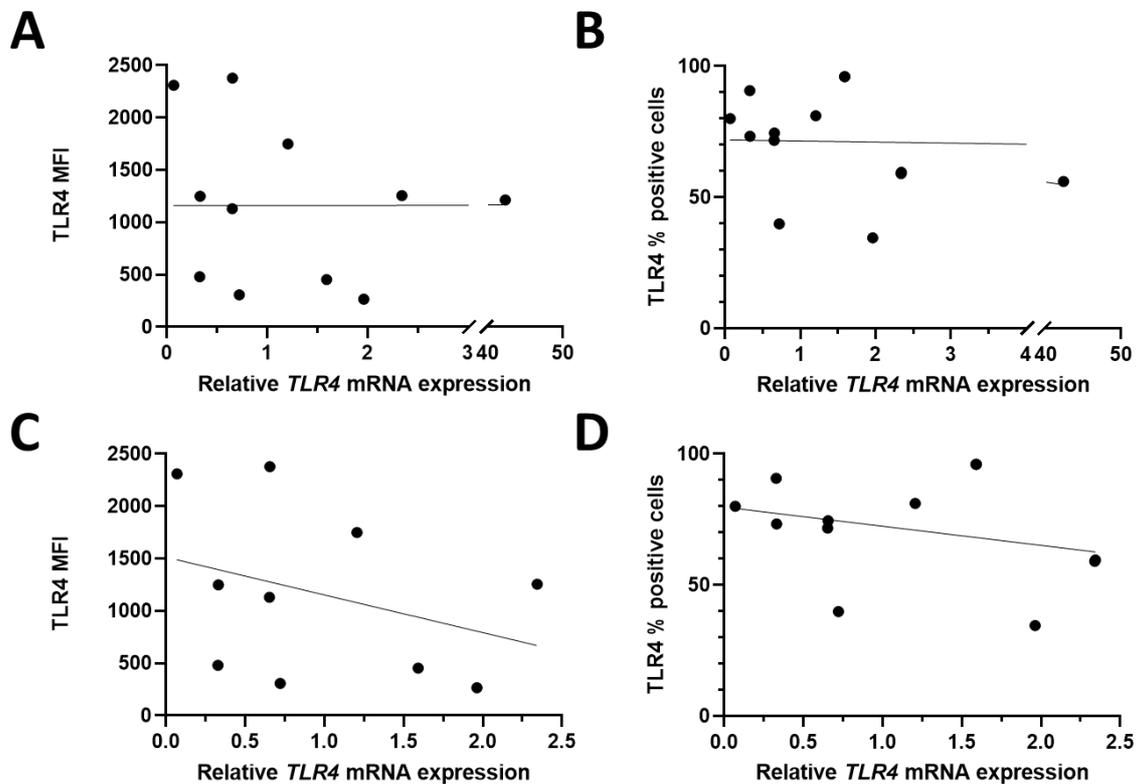
Figure 6-7. Correlations between CD49d, TLR4 and CD14 protein expression as determined by flow cytometry. MFI = mean fluorescence intensity. Individual samples are represented as dots. Line of best fit (simple linear regression) is plotted. There are negative values for CD14 MFI as the unstained total lymphocyte population had a higher MFI then the stained CLL population in some instances (and the difference is plotted). (A) TLR4 versus CD49d. (B) CD14 versus CD49d. (C) CD14 versus TLR4. (D) CD49d versus TLR4 in trisomy 12 subgroup alone.



6.3.2.3 Correlation of gene and protein expression

Unexpectedly, there was no correlation between relative *TLR4* mRNA expression and protein expression (measured as either mean MFI or percentage positive cells; see Figure 6-8A,B). The outlier (patient 21) was again removed from the data set but this did not change the outcome and no correlation was found (see Figure 6-8C,D).

Figure 6-8. Correlation between *TLR4* mRNA and protein expression. Individual samples are represented as dots. Line of best fit (simple linear regression) is plotted. MFI = mean fluorescence intensity. (A) *TLR4* MFI versus mRNA expression. (B) *TLR4* percentage positive cells versus mRNA expression. (C) *TLR4* MFI versus mRNA expression after removing the outlier, patient 21. (D) *TLR4* percentage positive cells versus mRNA expression after removing the outlier, patient 21.



6.3.3 Stimulating TLR4 with lipopolysaccharide (LPS)

Given the differences between TLR4 and CD14 expression at baseline (at least at the protein level), it was hypothesised that samples would respond differently to stimulation of the TLR4 pathway by LPS, a potent stimulator of the innate immune system. CLL patient samples were stimulated with LPS for 48-hours and subjected to cell viability analysis, immunophenotyping (including CD49d, TLR4 and CD14) and gene expression analysis via qRT-PCR (*ITGA4*, *TLR4*, *KMT2D*, *TNFAIP3* and *IL8*). A total of

14 samples were selected for the stimulation assay representing a range of CLL subtypes with at least 3 samples per category of: trisomy 12, disomy 12, CD49d+, CD49d-, IGHV-unmutated and IGHV-mutated. Data was not obtained on all 14 samples given low cell viability immediately post-thaw in 3 samples (<10%) and/or infected samples (flow cytometry results were obtained for 11/14 samples; see Table 9-17 in the Appendix). In addition to this, there was no reproducibility in amplification of the housekeeping gene *GUSB* in 3 of these 11 samples, so mRNA expression data was generated on 8 of 11 samples (see Table 9-18 in the Appendix).

6.3.3.1 RNA quality

Firstly, RNA quality was addressed following culturing of samples for 48-hours under control conditions given the variability in RNA quality observed directly post-thaw (see 5.3.1.5.5): it was hypothesised that culturing following thaw would increase RNA quality. RNA extracted from 8 samples after 48-hours in culture under control conditions was analysed. No conclusions could be drawn as to effect of culturing on RNA quality as none of the samples had a measurable RIN following 48-hours in culture. This was thought to be due to the low absolute quantity and high dilution of the RNA following culturing experiments – starting cell numbers were <1x10⁶ cells on all occasions at 48-hours.

6.3.3.2 Cell viability

Cell viability was assessed by Trypan blue exclusion assay immediately post-thaw. It was not assessed by flow cytometry at t=0h to maximise cell numbers available for the stimulation assay. Cell viability is presented in Table 9-17 and varies widely from 23-88% (n=11; mean±sd=44.1±22.7). After 48-hours cell viability (of the CLL cell population alone) was assessed by flow cytometry (using DilC₁(5)/PI; see example in Figure 6-9). There was no significant difference in cell viability at 48-hours between the control or LPS-stimulated samples (see Figure 6-9). This was true of all subgroups – trisomy 12, disomy 12, IGHV-unmutated and IGHV-mutated CLL. Viability after 48-hours was generally low with control means of 20% and 33% for the trisomy 12 and IGHV-unmutated cohorts, respectively.

6.3.3.3 Surface immunophenotype

Expression of surface TLR4, CD49d and CD14 was also analysed after 48-hours of LPS stimulation (see Figure 6-10 and Table 9-17). Again, there was no significant difference in expression of any of these markers between the control and LPS-stimulated samples for any subgroup (see Figure 6-10A,B,D,E). TLR4 expression was increased following LPS-stimulation in 5/11 samples (4/7 trisomy 12 samples, and 1/4 disomy 12 samples; see example for patient 73 in Figure 6-10C).

6.3.3.4 Gene expression

The gene expression changes following LPS stimulation are presented in Figure 6-11 and Table 9-18. There were no significant changes in mRNA expression in *TLR4*, *ITGA4*, *KMT2D*, *TNFAIP3* or *IL8* following stimulation with LPS in either the trisomy 12 or disomy 12 subgroups (see Figure 6-11A,C,D,E,F). No change was identified in TLR4 expression following LPS stimulation when comparing unmutated-IGHV and mutated-IGHV CLL (Figure 6-11B). There were, however, only 1 to 2 biological replicates in some comparator arms (see Figure 6-11), limiting the power of the analysis.

Figure 6-9. Cell viability after 48-hours stimulation with LPS. (A) Example dot plot of control sample at 48-hours (patient 7). 40.16% of CLL cells are viable (*DilC₁(5)* positive, PI negative), 8.46% are early apoptotic (*DilC₁(5)* negative, PI negative) and 51.14% are dead (*DilC₁(5)* negative, PI positive). (B) Example dot plot of LPS-stimulated sample at 48-hours (patient 7). There is an increase in viable cells to 66.36%. (C) Cell viability (%) of trisomy 12 and disomy 12 CLL control samples (blue) versus LPS-stimulated samples (red). (D) Cell viability (%) of IGHV-unmutated and IGHV-mutated CLL control samples (blue) versus LPS-stimulated samples (red). Individual samples are dots. Means±sd are plotted as columns±error bars. ns = not significant (2-way ANOVA analysis).

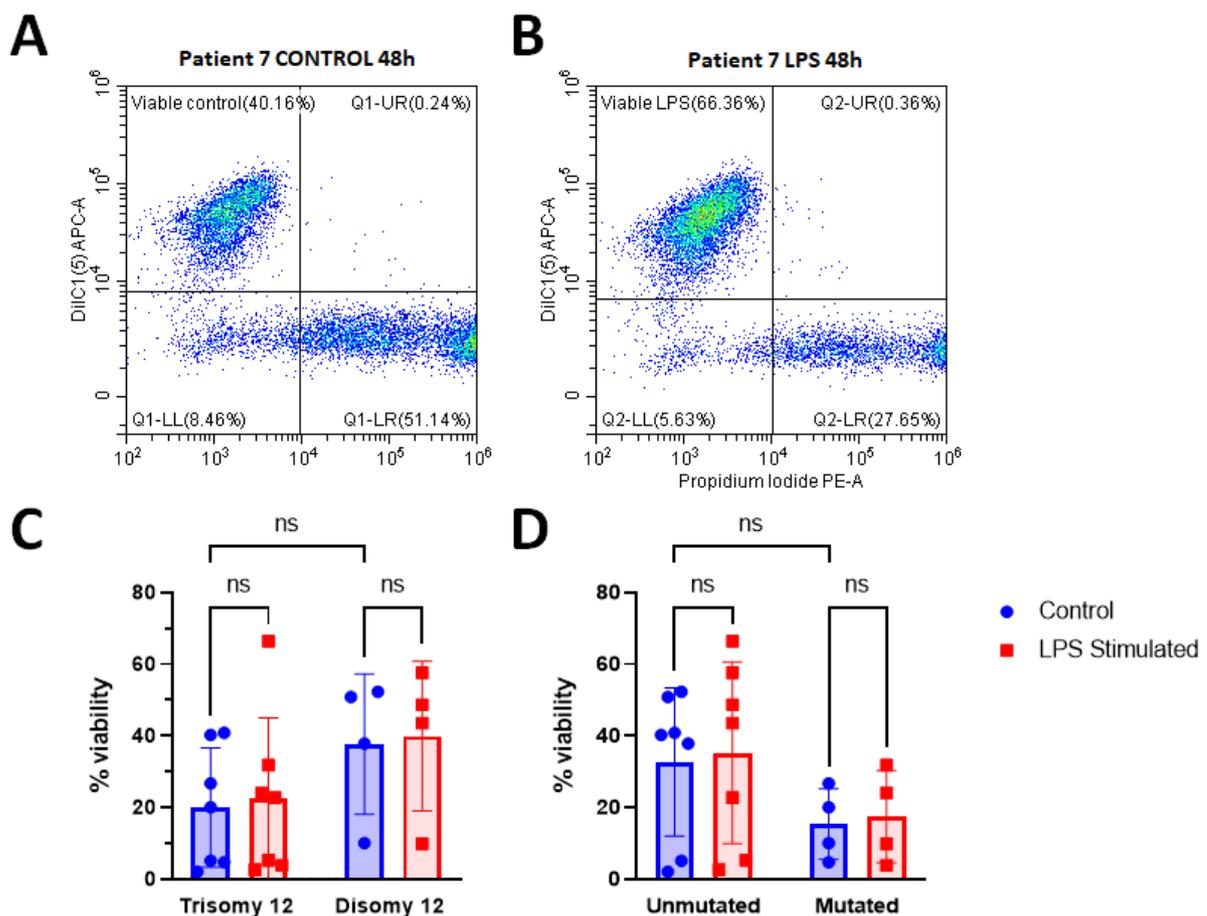


Figure 6-10. Change in TLR4, CD49d and CD14 protein expression after 48-hours of LPS-stimulation. Individual samples are dots. Means±sd are represented as columns with error bars. Data are normalised so the mean of the first control group =1.0. ns = not significant. MFI = mean fluorescence intensity. Control samples are in blue and LPS-stimulated samples in red. (A) TLR4 expression in trisomy 12 and disomy 12 CLL. (B) TLR4 expression in IGHV-unmutated and IGHV-mutated CLL. (C) Example histogram of TLR4 expression in patient 73 – red is control, green is LPS-stimulated. (D) CD49d expression in trisomy 12 and disomy 12 CLL. (E) CD14 expression in trisomy 12 and disomy 12 CLL.

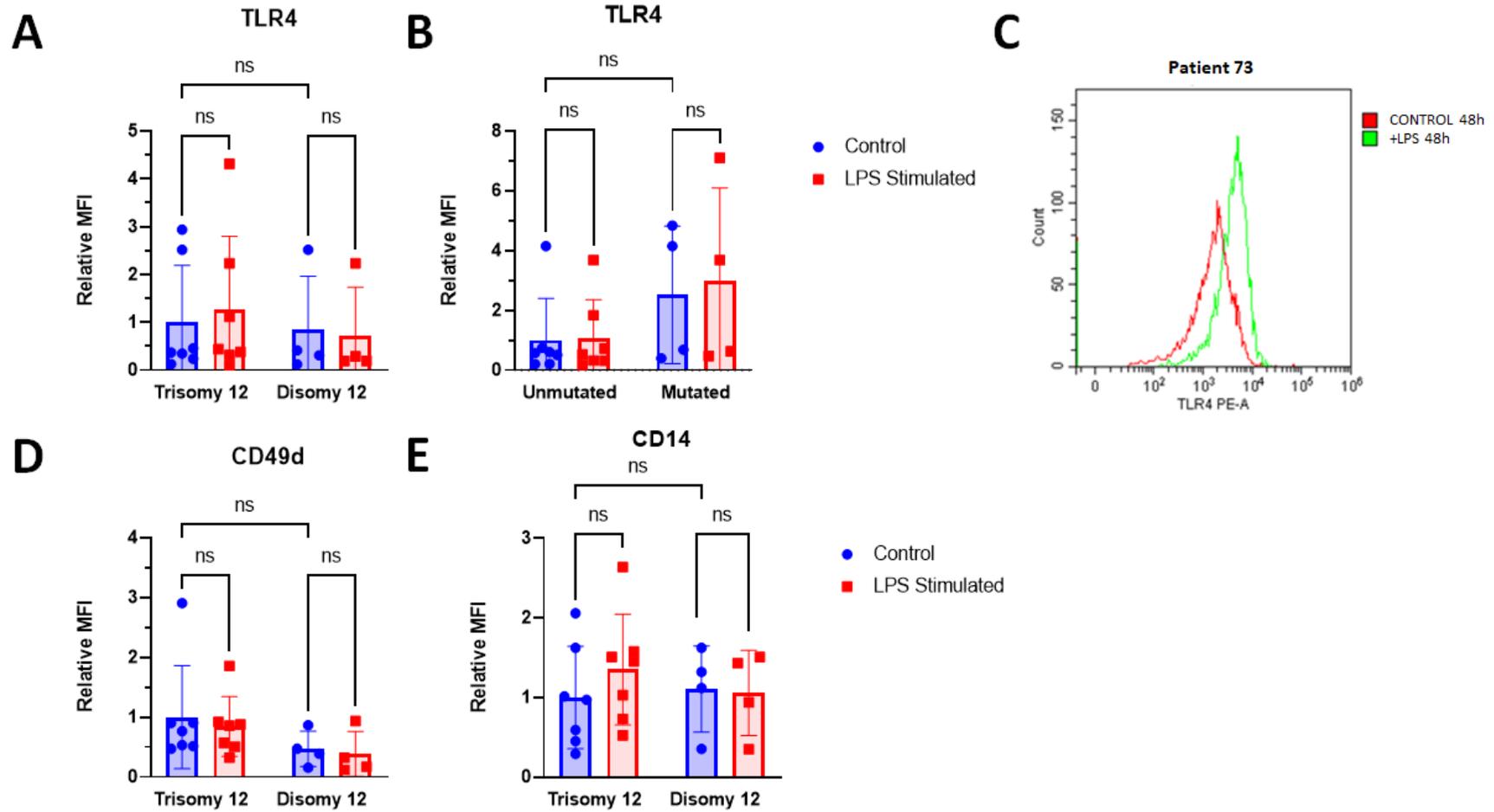
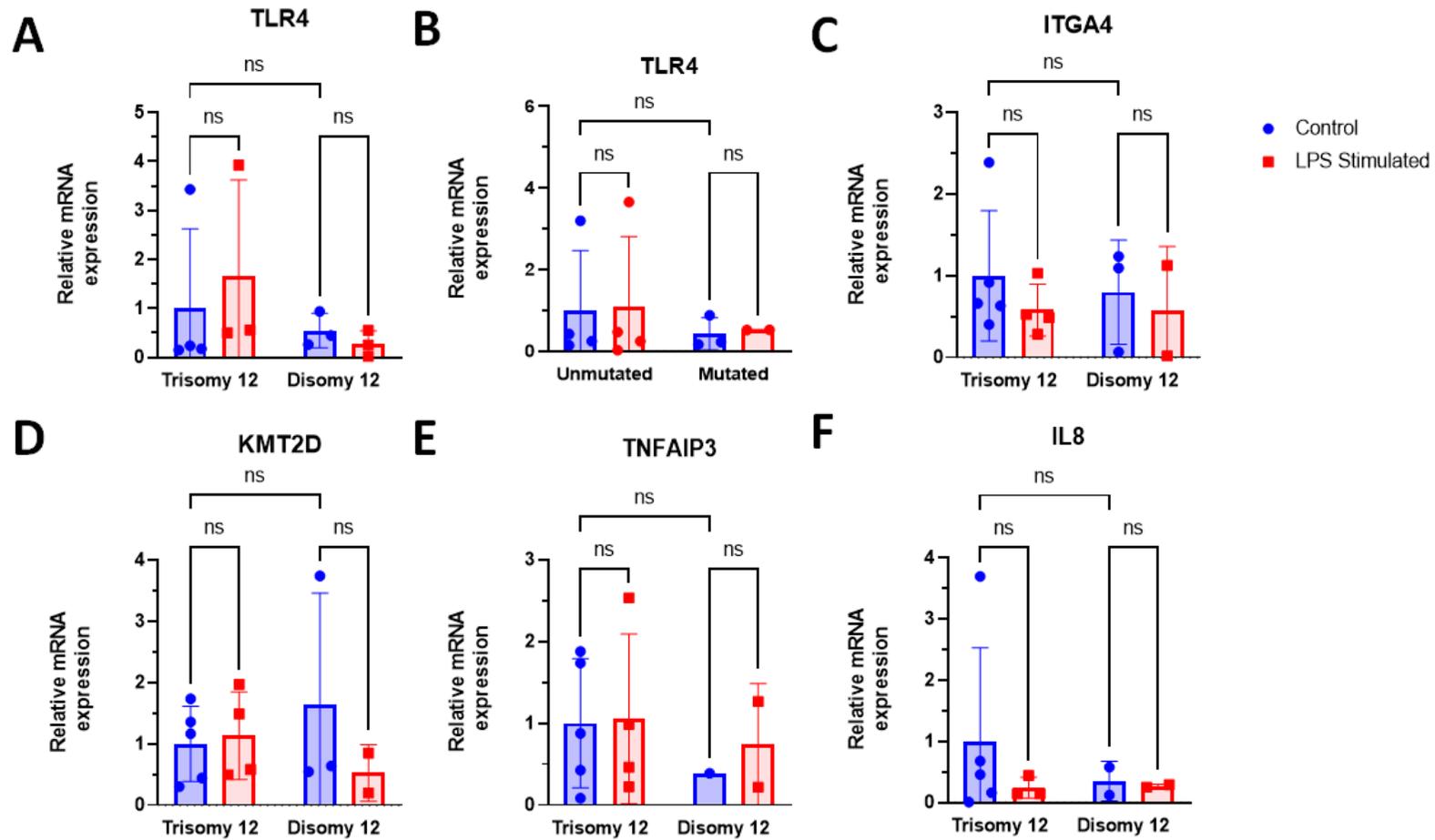


Figure 6-11. Change in TLR4, ITGA4, KMT2D, TNFAIP3 and IL8 mRNA expression after 48-hours of LPS stimulation. Individual samples are dots. Means±sd are represented as columns with error bars. Data are normalised so the mean of the first control group =1.0. ns = not significant. Control samples are in blue and LPS-stimulated samples in red. (A) TLR4 expression in trisomy 12 and disomy 12 CLL. (B) TLR4 expression in IGHV-unmutated and IGHV-mutated CLL. (C) ITGA4 expression in trisomy 12 and disomy 12 CLL. (D) KMT2D expression in trisomy 12 and disomy 12 CLL. (E) TNFAIP3 expression in trisomy 12 and disomy 12 CLL. (F) IL8 expression in trisomy 12 and disomy 12 CLL.



6.4 DISCUSSION

Toll-like receptor signalling is an important component of the innate immune system and has previously been implicated in the pathogenesis of CLL [240-242, 246]. The competency of the toll system varies widely in CLL, highlighting again the intrinsic biological heterogeneity of the disease, and may depend on the antigen experience and BCR structure of the CLL clone [241]. A relationship between toll-like signalling and genomic aberrations in CLL has not been investigated, and it was hypothesised that trisomy 12 CLL may have an enhanced dependence on TLR signalling based on the RNAseq pathway analysis from patient 6. The hypothesis is supported by reports in the literature of the strong microenvironmental dependence of trisomy 12 CLL: over-expression of CD49d involved in CLL cell homing to the microenvironmental niche and “nodal” clinical phenotype. Certainly, encounters with pathogenic antigens recognised by TLR may occur in the microenvironment in the first instance. Also, the finding of the *KMT2D* mutation in the trisomy 12 clone alone in patient 6, and the known regulatory function of *KMT2D* in TLR signalling, both support the potential relationship between trisomy 12 and the toll-like pathway. As such the hypothesis was tested in a broader cohort of CLL patients.

There was a significant difference in TLR4 protein (but not mRNA) and CD14 expression between the trisomy 12 and disomy 12 samples at baseline. Interestingly, the expression of TLR4 in the trisomy 12 group was similar to the three healthy control samples utilised. This suggests that trisomy 12 CLL cells are more closely related to normal B lymphocytes than disomic 12 CLL cells, which is supported by the unique epigenome of trisomy 12 cases, in which the chromatin signature approximates that of normal B cells [230]. The implication is that disomy 12 CLL may undergo a leukaemic reprogramming that downregulates TLR4 and CD14, rather than trisomy 12 actively upregulating these receptors. These findings could not, however, be reproduced in the stimulation assays after 48-hours in the control groups. The sample numbers were larger at baseline (0h) with 17 trisomy 12 samples and 11 disomy 12 samples compared to 7 and 5 cases (48h), which may account for the lack of statistical significance in the baseline group. Alternatively, it is possible TLR4 expression was modulated during 48-hours of cell culture, even without the addition of LPS. Whilst this data was not collected on the exact same samples at time point zero (to maximise cell numbers during the stimulation assay), TLR4 expression on different samples of the same patients was measured at baseline (see 6.3.2). Five of the 10 patients had a similar expression (MFI within 20%) of TLR4 at baseline and after 48-hours in culture, whilst the other 5 samples had varying expression – either up- or down-regulation of TLR4. There were no apparent biologic or clinical features that distinguished the two groups.

No upregulation of TLR4 at the mRNA or protein level was observed after 48-hours of stimulation with LPS. It is possible that TLR4 itself was stimulated and formed homodimers capable of activating other downstream members of the pathway without upregulating itself, but these critical pathway members were not assayed. Indeed, no difference in cell viability, nor difference in expression of *IL8*, *KMT2D* and its target *TNFAIP3*, were observed after stimulation. As in other chapters, the TLR stimulation assays were hampered by small sample numbers of variable quality available in the tissue-bank and low starting cell viability. This may be a result of sample processing and storage; however, intrinsic biological differences may also account for this: for example, some clones may be more sensitive to the freeze-thaw process than others. Nonetheless, because of low cell numbers during the assay, there were not enough cells to perform other end-point assays such as Western blots (for TNFAIP3 or NF- κ B intermediaries), proteomics or phospho-flow cytometry following cell fixation. It would be of interest to look at the end-result of TLR activation such as NF- κ B nuclear translocation and cell proliferation over a longer time course. Furthermore, if TLR signalling were only important in *KMT2D*-mutated or one stereotyped subset of CLL, for example, it would be unlikely that significant results would be observed in a sample size of 11 patients. Other variables that distinguished the trisomy 12 CLL clone in patient 6 may also account for the results from this patient sample, such as CD49d expression or the hypermutated IGHV. Large sample numbers may be required to dissect the effects of these different components. Alternatively, it is possible the case is unique and may not be extrapolated to the general CLL patient population. Indeed, despite the significant difference in TLR4 expression between the two clones in patient 6, the absolute expression of TLR4 was relatively low at baseline.

There was also a discrepancy in the expression of TLR4 at the mRNA and protein level, and no positive correlation was observed. This was also the case in other TLR members in the publication by Arvaniti et al. [241] in which “significant discrepancies” were observed between mRNA and protein expression. There are several possible explanations including post-translational regulatory mechanisms, the clonal architecture of CLL (with different subclones contributing differently), and technical reasons including contaminating monocytes in the RNA extractions (see later). Also, the product manual for the TLR4-PE antibody (ThermoFisher™) used recommends staining on lysed whole blood and that the Ficoll-gradient technique for PBMC isolation diminishes the antibody signal. This was not practical in this study due to the limited number of fresh peripheral blood samples available with which to prospectively analyse TLR4 expression. In addition to this, whilst the antibody recognised TLR4 alone, it did not recognise “active” dimerised forms necessarily and did not measure the whole complex in association with CD14 and the accessory protein MD-2. MD-2

binds to TLR4 in the cell cytoplasm and is required for trafficking of TLR4 to the cell surface and binding of the LPS ligand [5].

Whilst the TLR4 pathway was the focus of this research, it would be of interest to investigate other members of the TLR family. Certainly, the endosomal TLR8 was also upregulated in the trisomy 12 clone of patient 6 but was not investigated in the first instance due to its intra-cellular location. TLR9 is also highly expressed in CLL and is stimulated by unmethylated CpG oligonucleotides, which are used to stimulate CLL metaphases in the laboratory for chromosomal banding analysis. It has been recognised that stimulation via this pathway does not produce metaphases in all CLL samples [247], but the reason for this has not been explored and could potentially relate to the underlying genomic aberration. It would also be of interest to further analyse the TLR pathway (including TLR4 and CD14) in lymph node and bone marrow samples as well as the peripheral blood. It is possible that TLR4 and CD14 expression differs between the blood and microenvironmental compartments, however, this could not be investigated in this cohort due the lack of concurrent lymph node or bone marrow samples.

LPS itself varies between bacterial species with differing lengths and numbers of fatty acid chains attached to a polysaccharide backbone [5]. Different LPS serotypes from *Escherichia coli* have different potencies and stimulate lung cells *in vitro* to produce *IL8* to different extents [248]. A more potent serotype LPS may have yielded different results. Further optimisation experiments with different doses and potency LPS were not pursued. It should also be mentioned that the samples were not flow-sorted prior to the stimulation assays and there were a variable number of contaminating monocytes in the PBMC preparations. Monocytes express high levels of TLR4 and whilst they were excluded in the cell viability and immunophenotyping experiments, the RNA extraction would have included RNA from monocytes as well as the CLL cell population. This may have impacted upon the gene expression analyses. It would also be of interest to determine if there was any relationship between response to LPS *in vitro* and any clinical history of gram-negative sepsis and the severity of infection including incidence of septic shock. The sample size precluded this analysis and full clinical histories were not available for all the samples used in the stimulation assay. It is possible that antigen exposure primed the trisomy 12 clone in patient 6, upregulating TLR4 leading to increased *IL8* production, however, it is unclear whether this would have occurred before or after the development of outright CLL.

IL8 is not an ideal target for monitoring activation of the TLR4 pathway as it is non-specific and other pathways can lead to its expression. It was chosen as it was so highly expressed in the trisomy 12 clone of the index case. Certainly, there is a lot of crosstalk between many pathways

involved in the pathogenesis of CLL; stimulation of CLL cells via the BCR or CD40 was not performed in these experiments, but both also lead to activation of the NF- κ B pathway. The intersection of the NOTCH pathway with toll signalling is unclear in CLL, and the role of KMT2D, which has also been implicated in toll-like signalling, is also unclear. The mutational status of both genes was not known for all 11 patients; however, all 7 trisomy 12 patients were wildtype for *NOTCH1*. A more detailed genetic analysis of the samples studied may help dissect the effects of TLR4 signalling but a larger sample size is required to increase the power of the study.

In conclusion, it is not possible to determine whether TLR4 signalling is important in trisomy 12 CLL pathogenesis from the data presented. However, the relative “upregulation” in the pathway observed in the trisomy 12 CD49d+ clone appears to be reflective of a more normal activated B cell. It may be that the disomy 12 CD49d- clone in patient 6 is further removed from a healthy B cell counterpart than the trisomy 12 clone and that downregulation of the toll-like signalling occurs during leukaemogenesis of the disomy 12 clone. So, the clone with the more aggressive biological markers (*SF3B1* mutation, unmutated IGHV) downregulates expression of CD49d (normally observed in healthy B lymphocytes) and TLR4 and responds to different microenvironmental signals. This is supported by the data which suggest that the reference epigenome suggests of trisomy 12 CLL is more closely aligned with healthy B cells [208]. The data presented could not fully support or refute the hypothesised importance of TLR signalling in trisomy 12 CLL but suggests that further investigation in a larger cohort of patients is warranted.

7 CONCLUSIONS AND FUTURE DIRECTIONS

This thesis aimed to investigate the molecular pathogenesis of trisomy 12 CLL and explore its association with high CD49d expression. Initially, the local cohort of trisomy 12 were characterised, confirming higher CD49d expression in trisomy 12 CLL compared to disomy 12 CLL comparators. The opportunity provided by the unusual biclonal case was pursued in the remainder of the thesis and led to the identification of a novel putative driver of trisomy 12 CLL (*KMT2D*) and the potential dependence of trisomy 12 CLL on toll-like receptor signalling. This chapter will focus on key points arising from this thesis in the investigation of trisomy 12 CLL pathogenesis, in particular the association between trisomy 12 and CD49d expression.

7.1 CD49D IN CLL

CD49d, the alpha unit of the integrin heterodimer VLA-4, is a negative prognosticator in CLL with worse clinical outcomes are observed if the CLL malignant cell fraction expresses CD49d in greater than 30% of cells (with a hazard ratio for death of 1.88) [162]. This is somewhat surprising considering that CD49d is expressed at high levels in normal B lymphocytes and suggests that CD49d is actively downregulated in some CLL cases, rather than upregulated in poor-risk CLL. This begs the question as to what role maintenance of CD49d expression has in the development of poor prognosis and what the underlying mechanism is for downregulation in CD49d-negative CLL.

The mechanism of downregulation is not clear but does not appear to be solely due to hypermethylation of the *ITGA4* promoter as previously suggested [155]. In patient 21, there was no methylation of *ITGA4* in the CD49d- clone, suggesting an alternative mechanism of downregulation. The literature fails to provide an adequate explanation of the mechanism. The paper by Zucchetto *et al.* did demonstrate higher but variable methylation patterns in CD49d- cases but without a strong correlation with CD49d expression ($r^2 = 0.6$).

Furthermore, CD49d may not be a true driver of leukaemia but account for a malignant phenotype that becomes symptomatic earlier and is less responsive to chemoimmunotherapy. When expressed, CD49d may alter disease biology and account for some of the phenotypic changes observed in trisomy 12 CLL. For example, given its role in CLL cell homing, it is plausible that this is the reason trisomy 12 CLL more often present with a nodal phenotype, which has a shorter time to first treatment and may then harbour residual disease in the nodal compartments post-treatment resulting in a higher frequency of relapse [249]. This is also true of treatment with the novel agent

ibrutinib, in which CD49d expression identifies cases with an inferior nodal response and shortened progression-free survival [250].

This thesis demonstrated that CD49d expression did not necessarily co-exist with other poor prognostic markers, in line with its independent importance as a poor-prognostic marker [162]. In patient 6, the clone that had high expression of CD49d harboured the good-risk factor of a hypermutated IGHV, and conversely the clone with low expression of CD49d harboured the poor-risk factors of deletion 17p, an unmutated IGHV and an *SF3B1* mutation. In addition to this, CD49d may not be actively upregulated in poor-risk CLL and may be a “passenger” marker reflecting the degree of separation between the CLL tumour and normal B cells. This certainly fits with the epigenetic data showing that trisomy 12 CLL cells (that more often express CD49d at high levels) more closely resemble normal B cells [230]. This also fits with the data presented in this thesis showing that trisomy 12 CLL expresses TLR4 at similar levels to normal B cells as opposed to disomy 12 CLL and does not downregulate CD49d to the same extent as disomy 12 CLL.

CD49d also appears to drive a more nodal phenotype. Certainly, the lymph node microenvironment is where CLL cells are more likely to encounter antigens that stimulate toll-signalling and as such, it may be advantageous for trisomy 12 CLL to maintain CD49d and TLR4 expression and home to lymph nodes where it responds to a set of unknown potentially endogenous antigens. Trisomy 12 CLL is enriched for subset #8 (5% of cases) and conversely 70% of this subset have trisomy 12 [73]. This subset is known to have an aggressive course and possesses excessive antigen reactivity [251]. However, this still does not answer the question as to which factors on chromosome 12 are critical to drive leukaemia but paints a picture of a nodal phenotype which may be endogenously driven.

Despite the prognostic impact of CD49d, the inhibition of VLA-4 is unlikely to be a viable therapeutic strategy in CLL given its high expression in normal B lymphocytes (more so than non-trisomy 12 CLL B cells). Furthermore, the use of the anti-VLA4 monoclonal antibody natalizumab in multiple sclerosis is associated with significant adverse effects such as progressive multifocal leukoencephalopathy [252].

7.2 CLL HETEROGENEITY AND SAMPLING ISSUES

CLL is a heterogenous disease, exemplified by its complex clonal architecture, number of genomic alterations that occur within leukaemic cells and absence of a single unifying chromosomal aberration (in contrast to chronic myeloid leukaemia). This thesis highlighted the inherent heterogeneity of CLL, especially within the trisomy 12 subset, and the difficulty in analysing data

from small cohorts of patients. This is one of the foremost reasons that the two individual patients (patients 6 and 21) were analysed in such detail with the aim to eliminate confounding effects from background genetic variability and to gain an understanding of clonal evolution.

The variability of CLL was highlighted in the toll-like signalling experiments in which there was a wide range of expression of all markers at baseline and following stimulation with LPS, rendering it difficult to draw conclusions from the relatively small cohort. In addition to this, variability of sample quality also contributed to the lack of uniformity in results within similar subgroups. Whilst some of the variability was due to the method of sample storage, it is also possible that the inherent heterogeneity of the disease and intrinsic biological differences of subclones within a single sample affected the survival of cells during storage and thawing. It is also possible that certain clones that demonstrated very low viability post-thaw are more reliant on pro-survival microenvironmental signals encountered *in vivo* that were not replicated in the culture system. These sampling issues and biological heterogeneity underlie one of the fundamental difficulties of CLL research.

A more uniform system, such as cell-line work, would obviate the variability observed with primary sample research, however, the generalisability to the CLL disease state would not be assured. There is a CLL-like, EBV-transformed, patient-derived cell line with trisomy 12 (OSU-CLL) [253], however, it also contains trisomy 19 and has not been widely published.

7.3 ASSAY TECHNOLOGIES

A recurrent theme in this thesis has been the differing sensitivities, limitations, and applicability of different technologies in CLL research. Cytogenetic analysis of CLL prior to therapy (first-line or subsequent) is of critical importance in the clinic. It is recommended by the iwCLL and alters the treatment algorithm if del(17p) is present. Interphase FISH incorporating probes to the 17p13.1, 11q22.3, 12q and 13q14.2-q14.3 chromosomal regions is the backbone of CLL risk stratification. The technique is sensitive and can detect low-frequency subclones, however, is operator dependent and will not identify areas of uniparental disomy (copy-neutral loss of heterozygosity), which is important in *TP53*-mutated disease. Despite the sensitivity, FISH performed on patient 6 at diagnosis did not detect the del(17p) subclone (which required enrichment of the CD49d- clone with flow sorting) and did not recognise the copy-neutral loss of heterozygosity at 11q seen in patient 21. Whilst FISH assesses specific genomic regions alone, array-based comparative genomic hybridisation (array CGH) or single-nucleotide polymorphism (SNP) microarray allows for a whole-genome overview with sensitive resolution of small DNA copy number changes representing gains and losses of genetic material. It also allows for detection of loss of heterozygosity which is exemplified in the case of

patient 21: a loss of heterozygosity at 11q was found in parallel with a homozygous *BIRC3* mutation on 11q. In addition to this, additional genetic abnormalities were detected in patients that underwent SNP array as opposed to FISH at diagnosis including additional trisomies and del(14q). Both of these additional chromosomal abnormalities have been demonstrated to alter prognosis in trisomy 12 CLL and these data are lost in the targeted FISH analysis.

Chromosomal banding analysis (CBA) has been improved in CLL with new protocols for stimulating CLL cells into cycle that, interestingly, utilise toll agonists. CBA also allows for a whole-genome overview like SNP microarray; however, the resolution is much lower and small copy-number changes will not be detected. In contrast, it will detect balanced chromosomal translocations which are not readily detected by SNP microarray. This is of lesser importance in CLL in which such translocations are rare events. CBA was infrequently used in the cohort presented, however, complex karyotype, defined as ≥ 3 chromosomal alterations using CBA, has a negative prognostic significance and is predictive of treatment refractoriness. Even poorer outcomes are observed in patients with ≥ 5 abnormalities, however, this was again defined using CBA technology. It is unclear whether those patients with ≥ 3 chromosomal alterations using SNP microarray have the same prognosis, especially given the greater sensitivity of the technique to detect any changes from a normal diploid karyotype. The outcome of the patients with “complex karyotype” based on SNP microarray in this thesis was not investigated given small overall numbers limiting the power of this analysis.

Whole-exome sequencing is a powerful tool for detecting exome variants (even at low variant allele frequencies of 1%), enabling detection of clonal and subclonal mutations of known or potential clinical significance. It also allows for inferences to be made on clonal architecture and evolution, as demonstrated in patients 6 and 21 (see Figure 4-11 and Figure 4-12). However, it did not detect the subclonal deletion 17p in the CD49d- clone of patient 6 that was observed in the SNP microarray. This is of clinical significance as patients with aberrations of 17p are recommended a different upfront treatment, avoiding chemotherapy which can lead to chromosomal instability and early relapse or refractory disease. This is a key limitation of WES at present due to inherent bioinformatic difficulties in copy-number variant analysis of the sequencing data, especially in cancer which is a mosaic disease of multiple clones and subclones. WES remains a tool useful in clinical research but not in the day-to-day treatment and prognostication of CLL, especially given its expense and need for technical expertise in data analysis.

A next-generation sequencing method was utilised to identify the dominant IGHV rearrangement and mutational status in this thesis rather than the traditional Sanger sequencing

method, still in use in some diagnostic facilities. The NGS method employs a single optimised multiplex PCR master mix and preferentially amplifies the initial dominant clonal lymphoid population as well as identifying the DNA sequence of the rearranged IGHV gene contemporaneously. It also allows for visualisation of other clones present within the sample and their relative read frequencies, and data on the V and J segment utilisation. The main advantages over the standard Sanger sequencing method include the preferential amplification of the dominant clone, the obtainment of sequencing and frequency data at the same time, and the detection of subclones. Whilst the identification of multiple subclones is currently more academic than of practical use, the clear identification of a dominant clone and accurate quantification of the mutation frequency is valuable information. The subclone frequency and gene usage is additional information at no extra cost and improves the quality of the information obtained in the technical process. The sequencing information allows for the disease to be followed in time and for the detection of minimal residual disease (MRD). Whilst scrutiny of subclones and MRD detection using IGHV was not the focus of this work, the data generated is of value and could be used in future research projects.

7.4 PROGNOSTICATION IN CLL

There are numerous prognostic markers to aid in risk stratification of patients with CLL. As discussed above, the specific assay utilised will impact on the detection of certain markers and in practice it is recommended to use a combination of technologies to assay for complex karyotype, *TP53*-defective disease (either via mutational or deletional events), and the four most common genomic aberrations. Despite their prognostic relevance, detection of most prognostic markers does not currently alter the treatment algorithm (excepting *del(17p)/TP53* aberrations), however, provides a platform for their validation as potentially predictive markers to therapy in ongoing clinical studies. This thesis highlighted difficulties in prognostication in traditionally “low” to “intermediate” risk groups and the heterogeneity of trisomy 12 CLL. Even within the intermediate group prognosis varies widely. Certainly, patient 6 would have been described as intermediate risk at diagnosis with trisomy 12 without *del(11q)* or *del(17p)* on FISH. However, more detailed immunogenetic analysis revealed poor-risk features of clinical significance including the presence of a subclonal *del(17p)*, subclonal *SF3B1* mutation and expression of CD49d. This supports the growing evidence that extended genomic analysis of low- or intermediate-risk patients to detect important subclones may be of clinical utility and enhances our understanding of the biological underpinnings of CLL. The addition of CD49d to the diagnostic CLL flow cytometry panel would be a simple undertaking. Targeted next-generation sequencing and flow sorting of clones would still be reserved

for centres with the technical expertise now but would be of importance in risk stratification of patients and analysis of subgroup outcomes in prospective clinical trials.

7.5 GERMLINE PREDISPOSITION TO CLL

There is a familial predisposition to CLL [254], and genome-wide association studies have detected several susceptibility loci [255, 256], however, the heritable risk has not been fully explained. The presence of a germline *TET2* variant in patient 6 has not been described in CLL and its contribution to the development of two unique CLL clones in this patient is not clear. Acquisition of *TET2* variants in the haematopoietic compartment with ageing has been well described but is more closely associated with the development of myeloid malignancies such as acute myeloid leukaemia [257, 258]. Germline *TET2* haploinsufficiency has recently been described in a family with three cases of lymphoma, leading to global genome hypermethylation especially at active enhancer regions, so it seems plausible that the germline mutation in this patient might have predisposed him to the development of CLL.

It would be of interest to investigate for the presence of germline mutations that are typically associated with acquired clonal haematopoiesis (such as *TET2*) in patients with CLL, especially those that present with the disease at a young age or have a strong family history of lymphoproliferative disorders.

7.6 CLONAL EVOLUTION IN CLL

Trisomy 12 has been proposed as one of two founding lesions in CLL alongside del(13q). This thesis has neither confirmed nor refuted this, however, the presence of the somatic *MDC1* mutation in both leukaemic clones in patient 6 suggested the presence of a common progenitor cell prior to IGHV rearrangement and following B-lineage commitment. Although this is only a single case, it adds to the field of evidence that the debated cell of origin of CLL is not necessarily a fully differentiated B- cell and it may arise earlier in haematopoiesis [202].

Trisomy 12 was present at high clonal frequency (>60%) in the majority of cases. Whilst this may support it is a founding lesion, alternatively it could also mean that it is of such marked survival advantage after the development of CLL in certain conditions (that is, with a particular antigenic drive) that duplication of chromosome 12 is acquired in most of the malignant cells.

Its clonal frequency was not monitored over time. It would be of interest to assay for trisomy 12 in pre-malignant blood samples of individuals who go on to develop trisomy 12 CLL. A recent study of autosomal mosaic chromosomal aberrations in blood-derived DNA from large

population-based cohorts in Japan and the United Kingdom demonstrated that trisomy 12 was between 2- and 6-fold more common in the UK population than in Japan [259]. CLL is more common in the UK than Japan, and this study demonstrated that recurrent CLL genomic changes (such as trisomy 12 and del(13q)) were more common in haematopoietic clones in UK individuals prior to the development of overt CLL. It is thus proposed that different pressures select for trisomy 12 in the UK cohort prior to the development of leukaemia, however, the nature of these pressures is unclear. Exposure to different environmental antigens is an attractive hypothesis, however, equally there are intrinsic genetic differences between the populations, such as different human leukocyte antigen (HLA) haplotypes [260]. Another attractive hypothesis is a difference in TLR variants between the populations, however, this has not been extensively studied. For example, a rare toll-like receptor 6 (TLR6) variant (p.Val427Ala; SNP rs5743815) appears to be associated with susceptibility to CLL [261] and has a 75x higher incidence in Europeans compared to East Asians on the dbSNP database [262].

7.7 PATHOGENESIS OF TRISOMY 12 CLL

The pathogenesis of trisomy 12 CLL remains obscure. All samples presented in this thesis had duplication of a complete chromosome 12 and as such there was no “minimally triplicated” region in which to concentrate the investigation. Critical factors or pathways altered due to acquisition of chromosome 12 were not identified in the cohort, however, differential expression of members of toll-like signalling was detected between the disomy 12 and trisomy 12 clone in patient 6. A dependence on toll-signalling could not be confirmed in the wider trisomy 12 cohort, however, expression of TLR4 did approximate normal B cells in trisomy 12 CLL. Altogether, it is possible that trisomy 12 CLL arises from mature B cells that express CD49d and TLR4 and are exposed to different unknown antigenic selective pressures from their disomic 12 counterparts, which actively downregulate CD49d and TLR4. Investigating the mechanism by which this occurs and further interrogating the toll-like system in CLL would be valuable.

A novel nonsense mutation in the epigenetic regulator *KMT2D* on chromosome 12 was discovered in a single trisomy 12 case, however, overexpression of *KMT2D* in the wider trisomy 12 cohort was not demonstrated. Investigation of *KMT2D* mutations in the wider cohort, and epigenetic and functional studies were not performed due to practical constraints. Targeted sequencing of the gene is another avenue of future research to determine if *KMT2D* mutations indeed play a role in the pathogenesis of trisomy 12 CLL and would be an essential first step.

In conclusion, the aetiology and clonal evolution of trisomy 12 CLL remains unclear and efforts to dissect underlying pathogenic mechanisms are hampered by its inherent biological heterogeneity and methodological limitations, in particular sampling. This thesis presents a

comprehensive analysis of a trisomy 12 CLL cohort, an in-depth investigation of a unique biclonal case of CLL, data to challenge methylation-dependent regulation of *ITGA4*, data to support the closer alignment of trisomy 12 CLL with normal B lymphocytes in comparison to disomy 12 CLL, and implicates toll-like signalling and the epigenetic regulator KMT2D in the pathogenesis of trisomy 12 CLL.

8 REFERENCES

1. Australian Cancer Research Foundation Available from: <https://acrf.com.au/>.
2. Swerdlow, S.H., et al., eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 2008, International Agency for Research on cancer: Lyon.
3. Cerhan, J.R. and S.L. Slager, *Familial predisposition and genetic risk factors for lymphoma*. *Blood*, 2015. **126**(20): p. 2265-73.
4. LeBien, T.W. and T.F. Tedder, *B lymphocytes: how they develop and function*. *Blood*, 2008. **112**(5): p. 1570-1580.
5. Murphy, K. and C. Weaver, *Janeway's Immunobiology*. 9th ed. 2017, New York, NY: Garland Science.
6. Schroeder, H.W., Jr. and L. Cavacini, *Structure and function of immunoglobulins*. *J Allergy Clin Immunol*, 2010. **125**(2 Suppl 2): p. S41-52.
7. Perez-Andres, M., et al., *Human peripheral blood B-cell compartments: A crossroad in B-cell traffic*. *Cytometry Part B: Clinical Cytometry*, 2010. **78B**(S1): p. S47-S60.
8. Morbach, H., et al., *Reference values for B cell subpopulations from infancy to adulthood*. *Clin Exp Immunol*, 2010. **162**(2): p. 271-9.
9. Bain, B., *Blood Cells: A Practical Guide*. 5th edition ed. 2015: Wiley-Blackwell.
10. Chigrinova, E., et al., *Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome*. *Blood*, 2013. **122**(15): p. 2673-82.
11. Hallek, M., et al., *Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines*. *Blood*, 2008. **111**(12): p. 5446-56.
12. Almasri, N.M., et al., *Reduced expression of CD20 antigen as a characteristic marker for chronic lymphocytic leukemia*. *Am J Hematol*, 1992. **40**(4): p. 259-63.
13. Binet, J.L., et al., *A clinical staging system for chronic lymphocytic leukemia: prognostic significance*. *Cancer*, 1977. **40**(2): p. 855-64.
14. Rai, K., *A critical analysis of staging in CLL*, in *Chronic Lymphocytic Leukaemia: Recent Progress and Future Directions*, R.P. Gale and K. Rai, Editors. 1987, Liss: New York, NY. p. 253-264.
15. Marti, G.E., et al., *Diagnostic criteria for monoclonal B-cell lymphocytosis*. *Br J Haematol*, 2005. **130**(3): p. 325-32.
16. Rawstron, A.C., et al., *Different biology and clinical outcome according to the absolute numbers of clonal B-cells in monoclonal B-cell lymphocytosis (MBL)*. *Cytometry Part B: Clinical Cytometry*, 2010. **78B**(S1): p. S19-S23.
17. Rawstron, A.C., et al., *Monoclonal B-Cell Lymphocytosis and Chronic Lymphocytic Leukemia*. *New England Journal of Medicine*, 2008. **359**(6): p. 575-583.
18. Strati, P. and T.D. Shanafelt, *Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification*. *Blood*, 2015. **126**(4): p. 454-62.
19. Richter, M.N., *Generalized Reticular Cell Sarcoma of Lymph Nodes Associated with Lymphatic Leukemia*. *Am J Pathol*, 1928. **4**(4): p. 285-292 7.
20. Vitale, C. and A. Ferrajoli, *Richter Syndrome in Chronic Lymphocytic Leukemia*. *Curr Hematol Malig Rep*, 2016. **11**(1): p. 43-51.
21. Mescher, C., et al., *The impact of Agent Orange exposure on prognosis and management in patients with chronic lymphocytic leukemia: a National Veteran Affairs Tumor Registry Study*. *Leukemia & Lymphoma*, 2018. **59**(6): p. 1348-1355.
22. Blair, A., et al., *Chemical exposures and risk of chronic lymphocytic leukaemia*. *British Journal of Haematology*, 2007. **139**(5): p. 753-761.

23. Berndt, S.I., et al., *Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia*. Nat Genet, 2013. **45**(8): p. 868-76.
24. Crowther-Swanepoel, D., et al., *Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk*. Nat Genet, 2010. **42**(2): p. 132-6.
25. Di Bernardo, M.C., et al., *A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia*. Nat Genet, 2008. **40**(10): p. 1204-10.
26. Slager, S.L., et al., *Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia*. Blood, 2012. **120**(4): p. 843-6.
27. Speedy, H.E., et al., *A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia*. Nat Genet, 2014. **46**(1): p. 56-60.
28. Goldin, L.R., et al., *Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia*. Haematologica, 2009. **94**(5): p. 647-53.
29. Fabbri, G. and R. Dalla-Favera, *The molecular pathogenesis of chronic lymphocytic leukaemia*. Nat Rev Cancer, 2016. **16**(3): p. 145-62.
30. Kipps, T.J., et al., *Chronic lymphocytic leukaemia*. Nat Rev Dis Primers, 2017. **3**: p. 17008.
31. Spaargaren, M., et al., *BTK inhibitors in chronic lymphocytic leukemia: a glimpse to the future*. Oncogene, 2015. **34**(19): p. 2426-36.
32. Murphy, K.M. and C. Weaver, *Janeway's Immunobiology*. 9th ed. 2017: Garland Science.
33. Kipps, T.J., et al., *Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia*. Proc Natl Acad Sci U S A, 1989. **86**(15): p. 5913-7.
34. Vardi, A., et al., *Immunogenetic studies of chronic lymphocytic leukemia: revelations and speculations about ontogeny and clinical evolution*. Cancer Res, 2014. **74**(16): p. 4211-6.
35. Widhopf, G.F., 2nd, et al., *Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins*. Blood, 2004. **104**(8): p. 2499-504.
36. Baliakas, P., et al., *Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study*. Lancet Haematol, 2014. **1**(2): p. e74-84.
37. CATERA, R., et al., *Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation*. Mol Med, 2008. **14**(11-12): p. 665-74.
38. Chu, C.C., et al., *Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA*. Blood, 2008. **112**(13): p. 5122-9.
39. Chu, C.C., et al., *Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin*. Blood, 2010. **115**(19): p. 3907-15.
40. Hoogeboom, R., et al., *A mutated B cell chronic lymphocytic leukemia subset that recognizes and responds to fungi*. J Exp Med, 2013. **210**(1): p. 59-70.
41. Duhren-von Minden, M., et al., *Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling*. Nature, 2012. **489**(7415): p. 309-12.
42. Packham, G., et al., *The outcome of B-cell receptor signaling in chronic lymphocytic leukemia: proliferation or anergy*. Haematologica, 2014. **99**(7): p. 1138-48.
43. Stevenson, F.K., et al., *B-cell receptor signaling in chronic lymphocytic leukemia*. Blood, 2011. **118**(16): p. 4313-20.
44. Mockridge, C.I., et al., *Reversible anergy of slgM-mediated signaling in the two subsets of CLL defined by VH-gene mutational status*. Blood, 2007. **109**(10): p. 4424-31.
45. Kulis, M., et al., *Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia*. Nat Genet, 2012. **44**(11): p. 1236-42.
46. Klein, U., et al., *Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells*. J Exp Med, 2001. **194**(11): p. 1625-38.

47. Rosenwald, A., et al., *Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia*. J Exp Med, 2001. **194**(11): p. 1639-47.
48. Seifert, M., et al., *Cellular origin and pathophysiology of chronic lymphocytic leukemia*. J Exp Med, 2012. **209**(12): p. 2183-98.
49. Packham, G. and F. Stevenson, *The role of the B-cell receptor in the pathogenesis of chronic lymphocytic leukaemia*. Semin Cancer Biol, 2010. **20**(6): p. 391-9.
50. Ten Hacken, E. and J.A. Burger, *Microenvironment interactions and B-cell receptor signaling in Chronic Lymphocytic Leukemia: Implications for disease pathogenesis and treatment*. Biochim Biophys Acta, 2016. **1863**(3): p. 401-413.
51. Herishanu, Y., et al., *The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia*. Blood, 2011. **117**(2): p. 563-74.
52. Malavasi, F., et al., *CD38 and chronic lymphocytic leukemia: a decade later*. Blood, 2011. **118**(13): p. 3470-3478.
53. Deaglio, S., et al., *CD38 and CD100 lead a network of surface receptors relaying positive signals for B-CLL growth and survival*. Blood, 2005. **105**(8): p. 3042-50.
54. Granziero, L., et al., *CD100/Plexin-B1 interactions sustain proliferation and survival of normal and leukemic CD5+ B lymphocytes*. Blood, 2003. **101**(5): p. 1962-9.
55. Bagnara, D., et al., *A novel adoptive transfer model of chronic lymphocytic leukemia suggests a key role for T lymphocytes in the disease*. Blood, 2011. **117**(20): p. 5463-72.
56. Choi, M.Y., M.K. Kashyap, and D. Kumar, *The chronic lymphocytic leukemia microenvironment: Beyond the B-cell receptor*. Best Practice & Research Clinical Haematology, 2016. **29**(1): p. 40-53.
57. Seke Etet, P.F., L. Vecchio, and A.H. Nwabo Kamdje, *Interactions between bone marrow stromal microenvironment and B-chronic lymphocytic leukemia cells: Any role for Notch, Wnt and Hh signaling pathways?* Cellular Signalling, 2012. **24**(7): p. 1433-1443.
58. Hallek, M., et al., *Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial*. Lancet, 2010. **376**(9747): p. 1164-74.
59. Fischer, K., et al., *Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial*. Blood, 2016. **127**(2): p. 208-15.
60. Thompson, P.A., et al., *Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia*. Blood, 2016. **127**(3): p. 303-9.
61. Goede, V., et al., *Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions*. N Engl J Med, 2014. **370**(12): p. 1101-10.
62. Byrd, J.C., et al., *Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia*. N Engl J Med, 2014. **371**(3): p. 213-23.
63. O'Brien, S., et al., *Ibrutinib as initial therapy for elderly patients with chronic lymphocytic leukaemia or small lymphocytic lymphoma: an open-label, multicentre, phase 1b/2 trial*. Lancet Oncol, 2014. **15**(1): p. 48-58.
64. Farooqui, M.Z., et al., *Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: a phase 2, single-arm trial*. Lancet Oncol, 2015. **16**(2): p. 169-76.
65. Furman, R.R., et al., *Idelalisib and rituximab in relapsed chronic lymphocytic leukemia*. N Engl J Med, 2014. **370**(11): p. 997-1007.
66. Anderson, M.A., et al., *The BCL2 selective inhibitor venetoclax induces rapid onset apoptosis of CLL cells in patients via a TP53-independent mechanism*. Blood, 2016. **127**(25): p. 3215-24.
67. Roberts, A.W., et al., *Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia*. N Engl J Med, 2016. **374**(4): p. 311-22.

68. Stilgenbauer, S., et al., *Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study*. *Lancet Oncol*, 2016. **17**(6): p. 768-778.
69. Seymour, J.F., et al., *Four-Year Analysis of Murano Study Confirms Sustained Benefit of Time-Limited Venetoclax-Rituximab (VenR) in Relapsed/Refractory (R/R) Chronic Lymphocytic Leukemia (CLL)*. *Blood*, 2019. **134**(Supplement_1): p. 355-355.
70. Fischer, K., et al., *Venetoclax and Obinutuzumab in Patients with CLL and Coexisting Conditions*. *New England Journal of Medicine*, 2019. **380**(23): p. 2225-2236.
71. Damle, R.N., et al., *Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia*. *Blood*, 1999. **94**(6): p. 1840-7.
72. Hamblin, T.J., et al., *Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia*. *Blood*, 1999. **94**(6): p. 1848-54.
73. Sonia, J., et al., *Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL Study Group (GCLLSG)*. *Haematologica*, 2019. **105**(11): p. 2598-2607.
74. Rossi, D., et al., *Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome*. *Clin Cancer Res*, 2009. **15**(13): p. 4415-22.
75. Baliakas, P., et al., *Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations*. *Blood*, 2015. **125**(5): p. 856-9.
76. Hamblin, T.J., et al., *CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease*. *Blood*, 2002. **99**(3): p. 1023-9.
77. Crespo, M., et al., *ZAP-70 Expression as a Surrogate for Immunoglobulin-Variable-Region Mutations in Chronic Lymphocytic Leukemia*. *New England Journal of Medicine*, 2003. **348**(18): p. 1764-1775.
78. Orchard, J.A., et al., *ZAP-70 expression and prognosis in chronic lymphocytic leukaemia*. *Lancet*, 2004. **363**(9403): p. 105-11.
79. Wiestner, A., et al., *ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile*. *Blood*, 2003. **101**(12): p. 4944-51.
80. Dohner, H., et al., *Genomic aberrations and survival in chronic lymphocytic leukemia*. *N Engl J Med*, 2000. **343**(26): p. 1910-6.
81. Fabbri, G., et al., *Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation*. *J Exp Med*, 2011. **208**(7): p. 1389-401.
82. Landau, D.A., et al., *Evolution and impact of subclonal mutations in chronic lymphocytic leukemia*. *Cell*, 2013. **152**(4): p. 714-26.
83. Landau, D.A., et al., *Mutations driving CLL and their evolution in progression and relapse*. *Nature*, 2015. **526**(7574): p. 525-30.
84. Puente, X.S., et al., *Non-coding recurrent mutations in chronic lymphocytic leukaemia*. *Nature*, 2015. **526**(7574): p. 519-24.
85. Quesada, V., A.J. Ramsay, and C. Lopez-Otin, *Chronic lymphocytic leukemia with SF3B1 mutation*. *N Engl J Med*, 2012. **366**(26): p. 2530.
86. Wang, L., et al., *SF3B1 and other novel cancer genes in chronic lymphocytic leukemia*. *N Engl J Med*, 2011. **365**(26): p. 2497-506.
87. Puente, X.S., et al., *Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia*. *Nature*, 2011. **475**(7354): p. 101-5.
88. Edelmann, J., et al., *High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations*. *Blood*, 2012. **120**(24): p. 4783-94.
89. Klein, U., et al., *The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia*. *Cancer Cell*, 2010. **17**(1): p. 28-40.

90. Stankovic, T. and A. Skowronska, *The role of ATM mutations and 11q deletions in disease progression in chronic lymphocytic leukemia*. *Leuk Lymphoma*, 2014. **55**(6): p. 1227-39.
91. Rossi, D., et al., *Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia*. *Blood*, 2012. **119**(12): p. 2854-62.
92. Döhner, H., et al., *p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias*. *Blood*, 1995. **85**(6): p. 1580-9.
93. Döhner, H., et al., *Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia*. *New England Journal of Medicine*, 2000. **343**(26): p. 1910-1916.
94. Hallek, M., et al., *Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial*. *The Lancet*, 2010. **376**(9747): p. 1164-1174.
95. Zenz, T., et al., *Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial*. *Blood*, 2009. **114**(13): p. 2589-97.
96. Nadeu, F., et al., *Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia*. *Leukemia*, 2018. **32**(3): p. 645-653.
97. Fary, D., et al., *Biological and clinical implications of BIRC3 mutations in chronic lymphocytic leukemia*. *Haematologica*, 2020. **105**(2): p. 448-456.
98. Nadeu, F., et al., *Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia*. *Blood*, 2016. **127**(17): p. 2122-2130.
99. Jeromin, S., et al., *SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients*. *Leukemia*, 2014. **28**(1): p. 108-117.
100. Wang, L., et al., *SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia*. *New England Journal of Medicine*, 2011. **365**(26): p. 2497-2506.
101. Quesada, V., et al., *Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia*. *Nat Genet*, 2011. **44**(1): p. 47-52.
102. Rossi, D., et al., *Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness*. *Blood*, 2011. **118**(26): p. 6904-6908.
103. Dicker, F., et al., *The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype*. *Leukemia*, 2009. **23**(1): p. 117-124.
104. Malcikova, J., et al., *Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage*. *Blood*, 2009. **114**(26): p. 5307-5314.
105. Zenz, T., et al., *Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up*. *Blood*, 2008. **112**(8): p. 3322-3329.
106. Rossi, D., et al., *The Prognostic Value of TP53 Mutations in Chronic Lymphocytic Leukemia Is Independent of Del17p13: Implications for Overall Survival and Chemorefractoriness*. *Clinical Cancer Research*, 2009. **15**(3): p. 995.
107. Rossi, D., et al., *Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia*. *Blood*, 2014. **123**(14): p. 2139-2147.
108. *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*. 2021; Available from: <https://mitelmandatabase.isb-cgc.org>.
109. Nicholson, J.M. and D. Cimini, *Cancer karyotypes: survival of the fittest*. *Frontiers in oncology*, 2013. **3**: p. 148-148.
110. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instabilities in human cancers*. *Nature*, 1998. **396**(6712): p. 643-649.

111. Jeromin, S., et al., *SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients*. *Leukemia*, 2014. **28**(1): p. 108-17.
112. Rawstron, A.C., et al., *Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia*. *N Engl J Med*, 2008. **359**(6): p. 575-83.
113. Rossi, D., et al., *The prognosis of clinical monoclonal B cell lymphocytosis differs from prognosis of Rai 0 chronic lymphocytic leukaemia and is recapitulated by biological risk factors*. *Br J Haematol*, 2009. **146**(1): p. 64-75.
114. Shanafelt, T.D., et al., *Brief report: natural history of individuals with clinically recognized monoclonal B-cell lymphocytosis compared with patients with Rai 0 chronic lymphocytic leukemia*. *J Clin Oncol*, 2009. **27**(24): p. 3959-63.
115. Auer, R.L., et al., *The sequential analysis of trisomy 12 in B-cell chronic lymphocytic leukaemia*. *Br J Haematol*, 1999. **104**(4): p. 742-4.
116. Hjalmar, V., R. Hast, and E. Kimby, *Sequential fluorescence in situ hybridization analyses for trisomy 12 in chronic leukemic B-cell disorders*. *Haematologica*, 2001. **86**(2): p. 174-80.
117. Aamot, H.V., et al., *Non-Hodgkin lymphoma with t(14;18): clonal evolution patterns and cytogenetic-pathologic-clinical correlations*. *J Cancer Res Clin Oncol*, 2007. **133**(7): p. 455-70.
118. Antonelli, A., et al., *Cytogenetic features, clinical significance and prognostic impact of type 1 and type 2 papillary renal cell carcinoma*. *Cancer Genetics and Cytogenetics*, 2010. **199**(2): p. 128-133.
119. Kullendorff, C.-M., et al., *Cytogenetic findings and clinical course in a consecutive series of Wilms tumors*. *Cancer Genetics and Cytogenetics*, 2003. **140**(1): p. 82-87.
120. Matutes, E., et al., *Trisomy 12 defines a group of CLL with atypical morphology: correlation between cytogenetic, clinical and laboratory features in 544 patients*. *Br J Haematol*, 1996. **92**(2): p. 382-8.
121. Herishanu, Y., et al., *Integration of automated morphological features resolves a distinct group of atypical chronic lymphocytic leukemias with chromosomal aberrations*. *Leuk Res*, 2014. **38**(4): p. 484-9.
122. Oscier, D., et al., *The morphology of CLL revisited: the clinical significance of prolymphocytes and correlations with prognostic/molecular markers in the LRF CLL4 trial*. *Br J Haematol*, 2016. **174**(5): p. 767-75.
123. Melo, J.V., D. Catovsky, and D.A. Galton, *The relationship between chronic lymphocytic leukaemia and prolymphocytic leukaemia. II. Patterns of evolution of 'prolymphocytoid' transformation*. *Br J Haematol*, 1986. **64**(1): p. 77-86.
124. Schlette, E., et al., *CD79b expression in chronic lymphocytic leukemia. Association with trisomy 12 and atypical immunophenotype*. *Arch Pathol Lab Med*, 2003. **127**(5): p. 561-6.
125. Quijano, S., et al., *Impact of trisomy 12, del(13q), del(17p), and del(11q) on the immunophenotype, DNA ploidy status, and proliferative rate of leukemic B-cells in chronic lymphocytic leukemia*. *Cytometry B Clin Cytom*, 2008. **74**(3): p. 139-49.
126. Strati, P., et al., *Second cancers and Richter transformation are the leading causes of death in patients with trisomy 12 chronic lymphocytic leukemia*. *Clin Lymphoma Myeloma Leuk*, 2015. **15**(7): p. 420-7.
127. Cro, L., et al., *The clinical and biological features of a series of immunophenotypic variant of B-CLL*. *Eur J Haematol*, 2010. **85**(2): p. 120-9.
128. Kriston, C., et al., *Low CD23 expression correlates with high CD38 expression and the presence of trisomy 12 in CLL*. *Hematological Oncology*, 2017. **35**(1): p. 58-63.
129. Riches, J.C., et al., *Trisomy 12 chronic lymphocytic leukemia cells exhibit upregulation of integrin signaling that is modulated by NOTCH1 mutations*. *Blood*, 2014. **123**(26): p. 4101-10.
130. Bulian, P., et al., *Mutational status of IGHV is the most reliable prognostic marker in trisomy 12 chronic lymphocytic leukemia*. *Haematologica*, 2017. **102**(11): p. e443-e446.

131. Malavasi, F., et al., *CD38 and chronic lymphocytic leukemia: a decade later*. *Blood*, 2011. **118**(13): p. 3470-8.
132. Santos, F.P. and S. O'Brien, *Small lymphocytic lymphoma and chronic lymphocytic leukemia: are they the same disease?* *Cancer J*, 2012. **18**(5): p. 396-403.
133. Roos-Weil, D., et al., *Mutational and cytogenetic analyses of 188 CLL patients with trisomy 12: A retrospective study from the French Innovative Leukemia Organization (FILO) working group*. *Genes, Chromosomes and Cancer*, 2018. **57**(11): p. 533-540.
134. Strati, P., et al., *Eradication of bone marrow minimal residual disease may prompt early treatment discontinuation in CLL*. *Blood*, 2014. **123**(24): p. 3727-32.
135. Thompson, P.A., et al., *Trisomy 12 is associated with an abbreviated redistribution lymphocytosis during treatment with the BTK inhibitor ibrutinib in patients with chronic lymphocytic leukaemia*. *British Journal of Haematology*, 2015. **170**(1): p. 125-128.
136. Walter, H.S., et al., *Obinutuzumab-induced coagulopathy in chronic lymphocytic leukaemia with trisomy 12*. *Blood Cancer J*, 2016. **6**: p. e435.
137. Kipps, T.J., et al., *Long-Term Studies Assessing Outcomes of Ibrutinib Therapy in Patients With Del(11q) Chronic Lymphocytic Leukemia*. *Clinical Lymphoma Myeloma and Leukemia*, 2019. **19**(11): p. 715-722.e6.
138. Dietrich, S., et al., *Drug-perturbation-based stratification of blood cancer*. *J Clin Invest*, 2018. **128**(1): p. 427-445.
139. Rossi, D., et al., *Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia*. *Blood*, 2013. **121**(8): p. 1403-12.
140. Gonzalez-Gascon, Y.M.I., et al., *A high proportion of cells carrying trisomy 12 is associated with a worse outcome in patients with chronic lymphocytic leukemia*. *Hematol Oncol*, 2016. **34**(2): p. 84-92.
141. Baliakas, P., et al., *Additional trisomies amongst patients with chronic lymphocytic leukemia carrying trisomy 12: the accompanying chromosome makes a difference*. *Haematologica*, 2016. **101**(7): p. e299-302.
142. Cosson, A., et al., *14q deletions are associated with trisomy 12, NOTCH1 mutations and unmutated IGHV genes in chronic lymphocytic leukemia and small lymphocytic lymphoma*. *Genes Chromosomes Cancer*, 2014. **53**(8): p. 657-66.
143. Bulian, P., et al., *Mutational Status of IGHV is the most reliable prognostic Marker in Trisomy 12 Chronic Lymphocytic Leukemia*. *Haematologica*, 2017: p. haematol.2017.170340.
144. Baliakas, P., et al., *Additional trisomies amongst patients with chronic lymphocytic leukemia carrying trisomy 12: the accompanying chromosome makes a difference*. *Haematologica*, 2016. **101**(7): p. e299-e302.
145. Pospisilova, H., et al., *Interstitial del(14)(q) involving IGH: a novel recurrent aberration in B-NHL*. *Leukemia*, 2007. **21**(9): p. 2079-83.
146. Reindl, L., et al., *Biological and clinical characterization of recurrent 14q deletions in CLL and other mature B-cell neoplasms*. *Br J Haematol*, 2010. **151**(1): p. 25-36.
147. Vendramini, E., et al., *KRAS, NRAS, and BRAF mutations are highly enriched in trisomy 12 chronic lymphocytic leukemia and are associated with shorter treatment-free survival*. *Leukemia*, 2019. **33**(8): p. 2111-2115.
148. Hjalmar, V., *Sequential fluorescence in situ hybridization analysis for trisomy 12 in B-cell chronic lymphocytic leukemia*. *Methods Mol Med*, 2005. **115**: p. 231-40.
149. Garcia-Marco, J.A., C.M. Price, and D. Catovsky, *Interphase cytogenetics in chronic lymphocytic leukemia*. *Cancer Genet Cytogenet*, 1997. **94**(1): p. 52-8.
150. Porpacz, E., et al., *Gene expression signature of chronic lymphocytic leukaemia with Trisomy 12*. *Eur J Clin Invest*, 2009. **39**(7): p. 568-75.
151. Haslinger, C., et al., *Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status*. *J Clin Oncol*, 2004. **22**(19): p. 3937-49.

152. Maura, F., et al., *Insulin Growth Factor 1 Receptor Expression Is Associated with NOTCH1 Mutation, Trisomy 12 and Aggressive Clinical Course in Chronic Lymphocytic Leukaemia*. PLOS ONE, 2015. **10**(3): p. e0118801.
153. Decker, S., et al., *Trisomy 12 and elevated GLI1 and PTCH1 transcript levels are biomarkers for Hedgehog-inhibitor responsiveness in CLL*. Blood, 2012. **119**(4): p. 997-1007.
154. Zucchetto, A., et al., *CD38/CD31, the CCL3 and CCL4 chemokines, and CD49d/vascular cell adhesion molecule-1 are interchained by sequential events sustaining chronic lymphocytic leukemia cell survival*. Cancer Res, 2009. **69**(9): p. 4001-9.
155. Zucchetto, A., et al., *CD49d is overexpressed by trisomy 12 chronic lymphocytic leukemia cells: evidence for a methylation-dependent regulation mechanism*. Blood, 2013. **122**(19): p. 3317-21.
156. Benedetti, D., et al., *NOTCH1 mutations are associated with high CD49d expression in chronic lymphocytic leukemia: link between the NOTCH1 and the NF-kappaB pathways*. Leukemia, 2017.
157. Baumann, T., et al., *CD49d (ITGA4) expression is a predictor of time to first treatment in patients with chronic lymphocytic leukaemia and mutated IGHV status*. Br J Haematol, 2016. **172**(1): p. 48-55.
158. Ganghammer, S., et al., *CXCL12-induced VLA-4 activation is impaired in trisomy 12 chronic lymphocytic leukemia cells: a role for CCL21*. Oncotarget, 2015. **6**(14): p. 12048-60.
159. Fiorcari, S., et al.
160. Strati, P., et al., *CD49d associates with nodal presentation and subsequent development of lymphadenopathy in patients with chronic lymphocytic leukaemia*. British Journal of Haematology, 2017. **178**(1): p. 99-105.
161. Tissino, E., et al., *Functional and clinical relevance of VLA-4 (CD49d/CD29) in ibrutinib-treated chronic lymphocytic leukemia*. The Journal of experimental medicine, 2018. **215**(2): p. 681-697.
162. Dal Bo, M., et al., *CD49d prevails over the novel recurrent mutations as independent prognosticator of overall survival in chronic lymphocytic leukemia*. Leukemia, 2016. **30**(10): p. 2011-2018.
163. Balatti, V., et al., *NOTCH1 mutations in CLL associated with trisomy 12*. Blood, 2012. **119**(2): p. 329-31.
164. Baliakas, P., et al., *Recurrent mutations refine prognosis in chronic lymphocytic leukemia*. Leukemia, 2015. **29**(2): p. 329-36.
165. Del Giudice, I., et al., *NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL*. Haematologica, 2012. **97**(3): p. 437-41.
166. Rossi, D., et al., *Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia*. Blood, 2012. **119**(2): p. 521-9.
167. Gianfelici, V., *Activation of the NOTCH1 pathway in chronic lymphocytic leukemia*. Haematologica, 2012. **97**(3): p. 328-30.
168. Kopan, R. and M.X. Ilgan, *The canonical Notch signaling pathway: unfolding the activation mechanism*. Cell, 2009. **137**(2): p. 216-33.
169. Rosati, E., et al., *Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells*. Blood, 2009. **113**(4): p. 856-65.
170. Arruga, F., et al., *Functional impact of NOTCH1 mutations in chronic lymphocytic leukemia*. Leukemia, 2014. **28**(5): p. 1060-70.
171. Larrayoz, M., et al., *Non-coding NOTCH1 mutations in chronic lymphocytic leukemia; their clinical impact in the UK CLL4 trial*. Leukemia, 2017. **31**(2): p. 510-514.
172. Stilgenbauer, S., et al., *Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial*. Blood, 2014. **123**(21): p. 3247-54.

173. Pozzo, F., et al., *NOTCH1 mutations associate with low CD20 level in chronic lymphocytic leukemia: evidence for a NOTCH1 mutation-driven epigenetic dysregulation*. *Leukemia*, 2016. **30**(1): p. 182-9.
174. Arruga, F., et al., *Mutations in NOTCH1 PEST domain orchestrate CCL19-driven homing of chronic lymphocytic leukemia cells by modulating the tumor suppressor gene DUSP22*. *Leukemia*, 2017. **31**(9): p. 1882-1893.
175. Brochet, X., M.P. Lefranc, and V. Giudicelli, *IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis*. *Nucleic Acids Res*, 2008. **36**(Web Server issue): p. W503-8.
176. Bystry, V., et al., *ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy*. *Bioinformatics*, 2015. **31**(23): p. 3844-6.
177. Hertlein, E., et al., *Characterization of a new chronic lymphocytic leukemia cell line for mechanistic in vitro and in vivo studies relevant to disease*. *PLoS One*, 2013. **8**(10): p. e76607.
178. Rossi, D., et al., *Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia*. *Blood*, 2014. **123**(14): p. 2139-47.
179. Gahrton, G., et al., *Nonrandom chromosomal aberrations in chronic lymphocytic leukemia revealed by polyclonal B-cell-mitogen stimulation*. *Blood*, 1980. **56**(4): p. 640-7.
180. Hallek, M., et al., *iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL*. *Blood*, 2018. **131**(25): p. 2745-2760.
181. Forsberg, L.A., *Loss of chromosome Y (LOY) in blood cells is associated with increased risk for disease and mortality in aging men*. *Human Genetics*, 2017. **136**(5): p. 657-663.
182. Baliakas, P., et al., *Cytogenetic complexity in chronic lymphocytic leukemia: definitions, associations, and clinical impact*. *Blood*, 2019. **133**(11): p. 1205-1216.
183. Nadeu, F., et al., *Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia*. *Blood*, 2016. **127**(17): p. 2122-30.
184. Ghia, P., et al., *CD38 modifications in chronic lymphocytic leukemia: are they relevant?* *Leukemia*, 2004. **18**(10): p. 1733-1735.
185. Ghia, P., et al., *The pattern of CD38 expression defines a distinct subset of chronic lymphocytic leukemia (CLL) patients at risk of disease progression*. *Blood*, 2003. **101**(4): p. 1262-9.
186. Tissino, E., et al., *CD49d promotes disease progression in chronic lymphocytic leukemia: new insights from CD49d bimodal expression*. *Blood*, 2020. **135**(15): p. 1244-1254.
187. Heyman, B., A.D. Volkheimer, and J.B. Weinberg, *Double IGHV DNA gene rearrangements in CLL: influence of mixed-mutated and -unmutated rearrangements on outcomes in CLL*. *Blood Cancer J*, 2016. **6**(7): p. e440.
188. Rosenquist, R., et al., *Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations*. *Leukemia*, 2017. **31**(7): p. 1477-1481.
189. Langerak, A.W., et al., *Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases*. *Leukemia*, 2011. **25**(6): p. 979-84.
190. Stamatopoulos, B., et al., *Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia*. *Leukemia*, 2017. **31**(4): p. 837-845.
191. Oscier, D.G., et al., *The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial*. *Blood*, 2013. **121**(3): p. 468-75.
192. Yu, L.J., et al., *Comprehensive Genetic Characterization of 17p Deleted CLL Identifies Predictors of Overall Survival*. *Blood*, 2015. **126**(23).
193. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping*. *Nature Reviews Genetics*, 2011. **12**(5): p. 363-376.

194. Gunn, S.R., et al., *Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene*. *Leuk Res*, 2009. **33**(9): p. 1276-81.
195. Kolquist, K.A., et al., *Evaluation of chronic lymphocytic leukemia by oligonucleotide-based microarray analysis uncovers novel aberrations not detected by FISH or cytogenetic analysis*. *Molecular Cytogenetics*, 2011. **4**(1): p. 25.
196. Mestichelli, F., et al., *Array CGH analysis reveals deletion of chromosome 22q11 in CLL with normal karyotype and no fish alterations*. *Br J Haematol*, 2018. **183**(1): p. 152-155.
197. Mraz, M. and S. Pospisilova, *Detection of a deletion at 22q11 locus involving ZNF280A/ZNF280B/PRAME/GGTLC2 in B-cell malignancies: simply a consequence of an immunoglobulin lambda light chain rearrangement*. *British Journal of Haematology*, 2019. **186**(4): p. e91-e94.
198. Mraz, M., et al., *The origin of deletion 22q11 in chronic lymphocytic leukemia is related to the rearrangement of immunoglobulin lambda light chain locus*. *Leuk Res*, 2013. **37**(7): p. 802-8.
199. Zare, F., et al., *An evaluation of copy number variation detection tools for cancer using whole exome sequencing data*. *BMC Bioinformatics*, 2017. **18**(1): p. 286.
200. Hirsch, C.M., et al., *Pathogenic Relevance of Germ Line TET2 Alterations*. *Blood*, 2016. **128**(22).
201. Husby, S. and K. Grønbaek, *Mature lymphoid malignancies: origin, stem cells, and chronicity*. *Blood advances*, 2017. **1**(25): p. 2444-2455.
202. Damm, F., et al., *Acquired initiating mutations in early hematopoietic cells of CLL patients*. *Cancer Discov*, 2014. **4**(9): p. 1088-101.
203. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint*. *Nature*, 2003. **421**(6926): p. 961-966.
204. Salguero, I., et al., *MDC1 PST-repeat region promotes histone H2AX-independent chromatin association and DNA damage tolerance*. *Nature Communications*, 2019. **10**(1): p. 5191.
205. Bartkova, J., et al., *DNA damage response mediators MDC1 and 53BP1: constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours*. *Oncogene*, 2007. **26**(53): p. 7414-7422.
206. Minter-Dykhouse, K., et al., *Distinct versus overlapping functions of MDC1 and 53BP1 in DNA damage response and tumorigenesis*. *J Cell Biol*, 2008. **181**(5): p. 727-35.
207. Froimchuk, E., Y. Jang, and K. Ge, *Histone H3 lysine 4 methyltransferase KMT2D*. *Gene*, 2017. **627**: p. 337-342.
208. Beekman, R., et al., *The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia*. *Nat Med*, 2018. **24**(6): p. 868-880.
209. Tsagiopoulou, M., et al., *Chronic lymphocytic leukemias with trisomy 12 show a distinct DNA methylation profile linked to altered chromatin activation*. *Haematologica*, 2020: p. haematol.2019.240721.
210. Zhang, J., et al., *Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis*. *Nat Med*, 2015. **21**(10): p. 1190-8.
211. Gutierrez, A., et al., *The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia*. *Blood*, 2011. **118**(15): p. 4169-73.
212. Satterwhite, E., et al., *The BCL11 gene family: involvement of BCL11A in lymphoid malignancies*. *Blood*, 2001. **98**(12): p. 3413-20.
213. Punwani, D., et al., *Multisystem Anomalies in Severe Combined Immunodeficiency with Mutant BCL11B*. *N Engl J Med*, 2016. **375**(22): p. 2165-2176.
214. Hotinski, Anya K., et al., *A biclonal case of chronic lymphocytic leukaemia with discordant mutational status of the immunoglobulin heavy chain variable region and bimodal CD49d expression*. *British Journal of Haematology*, 2021. **192**(3): p. e77-e81.
215. Scheffold, A., et al., *IGF1R as druggable target mediating PI3K- δ inhibitor resistance in a murine model of chronic lymphocytic leukemia*. *Blood*, 2019. **134**(6): p. 534-547.

216. Davide, R., et al., *CD49d expression is an independent risk factor of progressive disease in early stage chronic lymphocytic leukemia*. *Haematologica*, 2008. **93**(10): p. 1575-1579.
217. Gattei, V., et al., *Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia*. *Blood*, 2008. **111**(2): p. 865-73.
218. Shanafelt, T.D., et al., *CD49d expression is an independent predictor of overall survival in patients with chronic lymphocytic leukaemia: a prognostic parameter with therapeutic potential*. *British journal of haematology*, 2008. **140**(5): p. 537-546.
219. Bulian, P., et al., *CD49d Is the Strongest Flow Cytometry–Based Predictor of Overall Survival in Chronic Lymphocytic Leukemia*. *Journal of Clinical Oncology*, 2014. **32**(9): p. 897-904.
220. Attia, H.R., et al., *ITGA4 gene methylation status in chronic lymphocytic leukemia*. *Future Science OA*, 2020. **6**(7): p. FSO583.
221. Robinson, J.T., et al., *Integrative genomics viewer*. *Nat Biotechnol*, 2011. **29**(1): p. 24-6.
222. Ortega-Molina, A., et al., *The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development*. *Nat Med*, 2015. **21**(10): p. 1199-208.
223. Ferreira, P.G., et al., *Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia*. *Genome research*, 2014. **24**(2): p. 212-226.
224. Mosquera Orgueira, A., et al., *Time to Treatment Prediction in Chronic Lymphocytic Leukemia Based on New Transcriptional Patterns*. *Frontiers in Oncology*, 2019. **9**(79).
225. Rendeiro, A.F., et al., *Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL*. *Nature Communications*, 2020. **11**(1): p. 577.
226. Brachtl, G., et al., *The pathogenic relevance of the prognostic markers CD38 and CD49d in chronic lymphocytic leukemia*. *Annals of Hematology*, 2014. **93**(3): p. 361-374.
227. Thurgood, L.A., et al., *Aberrant determination of phenotypic markers in chronic lymphocytic leukemia (CLL) lymphocytes after cryopreservation*. *Exp Hematol*, 2018. **63**: p. 28-32.e1.
228. Morin, R.D., et al., *Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma*. *Nature*, 2011. **476**(7360): p. 298-303.
229. Antonarakis, S.E., *Down syndrome and the complexity of genome dosage imbalance*. *Nature Reviews Genetics*, 2017. **18**(3): p. 147-163.
230. Beekman, R., et al., *The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia*. *Nature Medicine*, 2018. **24**(6): p. 868-880.
231. Shukla, V., et al., *A role for IRF4 in the development of CLL*. *Blood*, 2013. **122**(16): p. 2848-55.
232. Shaffer, A.L., et al., *IRF4 addiction in multiple myeloma*. *Nature*, 2008. **454**(7201): p. 226-31.
233. Iida, S., et al., *Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma*. *Nature Genetics*, 1997. **17**(2): p. 226-230.
234. Li, B. and W.-J. Chng, *EZH2 abnormalities in lymphoid malignancies: underlying mechanisms and therapeutic implications*. *Journal of Hematology & Oncology*, 2019. **12**(1): p. 118.
235. Papakonstantinou, N., et al., *The histone methyltransferase EZH2 as a novel prosurvival factor in clinically aggressive chronic lymphocytic leukemia*. *Oncotarget*, 2016. **7**(24).
236. Bödör, C., et al., *EZH2 mutations are frequent and represent an early event in follicular lymphoma*. *Blood*, 2013. **122**(18): p. 3165-8.
237. Morin, R.D., et al., *Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin*. *Nat Genet*, 2010. **42**(2): p. 181-5.
238. Decker, T., et al., *Immunostimulatory CpG-oligonucleotides cause proliferation, cytokine production, and an immunogenic phenotype in chronic lymphocytic leukemia B cells*. *Blood*, 2000. **95**(3): p. 999-1006.
239. Muthusamy, N., et al., *Enhanced detection of chromosomal abnormalities in chronic lymphocytic leukemia by conventional cytogenetics using CpG oligonucleotide in combination with pokeweed mitogen and phorbol myristate acetate*. *Cancer Genet*, 2011. **204**(2): p. 77-83.

240. Muzio, M., et al., *Expression and function of toll like receptors in chronic lymphocytic leukaemia cells*. British Journal of Haematology, 2009. **144**(4): p. 507-516.
241. Eleni, A., et al., *Toll-like receptor signaling pathway in chronic lymphocytic leukemia: distinct gene expression profiles of potential pathogenic significance in specific subsets of patients*. Haematologica, 2011. **96**(11): p. 1644-1652.
242. Rybka, J., et al., *The Expression of Toll-Like Receptors in Patients with B-Cell Chronic Lymphocytic Leukemia*. Archivum Immunologiae et Therapiae Experimentalis, 2016. **64**(1): p. 147-150.
243. Zhang, Q., et al., *Notch signal suppresses Toll-like receptor-triggered inflammatory responses in macrophages by inhibiting extracellular signal-regulated kinase 1/2-mediated nuclear factor κ B activation*. J Biol Chem, 2012. **287**(9): p. 6208-17.
244. Mansouri, L., et al., *NF- κ B activation in chronic lymphocytic leukemia: A point of convergence of external triggers and intrinsic lesions*. Semin Cancer Biol, 2016. **39**: p. 40-8.
245. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Research, 2000. **28**(1): p. 27-30.
246. Spaner, D.E. and A. Masellis, *Toll-like receptor agonists in the treatment of chronic lymphocytic leukemia*. Leukemia, 2007. **21**(1): p. 53-60.
247. Struski, S., et al., *Stimulation of B-cell lymphoproliferations with CpG-oligonucleotide DSP30 plus IL-2 is more effective than with TPA to detect clonal abnormalities*. Leukemia, 2009. **23**(3): p. 617-619.
248. Koyama, S., et al., *The potential of various lipopolysaccharides to release IL-8 and G-CSF*. Am J Physiol Lung Cell Mol Physiol, 2000. **278**(4): p. L658-66.
249. Kovacs, G., et al., *Minimal Residual Disease Assessment Improves Prediction of Outcome in Patients With Chronic Lymphocytic Leukemia (CLL) Who Achieve Partial Response: Comprehensive Analysis of Two Phase III Studies of the German CLL Study Group*. Journal of Clinical Oncology, 2016. **34**(31): p. 3758-3765.
250. Tissino, E., et al., *Functional and clinical relevance of VLA-4 (CD49d/CD29) in ibrutinib-treated chronic lymphocytic leukemia*. J Exp Med, 2018. **215**(2): p. 681-697.
251. Gounari, M., et al., *Excessive antigen reactivity may underlie the clinical aggressiveness of chronic lymphocytic leukemia stereotyped subset #8*. Blood, 2015. **125**(23): p. 3580-3587.
252. Butzkueven, H., et al., *Long-term safety and effectiveness of natalizumab treatment in clinical practice: 10 years of real-world data from the Tysabri Observational Program (TOP)*. Journal of Neurology, Neurosurgery & Psychiatry, 2020. **91**(6): p. 660-668.
253. Hertlein, E., et al., *Characterization of a new chronic lymphocytic leukemia cell line for mechanistic in vitro and in vivo studies relevant to disease*. PloS one, 2013. **8**(10): p. e76607-e76607.
254. Goldin, L.R., et al., *Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: results from the Swedish Family-Cancer Database*. Blood, 2004. **104**(6): p. 1850-1854.
255. Di Bernardo, M.C., et al., *A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia*. Nature Genetics, 2008. **40**(10): p. 1204-1210.
256. Berndt, S.I., et al., *Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia*. Nature Genetics, 2013. **45**(8): p. 868-876.
257. Genovese, G., et al., *Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence*. N Engl J Med, 2014. **371**(26): p. 2477-87.
258. Jaiswal, S., et al., *Age-related clonal hematopoiesis associated with adverse outcomes*. N Engl J Med, 2014. **371**(26): p. 2488-98.
259. Terao, C., et al., *Chromosomal alterations among age-related haematopoietic clones in Japan*. Nature, 2020. **584**(7819): p. 130-135.
260. Nakaoka, H. and I. Inoue, *Distribution of HLA haplotypes across Japanese Archipelago: similarity, difference and admixture*. Journal of Human Genetics, 2015. **60**(11): p. 683-690.

261. Cerhan, J.R., et al., *Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma*. *Blood*, 2007. **110**(13): p. 4455-63.
262. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation*. *Nucleic acids research*, 2001. **29**(1): p. 308-311.

9 APPENDIX

9.1 PRIMERS

Table 9-1. Primer sequences for Sanger sequencing.

Primer Set Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
<i>BCL11B</i>	CTCCTGAACCCCTTCCAG	CGAGCCGGCCTTGTGCAT
<i>BIRC3_1</i>	AGACTTCTGTTGCCTTGAAATGAG	ACACCTGGCTTCATGTTCCC
<i>BIRC3_2</i>	GATAAAAGCAAAGCCATGCACA	AACCAGCACGAGCAAGACTC
<i>BIRC3_3</i>	GGAGACAGAGTGGCTTGCTT	GCAGATTCAGTTTCTTACCCACA
<i>BIRC3_4</i>	TGCCGTGGAAATGGGCTTTA	GGACACAACGTCAGCTATCCA
<i>KMT2D</i>	TGGTCCACGGAGGTGTATGA	GGGAGCACTTGGTTAGCAGT
<i>NOTCH1</i>	TGCACACTATTCTGCCCCAG	ACTTGAAGGCCTCCGGAATG
<i>SF3B1</i>	GAGTATTTGGTTTTTCATGATGTTGC	CTGCTGCTCCCAAATTACCC
<i>TET2</i>	GTGCCCTTATCTGCTGCAAG	CTCAGCGTCTCGGTAAGCTC
<i>TP53</i>	GGTTGGCTCTGACTGTACCA	CCTGCTTGCTTACCTCGCTTA

Table 9-2. Primer sets used to confirm listed *BIRC3* mutations in Chapter 3.

Primer Set	<i>BIRC3</i> mutation(s) to confirm
<i>BIRC3_1</i>	c.1639delC; c.1664_1665insTT; c_1665_1666delA; c.1759A>T
<i>BIRC3_2</i>	c.73G>A
<i>BIRC3_3</i>	c.784T>C
<i>BIRC3_4</i>	c.1298delAAAinsA

Table 9-3. Primer sequences for qRT-PCR.

Primer Set Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
<i>GUSB</i>	GAAAATACGTGGTTGGAGAGCTCATT	CCGAGTGAAGATCCCCTTTTTA
<i>ITGA4</i>	TACAGATGCAGGATCGGAAAGA	AGGTTCTCCATTAGGGCTACC
<i>TLR4</i>	TTTGGACAGTTTCCACATTGA	AAGCATTCCACCTTTGTTGG
<i>IL8</i>	ACTGAGAGTGATTGAGAGTGGAC	AACCCTCTGCACCCAGTTTTTC
<i>EZH2</i>	CCCTGACCTCTGTCTTACTTGTGGA	ACGTCAGATGGTGCCAGCAATA
<i>IRF4</i>	GCGGTGCGCTTTGAACAAG	ACACTTTGTACGGGTCTGAGA

Table 9-4. Primer sequences for bisulfite-treated DNA. *ITGA4* primers 1&2 amplify product 1 (see gene map Figure 5-9). *ITGA4* primers 3&4 amplify product 2, and *ITGA4* primer 5 was used as a nested primer in product 1 for sequencing purposes.

Primer Set Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
<i>ITGA4 1&2 (product 1)</i>	TATTTAAATGTTTTTTAGGGGTTTT	CCAAACACAACAACAACATCA
<i>ITGA4 3&4 (product 2)</i>	GTGATGTTGTTGTTGTGTTTGG	AATCACTAACTCCTAATCCCATAC
<i>ITGA4 5 (nested product 1)</i>	TAGTTTGGGGTTATATAGTT	-

9.2 SUMMARY OF PATIENTS: SAMPLES & CLINICAL DETAILS

Table 9-5. Summary of entire patient cohort: identification, sample information and clinical details. Patients 1-67 were the initial cohort described in Chapter 3 (current to 2018). Patients 1-22 are the cohort with available samples fully characterised prospectively in Chapter 3. Patients 23-49 have available clinical & cytogenetic information, but no samples. Patients 50-67 have cytogenetic information but no samples or clinical information. Patients 68-73 (shaded grey) were added in 2020. Patients DIS1-DIS5 are CLL patients with /disomy 12 used as comparators in the immunophenotyping experiments of Chapter 3. Patients DIS6-DIS15 (shaded grey) are also disomy 12 CLL samples and were added in 2020. Note in bold: DIS2, DIS6 and DIS8 were excluded due to incorrect diagnosis (DIS2) or remission samples (DIS6 and DIS8). HEA1-3 are the healthy controls used for flow cytometry in Chapters 3 and 6. FMC =Flinders Medical Centre. SACRB = South Australian Research Biobank. CLL06 refers to the CLL06 clinical trial (samples stored at FMC). SLL = Small Lymphocytic Lymphoma. N/A = not applicable.

Patient ID in thesis	Age	Gender	FMC/SACRB/CLL06 ID	Sample		Clinical information				
				Date collected	Pre-treated at time of sample?	Date of diagnosis	Rai stage at diagnosis	Date of first treatment	Date of death	Cause of death
1	80	F	FMC30/CLL1536	21/10/2015	yes	21/03/2005	N/A (SLL)	28/03/2011	N/A	N/A
2	75	M	FMC34/CLL1407	02/04/2014	yes	4/03/2009	II	30/03/2014	1/09/2016	metastatic Ca
3	67	M	FMC69/CLL1415	26/06/2014	yes	29/10/2007	IV	30/10/2007	N/A	N/A
4	68	M	FMC263/CLL16121	19/08/2016	no	22/03/2016	I	30/08/2016	N/A	N/A
5	48	M	FMC136/CLL1023	03/08/2010	no	23/06/2010	II	27/09/2010	N/A	N/A
6	91	M	FMC143/CLL1324	17/04/2013	no	13/06/2008	III	1/05/2013	N/A	N/A
7	68	F	FMC258/CLL1626/ CLL1764/CLL1827	03/03/2016	no	14/08/2013	0	N/A	N/A	N/A
8	84	F	FMC232/CLL1430	30/09/2014	yes	6/02/2001	III	1/10/2010	1/10/2016	progressive CLL/caecal Ca
9	70	M	WES010	19/10/2011	no	Not available				
10	66	M	GOS033	17/07/2012	no					
11	71	F	GOS038	27/08/2012	no					
12	67	M	PMC051	19/11/2012	no					
13	62	M	AUS062	26/02/2013	no					
14	67	M	PMC075	01/07/2013	no					

Patient ID in thesis	Age	Gender	FMC/SACRB/CLL06 ID	Sample		Clinical information				
				Date collected	Pre-treated at time of sample?	Date of diagnosis	Rai stage at diagnosis	Date of first treatment	Date of death	Cause of death
15	61	M	CON120	23/06/2014	no	Not available				
16	72	M	H00002528	07/12/2015	yes	16/12/2005		1/12/2010	N/A	N/A
17	69	M	H00002629	20/01/2016	no	3/10/2008		12/02/2016	N/A	N/A
18	58	M	H00002007	16/03/2015	no	13/03/2016	II	1/06/2016	N/A	N/A
19	79	M	H00002617	18/01/2016	no	2/05/2007	III	1/02/2016	N/A	N/A
20	72	M	H00002527	07/12/2015	yes	01/08/2012	IV	1/09/2014	N/A	N/A
21	80	M	H00001407	22/05/2014	no	01/07/2007	IV	N/A	Died - date unclear	Unknown
22	44	M	FMC279/1723	30/03/2017	no	14/10/2016	0	N/A	N/A	N/A
23	62	M	-	No available cryopreserved samples		16/12/2009	Not recorded	14/03/2013	1/01/2016	Richter's syndrome
24	68	M	-			7/07/2004		Unclear	9/12/2016	progressive CLL/sepsis
25	68	M	-			7/12/2012		24/09/2015	N/A	N/A
26	63	M	-			15/12/2006		2/05/2012	N/A	N/A
27	66	F	-			1/06/2009		26/02/2015	N/A	N/A
28	60	F	-			8/11/2004		19/01/2005	N/A	N/A
29	64	M	-			Not available		Not available	Not available	N/A
30	66	M	-			30/06/2005		1/10/2007	N/A	N/A
31	73	M	-			Not available		Not available	Not available	N/A
32	78	F	-			3/02/2010		9/01/2017	N/A	N/A
33	79	F	FMC76			7/02/2005		3/01/2014	N/A	N/A
34	81	M	-			20/04/1998		Unclear	1/06/2016	progressive CLL/caecal Ca
35	56	F	-			6/01/2010		1/05/2013	1/02/2015	Richter's syndrome
36	70	F	FMC279			20/06/2014		1/04/2015	N/A	N/A
37	53	M	-			18/11/2015		19/04/2016	N/A	N/A
38	71	F	-			9/09/2013		23/08/2016	N/A	N/A
39	58	M	-	26/11/2010	16/03/2016	N/A	N/A			

Patient ID in thesis	Age	Gender	FMC/SACRB/CLL06 ID	Sample		Clinical information					
				Date collected	Pre-treated at time of sample?	Date of diagnosis	Rai stage at diagnosis	Date of first treatment	Date of death	Cause of death	
40	69	F	-	No available cryopreserved samples		30/07/2002	Not recorded	8/01/2003	N/A	N/A	
41	75	M	-			27/03/2015		8/03/2017	N/A	N/A	
42	78	M	FMC174			4/11/2015		19/12/2015	4/12/2016	progressive CLL	
43	76	M	FMC197			2/07/2015		N/A	N/A	N/A	
44	87	M	-			28/03/2014		13/04/2014	21/05/2014	progressive CLL	
45	80	F	-			4/08/2016		N/A	N/A	N/A	
46	80	F	-			17/09/2009		Unclear	10/12/2014	progressive CLL/sepsis	
47	90	M	-			6/05/1997		1/11/2001	2/04/2003	progressive CLL/sepsis	
48	82	F	-			20/09/2005		27/08/2015	N/A	N/A	
49	74	M	-			2/03/2013		N/A	1/07/2013	progressive CLL/sepsis	
50	51	F	-								
51	89	M	-								
52	80	M	-								
53	49	M	-								
54	61	F	-								
55	90	M	-								
56	89	F	-								
57	85	M	-								
58	82	M	-								
59	85	F	-								
60	68	F	-								
61	74	M	-								
62	37	M	-								
63	67	M	-								
64	72	F	-								
65	64	F	-								
66	72	M	-								

Patient ID in thesis	Age	Gender	FMC/SACRB/CLL06 ID	Sample		Clinical information				
				Date collected	Pre-treated at time of sample?	Date of diagnosis	Rai stage at diagnosis	Date of first treatment	Date of death	Cause of death
67	69	M	-	No available sample		No available clinical information				
68	65	M	FMC301/CLL1852	19/09/2018	no	Not audited				
69	65	M	FMC312/CLL1920	11/07/2018	no					
70	50	F	FMC304/CLL1843	07/08/2018	no					
71	63	M	FMC333/CLL2039	29/04/2020	no					
72	74	F	FMC313/CLL1677	19/04/2016	no					
73	66	M	CLL2065	05/08/2020	no					
DIS1	60	F	FMC20/CLL1658	30/03/2016	no	18/11/2013	0	05/05/2016	N/A	N/A
DIS2	84	M	FMC101/CLL1682	29/04/2016	yes	02/03/2005	N/A (SLL)	16/11/2012	N/A	N/A
DIS3	69	M	FMC140/CLL1687	11/05/2016	no	12/11/2007	0	N/A	N/A	N/A
DIS4	73	F	FMC17/CLL1697	06/06/2016	no	26/06/2015	0	N/A	N/A	N/A
DIS5	62	M	FMC268/CLL16148	21/12/2016	yes	01/04/2007	I	1/10/2009	N/A	N/A
DIS6	57	M	FMC19/CLL1812	14/03/2018	yes	Not audited				
DIS7	65	M	FMC10/CLL1810	14/03/2018	no					
DIS8	38	F	FMC272/CLL1750	23/08/2017	yes					
DIS9	48	M	FMC281/CLL1732	08/06/2017	no					
DIS10	56	M	FMC331/CLL1838	11/07/2018	no					
DIS11	49	M	FMC330/CLL1928	06/08/2019	no					
DIS12	77	F	FMC234/CLL1829	04/06/2018	no					
DIS13	75	F	FMC138/CLL1836	05/07/2018	no					
DIS14	67	M	FMC297/CLL1826	29/05/2018	no					
DIS15	71	M	FMC211/CLL1916	06/06/2019	yes					
HEA1	65	M	NOR1802	07/03/2018	N/A					
HEA2	73	F	NOR1803	08/03/2018						
HEA3	68	M	NOR1806	21/03/2018						

9.3 SUMMARY OF PATIENTS: CYTOGENETICS RESULTS

Table 9-6. Summary of cytogenetics results for entire trisomy 12 cohort & non-trisomy 12 CLL controls. Complex = three or more chromosomal abnormalities; + refers to chromosomal gain; N/A = not applicable; - refers to chromosomal loss; LOH = loss of heterozygosity; CBA = chromosomal banding analysis. Patients shaded in grey were added in 2020.

Patient ID	Cytogenetics				
	Modality	Date	Frequency of trisomy 12 clone (%)	Number of additional abnormalities	Nature of additional abnormalities
1	array	22/10/2015	47	>3	complex
2	array	25/08/2015	82	3	complex, +18
3	array	22/06/2015	63	0	N/A
4	array	15/08/2016	50	1	14q-
5	FISH	29/07/2010	28	0	N/A
6	FISH	29/04/2013	31	0	N/A
7	array	3/03/2016	47	0	N/A
8	array	17/08/2016	82	>3	complex
9	FISH	19/10/2011	82	0	N/A
10	FISH	17/07/2012	80	1	13q-
11	FISH	27/08/2012	83	0	N/A
12	FISH	19/11/2012	75	0	N/A
13	FISH	26/02/2013	35	1	13q-
14	FISH	1/07/2013	89	1	13q-
15	FISH	23/06/2014	60	0	N/A
16	CBA	25/03/2015	19	0	N/A
17	array	2/05/2016	86	0	N/A
18	array	17/03/2015	70	0	N/A
19	array	23/10/2015	90	1	LOH 1p
20	array	22/08/2014	99	0	N/A
21	array	28/01/2015	63	1	8q+
22	array	30/03/2017	78	1	2p+
23	FISH	25/02/2013	55	0	N/A
24	FISH	22/12/2012	44	0	N/A
25	array	9/10/2015	85	>3	complex
26	array	11/01/2017	78	0	N/A
27	CBA	22/08/2014	90	1	X-
28	array	21/10/2014	22	1	18q+
29	array	19/10/2016	unavailable	1	LOH 9q
30	CBA	28/06/2012	15	0	N/A
31	CBA	30/05/2014	6	1	Der17
32	array	22/06/2016	63	0	N/A
33	FISH	19/12/2013	85	0	N/A
34	array	28/08/2015	85	>3	complex with LOH at 12q
35	array	4/02/2015	90	>3	complex
36	array	24/03/2015	62	>3	complex with 17p-

Patient ID	Cytogenetics				
	Modality	Date	Frequency of trisomy 12 clone (%)	Number of additional abnormalities	Nature of additional abnormalities
37	array	4/12/2016	60	>3	complex
38	array	15/06/2016	82	>3	complex with 11q-, 14q-
39	array	23/09/2015	unavailable	0	N/A
40	CBA	23/08/2014	19	2	14q-, X-
41	array	20/12/2016	74	1	1q-
42	array	19/11/2015	15	2	2q-, 11q-
43	array	2/07/2015	22	2	7p-, 8p-
44	CBA	31/03/2014	40	0	N/A
45	array	8/05/2016	60	2	3q- 4q-
46	array	14/03/2013	20	>3	complex
47	CBA	1/01/2001	35	0	N/A
48	array	18/08/2015	86	2	7q+ 7q-
49	FISH	21/03/2013	54	0	N/A
50	FISH	27/06/2013	9	0	N/A
51	array	25/05/2016	unavailable	3	6p- 10q-
52	array	26/11/2015	unavailable	3	1q-, 13q-, +19
53	array	5/12/2016	unavailable	0	N/A
54	FISH	23/08/2014	70	1	13q-
55	CBA	30/05/2013	42	2	+18, +19
56	array	11/10/2015	unavailable	2	13q-, LOH Xq
57	array	20/07/2015	20	>3	complex with 11q-, 13q-, 14q-
58	array	6/09/2015	unavailable	0	N/A
59	array	24/07/2015	unavailable	2	4p-, 11q-
60	array	14/01/2016	unavailable	1	14q-
61	CBA	27/03/2014	16	0	N/A
62	array	6/06/1953	unavailable	1	14q-
63	array	1/06/2017	unavailable	0	N/A
64	array	3/12/2015	unavailable	0	N/A
65	array	14/01/2016	unavailable	0	N/A
66	array	18/05/2016	unavailable	>3	complex with 17p-
67	array	2/05/2016	82	0	N/A
68	array	31/08/2018	70	0	N/A
69	array	11/07/2019	40	1	LOH 11q
70	array	08/08/2018	33	0	N/A
71	array	30/11/2019	60	1	13q-
72	array	24/10/2018	41	1	13q-
73	array	20/08/2020	60	2	+16, 22q-
DIS1	array	13/01/2014	All normal diploid karyotype by SNP array with no acquired changes		
DIS2	array	09/12/2016			
DIS3	array	20/04/2016			
DIS4	array	18/08/2015			
DIS5	array	21/12/2016			
DIS6	array	14/03/2018	Complex karyotype		

Patient ID	Cytogenetics				
	Modality	Date	Frequency of trisomy 12 clone (%)	Number of additional abnormalities	Nature of additional abnormalities
DIS7	array	14/03/2018			Complex karyotype
DIS8	array	31/05/2017			Biallelic 13q-
DIS9	array	07/06/2017			11q-
DIS10	array	11/07/2018			Biallelic 13q-
DIS11	array	11/12/2018			Complex karyotype with 11q-
DIS12	array	17/11/2016			13q-
DIS13	array	01/03/2017			Normal diploid karyotype
DIS14	array	12/04/2018			Complex karyotype with 17p-
DIS15	array	06/06/2019			Complex karyotype with 14q-, 17p-

9.4 SUMMARY OF PATIENTS: IGHV & NGS RESULTS

Table 9-7. Summary of IGHV and NGS panel results for trisomy 12 cohort with available samples (patients 1-22). IGHV = immunoglobulin heavyvariable gene; NGS = next-generation sequencing. U = unmutated; M = mutated; wt = wildtype.

Patient ID	IGHV			NGS				
	IGHV-gene	U/M	Stereotype	ATM	BIRC3	NOTCH1	SF3B1	TP53
1	V3-07	M	no	wt	c.784T>C	wt	wt	wt
2	V1-08	U	no	c.7298-3C>T	c.1639delC	wt	wt	wt
3	V2-05*10	U	no	wt	wt	wt	wt	wt
4	V5-05*01	U	no	wt	wt	wt	wt	wt
5	V3-09*01	M	no	wt	wt	wt	wt	wt
6	V3-21*01 & V4-34*01	1 clone U/other M	unknown	c.8786+8A>C	wt	wt	wt	wt
7	V7-4*01	U	yes (CLL#1)	wt	wt	wt	wt	wt
8	V6-01*01	U	no	wt	wt	wt	wt	wt
9	V3-07*03	U	no	c.6919C>T	wt	wt	wt	wt
10	V4-39*07	U	no	wt	wt	wt	wt	wt
11	V4-39*01	U	yes (CLL#8)	wt	c.73G>A	c.7541_7542delCT	wt	wt
12	V3-07*03	M	no	wt	c.1665_1666delAA	wt	wt	wt
13	V3-73*03	M	no	c.2572T>C; c.3161C>G	wt	wt	wt	wt
14	V4-39*07	M	no	wt	wt	wt	c.2558T>C	wt
15	V1-69*01	U	no	wt	wt	wt	wt	wt
16	V3-49*03	U	no	wt	c.1664_1665insTT; c.1759A>T	wt	wt	wt
17	V4-39*01	U	yes (CLL#8)	c.2572T>C; c.3161C>G	wt	wt	wt	wt
18	V1-08	U	no	wt	wt	wt	wt	wt
19	V4-59*01	U	no	wt	wt	wt	wt	c.731G>A
20	V1-69*01	U	no	wt	wt	wt	wt	wt

Patient ID	IGHV			NGS				
	IGHV-gene	U/M	Stereotype	ATM	BIRC3	NOTCH1	SF3B1	TP53
21	V3-30*03	U	no	wt	c.1298_delAAAinsA	c.7541_7542delCT	wt	wt
22	V1-69*01	U	no	c.1363G>A	wt	wt	wt	wt
68	V1-02	U	yes (CLL#1)	Not performed				
69	V4-61	U	no					
70	V3-72 (productive) & V3-33 (unproductive)	M (both clones)	no					
71	V3-33*01	U	no					
72	V3-11	M	no					
73	V1-69	U	no					
DIS1	V6-01	U	no					
DIS2	V3-21	M	no					
DIS3	V3-30	M	no	Not performed				
DIS4	V4-59	M	no					
DIS5	V4-59	U	no					
DIS6	V3-73	U	no					
DIS7	V4-30*04	U	no					
DIS8	V3-30	M	no	Not performed (DIS14 harboured a TP53 mutation on Sanger sequencing performed at the treating clinician's request)				
DIS9	V1-69	U	no					
DIS10	V3-13	U	no					
DIS11	V3-15	U	no					
DIS12	V3-30	M	no					
DIS13	V3-30*03	U	no					
DIS14	V4-39	U	no					
DIS15	V2-05	M	no					

9.5 SUMMARY OF PATIENTS: TESTS PERFORMED

Table 9-8. Summary of tests performed on patient samples. Samples in grey were added in 2020. IGHV = immunoglobulin heavy variable gene; SNP = single nucleotide polymorphism; NGS = next-generation sequencing; WES = whole exome sequencing; RT-PCR = reverse-transcriptase PCR; TLR = toll-like receptor; LPS = lipopolysaccharide.

Patient ID	Chapter 3				Chapter 4		Chapter 5	Chapter 5 & 6	Chapter 6	
	IGHV	SNP array	NGS panel	Integrin flow panel	Sort	WES & RNAseq	Bisulfite sequencing	RT-PCRs	TLR flow panel	LPS stimulation assay
1	✓	✓	✓	✓						
2	✓	✓	✓	✓				✓	✓	
3	✓	✓	✓	✓						
4	✓	✓	✓	✓						
5	✓		✓	✓				✓	✓	✓
6	✓	✓	✓	✓	✓	✓	✓			
7	✓	✓	✓	✓				✓	✓	✓
8	✓	✓	✓	✓				✓		
9	✓		✓	✓				✓		
10	✓		✓	✓				✓	✓	
11	✓		✓	✓				✓	✓	
12	✓		✓	✓					✓	✓
13	✓		✓	✓						
14	✓		✓	✓				✓	✓	✓
15	✓		✓	✓				✓	✓	✓
16	✓		✓	✓						
17	✓	✓	✓	✓						
18	✓	✓	✓	✓					✓	✓
19	✓	✓	✓	✓						
20	✓	✓	✓	✓					✓	
21	✓	✓	✓	✓	✓		✓	✓	✓	
22	✓	✓	✓	✓				✓	✓	
68	✓	✓						✓	✓	
69	✓	✓								
70	✓	✓						✓	✓	
71	✓	✓							✓	
72	✓	✓								
73	✓	✓							✓	✓
DIS1	✓	✓		✓				✓	✓	
DIS2	✓	✓		✓						
DIS3	✓	✓		✓					✓	
DIS4	✓	✓		✓					✓	✓
DIS5	✓	✓		✓				✓	✓	✓
DIS6	✓	✓								
DIS7	✓	✓							✓	
DIS8	✓	✓								
DIS9	✓	✓							✓	
DIS10	✓	✓						✓	✓	
DIS11	✓	✓						✓	✓	

Patient ID	Chapter 3				Chapter 4		Chapter 5	Chapter 5 & 6	Chapter 6	
	IGHV	SNP array	NGS panel	Integrin flow panel	Sort	WES & RNAseq	Bisulfite sequencing	RT-PCRs	TLR flow panel	LPS stimulation assay
DIS12	✓	✓						✓	✓	
DIS13	✓	✓						✓	✓	
DIS14	✓	✓						✓	✓	
DIS15	✓	✓								✓

9.6 WES VARIANTS COMMON TO BOTH LEUKAEMIC CLONES (PATIENT 6)

Table 9-9. Variants common to both CD49d+ and CD49d- CLL clones detected in WES of sorted fractions of patient 6. None of the variants were detected in the T cell fraction. SNV = single nucleotide polymorphism; aa = amino acid; VAF = variant allele frequency; COSMIC = Catalogue of Somatic Mutations in Cancer (<https://cancer.sanger.ac.uk>); UTR = untranslated region; SNP = single nucleotide polymorphism.

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>GSTM4</i>	1p13.3	Missense SNV	c.523T>A	p.Tyr175Asn	48	3	153		
<i>OR2T2</i>	1p25.1	Frameshift variant	c.959_*80delTG ATCAGGAAGGG CTAGCAGGGAC TCCCACAGCATC AGAGTGGTGAC TGTGATCGGGA AGGATTAGCGG GGACTCCCAGA GCATCAGGGGT GGTGAC	p.Val320fs	6	2	58		
<i>PDE4DIP</i>	1q21.1	Synonymous SNV	c.276G>A	-	12	13	509		
<i>FLG2</i>	1q21.3	Missense SNV	c.6698C>G	p.Ala2233Gly	5	14	73		
<i>TOMM20</i>	1q42.3	Splice region & intron variant	c.394-8dupT	-	17	24	18	COSM1723322	Melanoma (1) & biliary ca (1)
<i>XPO1</i>	2p15	Intron variant	c.760-209T>G	-	50	67	6		Not the mutation described in CLL

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>ZRANB3</i>	2q21.3	Splice region & intron variant	n.498-8delT	-	32	24	37		
<i>TTN</i>	2q31.2	Synonymous SNV	c.39063A>G	-	9	20	66		
<i>TTN</i>	2q31.2	Intron variant	c.100766-12_100766-10delTTT	-	50	71	8	COSM3729017	Stomach ca (1)
<i>SESTD1</i>	2q31.2	Intron variant	c.55+90delT	-	8	11	37		
<i>ABCA12</i>	2q35	Intron variant	c.7105-66delT	-	8	45	13		
<i>MROH2A</i>	2q37.1	Intron variant	c.2173-93_2173-92delGT	-	11	15	75		
<i>ANKRD28</i>	3p25.1	3' UTR variant	c.*1914delA	-	21	5	19		
<i>UCHL1</i>	4p13	Splice region & intron variant	c.326-4delT	-	8	14	75		
<i>MGARP</i>	4q31.1	3' UTR variant	c.*91delA	-	25	32	29		
<i>HCN1</i>	5p12	Intron variant	c.1230+105delA	-	30	26	31		
<i>KIF2A</i>	5q12.1	Intron variant	c.65-90delT	-	9	27	22		
<i>POLR3G</i>	5q14.3	Intron variant	c.304+69delT	-	10	16	44		
<i>HLA-DRB5</i>	6p21.32	Intron variant	c.100+159G>A	-	7	19	16		
<i>MDC1</i>	6p21.33	Inframe	c.4824_4946del	p.Pro1609_Thr1649	47	42	120		

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
		deletion		del					
<i>CCDC129</i>	7p14.3	Frameshift variant	c.3106delT	p.Ser1036fs	21	25	38		
<i>MICALL2</i>	7p22.3	Synonymous SNV	c.1983A>C	-	25	24	126	COSM3745609	Now reported as SNP
<i>STAG3L4</i>	7q11.21	Splice region & intron variant	n.266-6delT	-	17	5	54		
<i>BCL7B</i>	7q11.23	Splice region & intron variant	n.93-4delT	-	29	20	35	COSM5446608	Thyroid ca (1)
<i>SLC13A1</i>	7q31.32	Intron variant	c.1513-59_1513-58delTT	-	21	33	16		
<i>STAU2</i>	8q21.11	Intron variant	c.1162-53C>A	-	40	18	18		
<i>BTA1F1</i>	10q23.32	Splice region & intron variant	c.2428-6dupT	-	19	4	51		
<i>MAPK8</i>	10q11.22	Intron variant	c.688+307delT	-	24	25	70		
<i>KCNIP2</i>	10q24.32	5' UTR variant	c.-32A>C	-	25	29	64		
<i>MTG1</i>	10q26.3	Splice region variant	n.413delA	-	17	18	49		
<i>KIF18A</i>	11p14.1	Intron variant	c.483+165delT	-	14	43	7		
<i>FKBP2</i>	11q13.1	3' UTR variant	c.*63delA	-	17	31	18		
<i>ABCD2</i>	12q12	Intron variant	c.940-57_940-56delAA	-	33	44	25		

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>GNPTAB</i>	12q23.2	Intron variant	c.1613-25delA	-	11	9	81		
<i>SSTR1</i>	14q21.1	3' UTR variant	c.*20_*43dupT CTGAGCCCCGGG CCACGCAGGGG C	-	34	45	29	COSM4603501 Head & neck ca (12)	
<i>MAX</i>	14q23.3	3' UTR variant	c.*128delT	-	15	2	43		
<i>PPP2R5C</i>	14q32.31	Intron variant	c.94-9113delT	-	15	15	20		
<i>IGHJ3</i>	14q32.33	Synonymous SNV	c.39C>T	-	99	100	274		
<i>PDIA3</i>	15q15.3	Intron variant	c.1137+111delT	-	30	30	24		
<i>CACNA1H</i>	16p13.3	Intron variant	c.2452-84_2452-42delACGGGTG GGGGCCCCAGA TCAGTGCCGGT GAGGGGTGGGA GCC	-	100	100	27		
<i>PLIN4</i>	19p13.3	Missense SNV	c.2704G>A	p.Gly902Ser	8	3	145		
<i>ELAVL3</i>	19p13.2	3' UTR variant	c.*53_*56delGA GA	-	33	29	21		
<i>SUGP2</i>	19p13.11	Intron variant	c.122-18_122-17delTT	-	19	16	34		
<i>ZNF880</i>	19q13.41	Synonymous SNV	c.1689G>T	-	7	18	190		
<i>ZNF765</i>	19q13.42	Synonymous SNV	c.978G>T	-	32	14	44		

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>RYR1</i>	19q13.2	Intron variant	c.7028-190_7028-146delGGGAGCCTGGTGTACCCCTAGAGGTGTTGGGTCCTGGGGCTGGC	-	56	21	25		
<i>SSC5D</i>	19q13.42	Inframe deletion	c.3855_3977del	p.Thr1286_Thr1326del	37	43	46		
<i>SEMG2</i>	20q13.12	Missense SNV	c.1195T>A	p.Tyr399Asn	13	17	98		
<i>BAGE2</i>	21p11.1	Non-coding exon variant	n.810G>A	-	18	11	565		
<i>KRTAP10-2</i>	21q22.3	Splice region & intron variant	n.366T>C	-	6	12	32		
<i>POM121L7P</i>	22q11.21	Missense SNV	c.365T>C	p.Ile122Thr	12	15	52	COSM4591810	Now reported as SNP
<i>TFIP11</i>	22q12.1	Splice region variant	n.561delT	-	28	16	19		
<i>UTY</i>	Yq11.221	Splice region & intron variant	c.1771-3delT	-	17	23	18		

9.7 WES VARIANTS IN CD49D+ CLONE ALONE (PATIENT 6)

Table 9-10. Variants present in the CD49d+ CLL clone alone as detected in WES of sorted fractions of patient 6. SNV = single nucleotide polymorphism; aa = amino acid; VAF = variant allele frequency; COSMIC = Catalogue of Somatic Mutations in Cancer (<https://cancer.sanger.ac.uk>); UTR = untranslated region; SNP = single nucleotide polymorphism.

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>FBLIM1</i>	1p36.21	Intron variant	c.250+86_250+89delCATT	-	19	0	22		
<i>SSX2IP</i>	1p22.3	3' UTR variant	c.*216G>A	-	70	0	10		
<i>TCHH</i>	1q21.3	Missense SNV	c.2923G>A	p.Gly975Arg	40	0	212	COSM896002	Endometrial ca (1)
<i>IGFN1</i>	1q32.1	Missense SNV	c.6043G>A	p.Gly2015Ser	9	0	78		
<i>IGFN1</i>	1q32.1	Synonymous SNV	c.6048C>T	p.Phe2016Phe	9	0	72		
<i>TARBP1</i>	1q42.2	Splice region & intron variant	c.1249-1G>T	-	47	0	37		
<i>NTSR2</i>	2q13	Missense SNV	c.886G>A	p.Val296Ile	23	0	46		
<i>RGPD8</i>	2q21.1	Splice region & intron variant	c.2603-5_2603-4delTT	-	10	0	76		
<i>HOXD12</i>	2q21.1	Missense SNV	c.433G>T	p.Ala145Ser	47	0	57		
<i>TTN</i>	2q31.1	Missense SNV	c.38975A>T	p.Lys12992Ile	17	0	140		
<i>STAT1</i>	2q31.2	Intron variant	c.2239-324T>G	-	43	0	8		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>EDEM1</i>	2q32.2	Intron variant	c.859-30G>C	-	40	0	47		
<i>CSPG5</i>	3p21.31	Missense SNV	c.1493C>T	p.Ser498Phe	38	0	49		
<i>FEZF2</i>	3p14.2	3' UTR variant	c.*91dupA	-	19	0	42		
<i>ALDH1L1</i>	3q21.3	Nonsense SNV	c.2143G>T	p.Glu715*	56	0	40		
<i>PLSCR2</i>	3q24	Splice region & intron variant	c.-83-3delT	-	27	0	31		
<i>PFN2</i>	3q25.1	Intron variant	c.133-31_133-28delGTTT	-	16	0	88		
<i>DGKG</i>	3q27.3	3' UTR variant	c.*104_*105dupCA	-	9	0	27		
<i>FETUB</i>	3q27.3	Frameshift variant	c.437delA	p.Lys146fs	16	0	45	COSM1238889	5 cancers, large intestine (18) most frequent
<i>MB21D2</i>	3q29	Intron variant	c.211+43_211+44delTG	-	8	0	36		
<i>RPL9</i>	4p14	Missense SNV	c.32G>A	p.Arg11Gln	20	0	84		
<i>UBE2K</i>	4p14	Intron variant	c.63+187T>G	-	37	0	13		
<i>EXOC1</i>	4q12	Intron variant	c.1074+901C>A	-	26	0	48		
<i>POLR2B</i>	4q12	Intron variant	c.1098-15_1098-14insC	-	25	0	70		
<i>LARP1BP1</i>	4q13.1	Non-coding exon variant	n.327A>T	-	20	0	46		
<i>DSPP</i>	4q22.1	Inframe deletion	c.2469_2573del	p.Ser823_Asn857del	24	0	207		
<i>FREM3</i>	4q31.21	Splice region & intron	c.5186-3delT	-	13	0	38		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		variant							
<i>TRIM2</i>	4q31.3	Intron variant	c.454-1216_454-1214delATC	-	13	0	37		
<i>DCHS2</i>	4q31.3	Missense SNV	c.4367C>A	p.Pro1456Gln	21	0	37		
<i>IRX1</i>	5p15.33	Intron variant	c.277-60_277-59delTC	-	12	0	56		
<i>MYO10</i>	5p15.1	Synonymous SNV	c.1998G>C	-	48	0	39		
<i>PRDM9</i>	5p14.2	Inframe deletion	c.2593_2676del TGTGGGCGGGG CTTAGCGATAG GTCAAGCCTCTG CTATCACCAGAG GACACACACAG GGGAGAAGCCC TACGTCTGCAGG GAG	p.Cys865_Glu892del	14	0	81		
<i>MAST4</i>	5q12.3	Intron variant	c.643-29355T>G	-	27	0	18		
<i>ZMAT2</i>	5q31.3	3' UTR variant	c.*132T>G	-	42	0	21		
<i>TENM2</i>	5q34	Intron variant	c.1309+241A>G	-	67	0	20		
<i>CNOT6</i>	5q35.3	3' UTR variant	c.*175T>A	-	46	0	21		
<i>TRIM39</i>	6p22.1	Intron variant	c.1010-217G>A	-	16	0	41		
<i>HLA-DRB1</i>	6p21.32	Missense SNV	c.40G>C	p.Ala14Pro	11	0	169		
<i>TFAP2B</i>	6p12.3	Intron variant	c.82-967A>T	-	45	0	18		
<i>IL17F</i>	6p12.2	Missense SNV	c.137G>T	p.Ser46Ile	57	0	37		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		Splice region & intron variant							
<i>L3MBTL3</i>	6q23.1	Splice region & intron variant	c.1408-8delT	-	13	0	40		
<i>PEX7</i>	6q23.3	Splice region & intron variant	c.748-4delT	-	19	0	16		
<i>ARID1B</i>	6q25.3	Inframe deletion	c.360_362delGCA	p.Gln121del	10	0	68		
<i>STK31</i>	7p15.3	Intron variant	c.483+223T>G	-	71	0	8		
<i>HNRNPA2B1</i>	7p15.2	Intron variant	c.*22-24delT	-	12	0	40		
<i>MUC17</i>	7q22.1	Missense SNV	c.7282G>A	p.Val2428Ile	22	0	130	COSM3784028	4 separate cancers including AML (1)
<i>CCDC136</i>	7q32.1	Missense SNV	c.694C>T	p.Arg232Cys	45	0	50		
<i>UBE2H</i>	7q32.2	5' UTR variant	c.-186A>C	-	29	0	14		
<i>RP1L1</i>	8p23.1	Missense SNV	c.6237C>G	p.His2079Gln	50	0	156		
<i>EGR3</i>	8p21.3	Missense SNV	c.955C>A	p.His319Asn	28	0	63		
<i>NECAB1</i>	8q21.3	3' UTR variant	c.*3472T>G	-	36	0	36		
<i>PABPC1</i>	8q22.3	3' UTR variant	c.*121A>G	-	25	0	6		
<i>SH3GL2</i>	9p22.2	5' UTR variant	c.-92C>A	-	45	0	58		
<i>CNTFR</i>	9p13.3	3' UTR variant	c.*155T>G	-	23	0	14		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>PAX5</i>	9p13.2	Intron variant	c.47-5725A>C	-	46	0	49		Mutations in coding sequence reported in lymphoid malignancy including CLL (3)
<i>PRUNE2</i>	9q21.2	Intron variant	c.8728+2847_8728+2848delGT	-	27	0	25		
<i>SLC34A3</i>	9q34.3	Intron variant	c.925+20_926-48del	-	18	0	350		
<i>ZEB1</i>	10p11.22	Inframe deletion	c.3210_3212delGGA	p.Glu1071del	10	0	40		
<i>NRP1</i>	10p11.22	Missense SNV	c.2334G>T	p.Gln778His	55	0	44		
<i>DLG5</i>	10q22.3	Intron variant	c.865-142G>C	-	47	0	22		
<i>LZTS2</i>	10q24.31	Missense SNV	c.871G>T	p.Gly291Trp	47	0	45		
<i>BTBD16</i>	10q26.13	Intron variant	c.242-22C>T	-	49	0	94		
<i>JAKMIP3</i>	10q26.3	Missense SNV	c.976C>G	p.Arg326Gly	40	0	52		
<i>CAPRIN1</i>	11p13	Synonymous SNV	c.84G>T	-	18	0	56		
<i>NRXN2</i>	11q13.1	Frameshift variant	c.30delG	p.Cys11fs	15	0	35		
<i>RBM14</i>	11q13.2	Intron variant	c.1803-237A>C	-	30	0	7		
<i>PPFIA1</i>	11q13.3	Intron variant	c.606+40delG	-	29	0	42		
<i>KRTAP5-10</i>	11q13.4	Inframe deletion	c.48_68delTTGTGGCTCCGGCTGTGGGGG	p.Cys17_Gly23del	87	0	36		
<i>RNF121</i>	11q13.4	Nonsense	c.910C>T	p.Arg304*	52	0	49		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		SNV							
<i>SORL1</i>	11q24.1	Synonymous SNV	c.4596G>A	-	46	0	36		
<i>PEX5</i>	12p13.31	Intron variant	c.192+32_192+76delCAGCCTCTG AGGCAGTGAGT GTTCTTGAGGTG GAAAGCCCAGG TG	-	100	0	98		
<i>M6PR</i>	12p13.31	Synonymous SNV	c.423G>C	-	29	0	39		
<i>BHLHE41</i>	12p12.1	Synonymous SNV	c.900G>C	-	5	0	57		
<i>FKBP11</i>	12q13.12	Splice region & intron variant	n.4957delA	-	71	0	7		
<i>KMT2D</i>	12q13.12	Nonsense SNV	c.15256C>T	p.Arg5086*	67	0	51	COSM5916243	Lymphoid neoplasm (1) & skin ca (2)
<i>SMARCC2</i>	12q13.2	Inframe deletion	c.3274_3276del CCT	p.Pro1092del	13	0	45	COSM1363038	3 cancers including large intestine (7)
<i>WIF1</i>	12q14.3	Intron variant	c.827-142delA	-	20	0	29		
<i>LEMD3</i>	12q14.3	3' UTR variant	c.*83T>G	-	32	0	40		
<i>POC1B</i>	12q21.33	Intron variant	c.1332+207_1332+208delAC	-	25	0	12		
<i>LUM</i>	12q21.33	Intron variant	c.-21-74A>T	-	20	0	39		
<i>MYO1H</i>	12q24.11	Intron variant	c.242+131_242+142delGCCCCACC CATCC	-	43	0	10		
<i>WSB2</i>	12q24.23	Intron variant	c.1053-241_1053-	-	50	0	2		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
			236delTGTGCC						
<i>ANHX</i>	12q24.33	Intron variant	c.839+1002_839+1003delGT	-	67	0	16		
<i>PSPC1</i>	13q12.11	3' UTR variant	c.*254C>T	-	60	0	8		
<i>SPART</i>	13q13.3	Intron variant	c.1164+210T>C	-	53	0	23		
<i>MYCBP2</i>	13q22.3	Intron variant	c.2268-18_2268-15delTTTT	-	100	0	13		
<i>SLC7A7</i>	14q11.2	Missense SNV	c.833A>G	p.Asn278Ser	42	0	45		
<i>MYH7</i>	14q11.2	Intron variant	c.4645-170T>A	-	36	0	20		
<i>NKX2-1</i>	14q13.3	Intron variant	c.78-39C>T	-	56	0	66		
IGHV3-21	14q32.33	Synonymous SNV	c.333T>C	-	18	0	77	IGH locus - known area of somatic hypermutation in B cells	
IGHV4-34	14q32.33	Missense SNV	c.326G>C	p.Ser109Thr	10	0	56		
IGHV4-34	14q32.33	Missense SNV	c.268A>C	p.Lys90Gln	28	0	56		
IGHV4-34	14q32.33	Synonymous SNV	c.201A>G	-	52	0	56	IGH locus - known area of somatic hypermutation in B cells	
IGHV4-34	14q32.33	Missense SNV	c.182G>A	p.Ser61Asn	51	0	56		
IGHV4-34	14q32.33	Missense SNV	c.170G>A	p.Gly57Asp	49	0	56		
IGHV4-34	14q32.33	Synonymous SNV	c.159G>A	-	51	0	56		
IGHV4-34	14q32.33	Missense SNV	c.158G>A	p.Gly53Glu	51	0	56		
<i>BCL11B</i>	14q32.2	Missense SNV	c.1387G>A	p.Ala463Thr	53	0	242	Different mutations in this gene reported in	

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>BCL11B</i>	14q32.2	5' UTR variant	c.-99_-97delGGC	-	37	0	41		CLL (8); associated with ALL
<i>HERC2P9</i>	15q13.1	Non-coding exon variant	n.327G>C	-	27	0	11		
<i>UBR1</i>	15q15.2	Intron variant	c.3998-11delT	-	17	0	39		
<i>NMNAT1 P5</i>	15q22.2	Splice region & intron variant	n.186-7delA	-	40	0	10		
<i>KIF23</i>	15q23	Splice region & intron variant	c.564-10_564-7delCTTT	-	10	0	93		
<i>SYNM</i>	15q26.3	Synonymous SNV	c.2985C>T	-	49	0	235		
<i>WASH3P</i>	15q26.3	Splice region & intron variant	n.1220-7G>A	-	50	0	8	COSM3932019	Liver ca (3) & bladder ca (1)
<i>CY5B</i>	16q22.1	Intron variant	c.333+230delT	-	33	0	10		
<i>NCOR1</i>	17p11.2	Intron variant	c.790-92_790-91insAAAAA	-	19	0	25		
<i>RHOT1</i>	17q11.2	Intron variant	c.1836-75delT	-	12	0	53		
<i>TBC1D3F</i>	17q12	5' UTR variant (premature start codon)	c.-803C>T	-	16	0	44		
<i>STAT3</i>	17q21.2	Intron variant	c.2145-57delT	-	57	0	7		Mutations in coding sequence reported in lymphoid malignancy (476); associated with T-LGL

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>ADAM11</i>	17q21.31	Intron variant	c.1485+70T>G	-	50	0	14		
<i>PTPRM</i>	18p11.23	Intron variant	c.2167+7291G>T	-	25	0	38		
<i>PHLPP1</i>	18q21.33	Frameshift variant	c.77_78insTCTGG	p.Ala27fs	19	0	17		
<i>GIPC3</i>	19p13.3	3' UTR variant	c.*848_*973del	-	24	0	76		
<i>LRRC8E</i>	19p13.2	3' UTR variant	c.*160T>G	-	37	0	16		
<i>IQCN</i>	19p13.11	Synonymous SNV	c.813C>T	-	51	0	144		
<i>ZNF714</i>	19p12	Synonymous SNV	c.1236C>T	-	33	0	38		
<i>ZNF208</i>	19p12	Missense SNV	c.745G>T	p.Val249Leu	11	0	110		
<i>ZNF254</i>	19p12	Missense SNV	c.1361G>T	p.Arg454Ile	24	0	49	COSM994266	Endometrial ca (1)
<i>CATSPERG</i>	19q13.2	Splice region & intron variant	c.670-7T>C	-	42	0	41		
<i>HNRNPL</i>	19q13.2	3' UTR variant	c.*260delT	-	60	0	6		
<i>PAK4</i>	19q13.2	Intron variant	c.-322+119T>G	-	32	0	37		
<i>PPFIA3</i>	19q13.33	Missense SNV	c.530C>T	p.Ala177Val	23	0	88		
<i>ZNF304</i>	19q13.43	Missense SNV	c.1642T>A	p.Leu548Met	24	0	42		
<i>LPIN3</i>	20q12	Intron variant	c.558-102C>A	-	45	0	41		
<i>SALL4</i>	20q13.2	Missense SNV	c.986G>A	p.Arg329His	47	0	234	COSM2763118	Lung ca (1)
<i>TFAP2C</i>	20q13.31	Intron variant	c.49-798delG		13	0	45		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of	
<i>TPTE</i>	21p11.1	Missense SNV	c.629A>T	p.His210Leu	18	0	204			
<i>RRP1B</i>	22q22.3	Intron variant	c.358-26_358-24delTTT	-	27	0	30			
<i>COL6A1</i>	22q22.3	Frameshift variant	c.1568_1575+1delGCTTCCCCG	p.Phe524fs	41	0	145			
<i>ZNF74</i>	22q11.21	Missense SNV	c.1667T>G	p.Phe556Cys	50	0	99			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.79C>T	p.Pro27Ser	45	0	101		IGL locus - known area of somatic hypermutation in B cells	
<i>IGLV3-1</i>	22q11.22	Synonymous SNV	c.147A>G	-	57	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.149A>C	p.Tyr50Ser	55	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.155G>C	p.Cys52Ser	41	0	59			
<i>IGLV3-1</i>	22q11.22	Nonsense SNV	c.163C>T	p.Gln55*	41	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.165G>C	p.Gln55His	39	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.172C>G	p.Pro58Ala	59	0	59			
<i>IGLV3-1</i>	22q11.22	Synonymous SNV	c.174A>G	-	39	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.209G>C	p.Ser70Thr	42	0	59			
<i>IGLV3-1</i>	22q11.22	Synonymous SNV	c.210C>T	-	41	0	59			
<i>IGLV3-1</i>	22q11.22	Missense SNV	c.212A>G	p.Lys71Arg	42	0	59			
<i>IGLV3-1</i>	22q11.22	Synonymous SNV	c.234G>A	-	47	0	59			
<i>TRIOBP</i>	22q13.1	Inframe deletion	c.2060_2203del	p.Pro687_Ser734del	66	0	69			

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>GRAMD4</i>	22q13.31	Intron variant	c.-49-144_-49-36del	-	11	0	15		
<i>PDHA1</i>	Xp22.12	Synonymous SNV	c.69C>G	-	35	0	170		
<i>DDX3X</i>	Xp11.4	3' UTR variant	c.*137dupT	-	15	0	67		Different mutations in this gene reported in CLL (12)
<i>CCNB3</i>	Xp11.22	Missense SNV	c.2497T>A	p.Leu833Met	61	0	38		
<i>SSXP1</i>	Xp11.22	Splice region & intron variant	n.70-8dupT	-	25	0	27		
<i>ARHGEF9</i>	Xq11.2	5' UTR variant	c.-97dupA	-	45	0	20		

9.8 WES VARIANTS IN CD49D- CLONE ALONE (PATIENT 6)

Table 9-11. Variants present in the CD49d- CLL clone alone as detected in WES of sorted fractions of patient 6. SNV = single nucleotide polymorphism; aa = amino acid; VAF = variant allele frequency; COSMIC = Catalogue of Somatic Mutations in Cancer (<https://cancer.sanger.ac.uk>); UTR = untranslated region; SNP = single nucleotide polymorphism.

Gene	Chromosome band	Mutation			VAF %		Median depth	COSMIC ID	COSMIC type of cancer (no. of samples); other
		Type	Change	Predicted aa change	CD49d positive clone	CD49d negative clone			
<i>TMEM51</i>	1p36.21	Synonymous SNV	c.45C>T	-	0	13	52		
<i>ZMYM4</i>	1p34.3	Missense SNV	c.182G>T	p.Gly61Val	0	31	39		
<i>RNPC3</i>	1p21.1	Frameshift variant	c.357_358delA A	p.Arg120fs	0	25	20	COSM4610875	Large intestine ca (12) & lung ca (1)
<i>HRNR</i>	1q21.3	Frameshift variant	c.8291_8292dup pCT	p.Gly2765fs	0	9	111		
<i>FLG</i>	1q21.3	Missense SNV	c.10648G>C	p.Glu3550Gln	0	13	235		
<i>KHDC4</i>	1q22	Frameshift variant	c.1546_1547del AG	p.Arg516fs	0	9	39	COSM265410	5 cancers, large intestine ca (4) & melanoma (4) most frequent
<i>RFWD2</i>	1q25.2	Intron variant	c.1531-64dupT	-	0	28	39		
<i>LHX9</i>	1q31.3	Missense SNV	c.806G>A	p.Arg269His	0	35	60	COSM2125830	Large intestine ca (3) & liver ca (1)
<i>TGFB2</i>	1q41	Missense SNV	c.1001C>T	p.Ala334Val	0	14	40	COSM5006673	Large intestine ca (1)
<i>ZFP36L2</i>	2p21	3' UTR variant	c.*213delT	-	0	16	31		
<i>ANKRD39</i>	2q11.2	Intron variant	c.100+112_100 +118dupTTGGG GC	-	0	67	13		
<i>LIPT1</i>	2q11.2	Splice region	c.*78delA	-	0	12	61		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		variant							
<i>TTN</i>	2q31.2	Synonymous SNV	c.39087A>C	-	0	15	150		
<i>SF3B1</i>	2q33.1	Missense SNV	c.1866G>T	p.Glu622Asp	0	49	129	COSM110693	CLL (11) & MDS (22); recurrent mutation in CLL
<i>ABCA12</i>	2q35	Intron variant	c.3625-258T>C	-	0	60	5		
<i>SCLY</i>	2q37.3	3' UTR variant	c.*213_*254del TGCCCACATGG GACCGCCCACA TAGGACCGCCC ACATAGGAC	-	0	100	14		
<i>SS18L2</i>	3p22.1	Missense SNV	c.44A>C	p.Glu15Ala	0	16	50		
<i>SETD2</i>	3p21.31	Splice region & intron variant	c.4587-1G>A	-	0	36	41		
<i>GPX1</i>	3p21.31	5' UTR variant (premature start codon)	c.-103G>T	-	0	57	22		
<i>GRK7</i>	3q23	3' UTR variant	c.*64C>A	-	0	11	44		
<i>MBNL1</i>	3q25.1	Intron variant	c.861+64delT	-	0	14	38		
<i>P3H2</i>	3q28	5' UTR variant	c.-59C>G	-	0	31	48		
<i>GBA3</i>	4p15.2	Splice region & intron variant	n.1084-3dupT	-	0	12	50		
<i>PTPN13</i>	4q21.3	Intron variant	c.4345-137delT	-	0	15	23		
<i>DUX4L4</i>	4q35.2	Missense SNV	c.1130C>A	p.Pro377His	0	16	36	COSM4593687	Head & neck ca (2)
<i>DUX4L4</i>	4q35.2	Missense SNV	c.1153G>A	p.Glu385Lys	0	18	38		
<i>DUX4L4</i>	4q35.2	Missense SNV	c.1163C>A	p.Ala388Asp	0	19	37		
<i>SLC6A19</i>	5p15.33	Inframe deletion	c.664-223_666del	p.Ala222_Val223del	0	13	27		
<i>DMGDH</i>	5q14.1	Synonymous SNV	c.69C>T	-	0	31	41		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>DMXL1</i>	5q23.1	Intron variant	c.88-133delT	-	0	20	20		
<i>CATSPER3</i>	5q31.1	Missense SNV	c.619C>T	p.His207Tyr	0	17	52		
<i>GRIA1</i>	5q33.2	Intron variant	c.112+163T>G	-	0	41	36		
<i>LARP1</i>	5q33.2	Intron variant	c.205+4761_205+4762delAC	-	0	11	35		
<i>ATP10B</i>	5q34	Missense SNV	c.1339C>T	p.Arg447Cys	0	40	44	COSM250767	Liver ca (2) & angiosarcoma (1)
<i>KIF6</i>	6p21.2	Missense SNV	c.509C>T	p.Pro170Leu	0	12	43	COSM5007484	Large intestine ca (1)
<i>BICRAL</i>	6p21.1	Missense SNV	c.600T>A	p.His200Gln	0	23	57		
<i>LRRC73</i>	6p21.1	5' UTR variant	c.-83A>C	-	0	37	54		
<i>PDE7B</i>	6q23.3	Intron variant	c.83-70484_83-70482delGCA	-	0	12	40		
<i>PPP1R14C</i>	6q25.1	Intron variant	c.306+168T>G	-	0	24	29		
<i>DPY19L2P3</i>	7p14.3	Splice region & intron variant	n.612-2delA	-	0	33	23		
<i>NACAD</i>	7p13	Missense SNV	c.2714C>T	p.Pro905Leu	0	29	166	COSM3718680	Now reported as SNP
<i>RUNDC3B</i>	7q21.12	Splice region & intron variant	c.239-5delT	-	0	22	32		
<i>MUC12</i>	7q22.1	Missense SNV	c.5883T>A	p.His1961Gln	0	41	1032		
<i>LMOD2</i>	7q31.32	Inframe deletion	c.1284_1343del TCCTCCTCCCC TCCTTCTCCCA AAGGCTGCCAC CACCTCCTCCTC CTCCCCCTCCTC C	p.Pro429_Pro448del	0	21	109		
<i>DGKI</i>	7q33	Missense SNV	c.2665C>T	p.Arg889Trp	0	36	39		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
<i>PEBP4</i>	8p21.3	3' UTR variant	c.*109delA	-	0	15	27		
<i>CYP7B1</i>	8q12.3	Synonymous SNV	c.105G>A	-	0	34	45		
<i>MYBL1</i>	8q13.1	Intron variant	c.292-118delT	-	0	14	22		
<i>VCIPI1</i>	8q13.1	Frameshift variant	c.361_373delG GCCTTCCAAC	p.Gly121fs	0	36	74		
<i>UBE2W</i>	8q21.11	Intron variant	c.103-84C>A	-	0	21	28		
<i>CNTNAP3</i>	9p13.1	Missense SNV	c.3640C>T	p.Arg1214Trp	0	11	63	COSM4592383	Now reported as SNP
<i>CNTNAP3</i>	9p13.1	Synonymous SNV	c.3639G>T	-	0	11	63	COSM5094831	Thymoma (1)
<i>CNTNAP3</i>	9p13.1	Synonymous SNV	c.3627C>T	-	0	11	64	COSM5956705	Now reported as SNP
<i>TRIM14</i>	9q22.33	Synonymous SNV	c.1137C>T	-	0	11	48		
<i>ZNF618</i>	9q32	Missense SNV	c.40G>A	p.Gly14Arg	0	14	51		
<i>USP6NL</i>	10p14	Splice region & intron variant	c.74-4delT	-	0	37	10		
<i>ARHGAP12</i>	10p11.22	3' UTR variant	c.*120dupT	-	0	12	30		
<i>ZNF487</i>	10q11.21	Missense SNV	c.1067C>A	p.Ala356Glu	0	24	46		
<i>ZNF487</i>	10q11.21	Missense SNV	c.1082G>T	p.Gly361Val	0	25	39		
<i>KCNIP2</i>	10q24.32	Intron variant	c.169+150T>G	-	0	31	13		
<i>VTI1A</i>	10q25.2	Intron variant	c.343-94T>G	-	0	25	38		
<i>KNDC1</i>	10q26.3	Nonsense SNV	p.Gln2*/c.4C>T	-	0	18	59		
<i>CDHR5</i>	11p15.5	Inframe deletion	c.1602_1787del	p.Pro535_Thr596del	0	22	118		
<i>USH1C</i>	11p15.1	Synonymous SNV	c.1110T>C	-	0	37	45		
<i>ANO3</i>	11p14.2	Intron variant	c.2658-33_2658-	-	0	13	112		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
			30delATTT						
<i>SCYL1</i>	11q13.1	Splice region & intron variant	c.252+1G>A	-	0	24	47		
<i>ARAP1</i>	11q13.4	Intron variant	c.2167+186C>A	-	0	12	24		
<i>PPME1</i>	11q13.4	Intron variant	c.964+225T>G	-	0	37	8		
<i>PAK1</i>	11q14.1	Intron variant	c.1116+183delT	-	0	16	19		
<i>CWC15</i>	11q21	Intron variant	c.442-75delT	-	0	23	25		
<i>TRAPPC4</i>	11q23.3	5' UTR variant	c.-97_-95delAGG	-	0	23	49		
<i>A2ML1</i>	12p13.31	Intron variant	c.4061+79_4061+80delGT	-	0	14	58		
<i>KRT128P</i>	12q13.13	Non-coding exon variant	n.443_451delG GCAGCTAG	-	0	60	5		
<i>METTL17</i>	14q11.2	Intron variant	c.528+110delT	-	0	20	51		
<i>NPAS3</i>	14q13.1	3' UTR variant	c.*105G>A	-	0	16	45		
<i>LINC01599</i>	14q21.3	Missense SNV	c.319G>C	p.Glu107Gln	0	51	38		
<i>EXOC5</i>	14q22.3	Splice region & intron variant	c.-139+2T>G	-	0	50	6		
<i>PPM1A</i>	14q23.1	Intron variant	c.1172-132delT	-	0	36	25		
<i>PCNX1</i>	14q24.2	Splice region & intron variant	c.4853-5_4853-4delCT	-	0	9	50		
<i>TRIP11</i>	14q32.12	Intron variant	c.4699-72T>A	-	0	55	42		Mutations in coding sequence reported in lymphoid malignancy including CLL (1)

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
IGHJ6	14q32.33	Synonymous SNV	c.12C>T	-	0	49	132		Known area of somatic hypermutation
IGHV3-21	14q32.33	Missense SNV	c.149G>C	p.Ser50Thr	0	46	77		Known area of somatic hypermutation
<i>CCPG1</i>	15q21.3	Intron variant	c.-9-43_-9-20delAAAAAAAA TGTGTCTCATG AGTCAA	-	0	15	44		
<i>CYP11A1</i>	15q24.1	Intron variant	c.270-103_270-82delTCCCACA GTCAGCAAGTT GTGG	-	0	39	46		
<i>IL16</i>	15q25.1	Inframe deletion	c.2512_2514del TCC	p.Ser838del	0	8	60	COSM293750	4 cancers, large intestine most frequent (2)
<i>C16orf45</i>	16p13.11	Intron variant	c.106+93T>G	-	0	31	15		
<i>PAPD5</i>	16q12.1	Missense SNV	c.1088A>G	p.Tyr363Cys	0	17	48	COSM3937079	Oesophageal ca (1)
<i>CDH8</i>	16q21	Intron variant	c.1655-38G>A	-	0	34	77		
<i>CALB2</i>	16q22.2	Intron variant	c.400-92_400-91delCA	-	0	10	36		
<i>VPS53</i>	17p13.3	Splice region & intron variant	n.732delA	-	0	31	25		
<i>CTDNEP1</i>	17p13.1	3' UTR variant	c.*434T>G	-	0	21	21		
<i>TBC1D26</i>	17p12	Splice region & intron variant	n.-281T>C	-	0	33	12		
<i>TVP23B</i>	17p11.2	5' UTR variant	c.-196C>T	-	0	22	19		
<i>TVP23B</i>	17p11.2	5' UTR variant (premature start codon)	c.-184C>T	-	0	21	19		
<i>KRTAP4-7</i>	17q21.2	Synonymous	c.183A>G	-	0	12	117		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		SNV							
<i>KRTAP4-7</i>	17q21.2	Synonymous SNV	c.186C>A	-	0	12	112		
<i>KRTAP4-5</i>	17q21.2	Inframe deletion	c.206_220delA TTGCTGTGAAT CCA	p.Tyr69_Ser74del nsCys	0	18	116	COSM1383091	3 cancers, large intestine ca most frequent (9)
<i>KAT2A</i>	17q21.2	Inframe deletion	c.1181- 110_1186del	p.Ala394_Ser395d el	0	27	55		
<i>ARHGAP 27P1- BPTFP1- KPNA2P3</i>	17q24.1	Non-coding exon variant	n.2240C>A	-	0	25	51		
<i>TNRC6C</i>	17q25.3	Splice region & intron variant	c.-625-4delT	-	0	50	15		
<i>DSG3</i>	18q12.1	Splice region & intron variant	c.84+1_84+3del GTA	-	0	12	43		
<i>CXXC1</i>	18q21.1	Inframe deletion	c.481_483delC AG	p.Gln161del	0	11	47	COSM5854235	Melanoma (5) & nasopharyngeal ca (1)
<i>MIER2</i>	19p13.3	Intron variant	c.889+12_889+ 68delTCTGGGC CTTCCTCCGCT GCGGCCCGCC CTGGGCACTGC TGACCGTTCTCC CCGG	-	0	27	54		
<i>CIRBP</i>	19p13.3	3' UTR variant	c.*215dupT	-	0	15	20		
<i>PTPRS</i>	19p13.3	Intron variant	c.707-141delT	-	0	44	21		
<i>KHSRP</i>	19p13.3	3' UTR variant	c.*344_345del AT	-	0	9	103		
<i>CLEC4G</i>	19p13.2	Splice region & intron	c.9-4T>C	-	0	70	8		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
		variant							
<i>ELAVL3</i>	19p13.2	Intron variant	c.9+90delG	-	0	25	68		
<i>ABHD8</i>	19p13.11	Intron variant	c.-9+182T>G	-	0	36	10		
<i>ZNF66</i>	19p12	Frameshift variant	c.285_286dupT T	p.Ser96fs	0	27	33		
<i>ZNF43</i>	19p12	Synonymous SNV	c.1932C>T	-	0	27	17		
<i>ZNF724</i>	19p12	Missense SNV	c.400A>T	p.Asn134Tyr	0	14	194		
<i>ZNF726</i>	19p12	Missense SNV	c.1549G>T	p.Ala517Ser	0	17	42		
<i>TMEM145</i>	19q13.2	Frameshift variant (stop codon lost)	c.1462_1465del TTTT	p.Phe488fs	0	100	7		
<i>SYT3</i>	19q13.33	Intron variant	c.1402+90_1402+163delTGAG GGAGGAGGAG CTGGGGATCTG GACTCCTGGGT CTGAGGGAGG AGGGGCCAGG GGCCTGGACTC CTGGGTC	-	0	80	63		
<i>FRG1BP</i>	20q11.21	Missense SNV	c.541C>A	p.Pro181Thr	0	12	25	COSM1025524	Now reported as SNP
<i>RBM39</i>	20q11.22	Splice region & intron variant	n.704dupA	-	0	30	10		
<i>MICAL3</i>	22q11.21	Inframe deletion	c.2994_2996del GGA	p.Glu999del	0	30	20	COSM1414806	Large intestine ca (1)
<i>IGLV4-60</i>	22q11.22	Synonymous SNV	c.198G>C	-	0	34	178		IGL locus - known area of somatic hypermutation in B cells
<i>IGLV4-60</i>	22q11.22	Missense SNV	c.217G>C	p.Gly73Arg	0	32	178		
<i>IGLV4-60</i>	22q11.22	Missense SNV	c.220A>G	p.Ser74Gly	0	32	178		
<i>IGLV4-60</i>	22q11.22	Missense SNV	c.221G>A	p.Ser74Asn	0	32	178		

Gene	Chromosome	Mutation			VAF %		Median	COSMIC ID	COSMIC type of
IGLV4-60	22q11.22	Missense SNV	c.227G>C	p.Ser76Thr	0	32	178		
IGLV4-60	22q11.22	Missense SNV	c.272G>A	p.Ser91Asn	0	25	161		
IGLV4-60	22q11.22	Missense SNV	c.275C>T	p.Ser92Phe	0	25	196		
IGLL5	22q11.22	Missense SNV	c.108T>G	p.His36Gln	0	98	45		
IGLL5	22q11.22	Synonymous SNV	c.174C>T	-	0	93	75	COSM5948183	
<i>MTMR3</i>	22q12.2	Intron variant	c.558-62_558-61delTC	-	0	12	44		
<i>TRIOBP</i>	22q13.1	Synonymous SNV	c.3219G>C	-	0	14	55		
<i>RANGAP1</i>	22q13.2	Inframe deletion	c.1101_1103delGGA	p.Glu368del	0	11	77	COSM5089452	Melanoma (5) & nasopharyngeal ca (1)
<i>FBLN1</i>	22q13.31	Intron variant	c.185+276_185+277insCTCTGC CACAGCTGGTC AGCTG	-	0	60	12		
<i>SHANK3</i>	22q13.33	Missense SNV	c.2854G>A	p.Ala952Thr	0	52	48		
<i>TIMM17B</i>	Xp11.23	3' UTR variant	c.*138T>G	-	0	33	25		
<i>ZNF75D</i>	Xq26.3	Intron variant	c.604+152T>G	-	0	33	22		
<i>MAP7D3</i>	Xq26.3	Frameshift variant	c.922delC	p.Gln308fs	0	12	47	COSM5831229	Large intestine (7) & melanoma (6)
<i>SLC25A15P1</i>	Yq11.23	Splice region & intron variant	n.55-8dupT	-	0	22	25		

9.9 CURATED GENE LISTS FROM RNASEQ (PATIENT 6)

Table 9-12. Genes downregulated 10 or more-fold (with absolute reads of >100) in the CD49d+ clone in patient 6

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>LOC100505841</i>	4	4023	-9.79
<i>LOC100128252</i>	0	265	-9.05
<i>MYLK</i>	2	838	-8.37
<i>SCN8A</i>	1	356	-7.88
<i>RAI2</i>	1	248	-7.36
<i>FGF2</i>	4	584	-7.00
<i>APOD</i>	20	2529	-6.93
<i>WNT5B</i>	1	177	-6.87
<i>PRSS1</i>	9	984	-6.68
<i>ANGPT2</i>	1	151	-6.65
<i>MOCS1</i>	1	148	-6.62
<i>STK32B</i>	11	1070	-6.52
<i>PLEKHG4B</i>	1	137	-6.51
<i>ZNF471</i>	5	494	-6.47
<i>CABLES1</i>	7	637	-6.39
<i>GTF2IRD1</i>	1	125	-6.38
<i>ZMAT1</i>	71	5736	-6.33
<i>AMOT</i>	2	201	-6.32
<i>PSD3</i>	7	606	-6.32
<i>LOC100130872</i>	12	997	-6.30
<i>DMD</i>	25	2028	-6.30
<i>CEACAM19</i>	2	187	-6.21
<i>JAM3</i>	29	2202	-6.20
<i>WWC1</i>	6	481	-6.19
<i>ZFP28</i>	6	462	-6.14
<i>SGSM1</i>	42	2919	-6.12
<i>RGS13</i>	22	1503	-6.04
<i>FRMD5</i>	2	160	-5.99
<i>SNCAIP</i>	6	409	-5.96
<i>B3GAT1</i>	4	282	-5.95
<i>SLC22A17</i>	7	445	-5.88
<i>DMRTA1</i>	3	204	-5.86
<i>KLHDC8B</i>	14	827	-5.82
<i>ABCB1</i>	71	3848	-5.75
<i>ZNF667</i>	15	836	-5.74
<i>CD9</i>	8	453	-5.72
<i>ZNF558</i>	23	1244	-5.71
<i>TEKT4P2</i>	7	395	-5.70

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>E2F7</i>	6	331	-5.66
<i>MAP6</i>	4	225	-5.63
<i>LAG3</i>	4	224	-5.63
<i>TOX2</i>	26	1304	-5.60
<i>KLRC4</i>	2	121	-5.59
<i>KLK1</i>	6	305	-5.54
<i>ZNF208</i>	18	869	-5.54
<i>IGF1R</i>	20	963	-5.54
<i>GRIK1-AS2</i>	13	628	-5.52
<i>HOXB2</i>	19	891	-5.50
<i>FGFR1</i>	9	425	-5.47
<i>PDZD2</i>	24	1073	-5.44
<i>SMARCA1</i>	3	152	-5.43
<i>DMKN</i>	17	749	-5.40
<i>IL17RB</i>	3	148	-5.39
<i>DLGAP1</i>	8	360	-5.39
<i>C8orf31</i>	2	105	-5.38
<i>GRIK1-AS1</i>	3	146	-5.37
<i>SLC29A4</i>	2	101	-5.33
<i>PHF16</i>	105	4190	-5.31
<i>ARHGAP6</i>	7	285	-5.24
<i>PVR</i>	16	580	-5.12
<i>CHDH</i>	16	570	-5.09
<i>PRSS16</i>	6	211	-5.01
<i>DUSP26</i>	4	145	-5.00
<i>CTLA4</i>	696	21941	-4.98
<i>PTPRS</i>	25	807	-4.97
<i>LILRA4</i>	30	944	-4.94
<i>BEX5</i>	3	106	-4.91
<i>MERTK</i>	8	249	-4.86
<i>KLK2</i>	28	820	-4.83
<i>UGGT2</i>	13	373	-4.77
<i>GTSF1L</i>	4	122	-4.76
<i>NAP1L2</i>	10	286	-4.76
<i>FMOD</i>	482	12568	-4.70
<i>LDLRAD2</i>	11	302	-4.70
<i>C17orf28</i>	10	272	-4.68
<i>LOC645638</i>	146	3724	-4.67

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>EPB41L1</i>	5	140	-4.66
<i>GRAMD1B</i>	16	409	-4.62
<i>CLDN7</i>	6	152	-4.54
<i>WNT3</i>	99	2310	-4.53
<i>FAM50B</i>	4	103	-4.50
<i>FER1L4</i>	45	995	-4.47
<i>CYP2C8</i>	10	229	-4.43
<i>CLEC4G</i>	9	203	-4.41
<i>ZNF876P</i>	67	1409	-4.39
<i>RGS7</i>	5	111	-4.32
<i>VWA5A</i>	20	409	-4.30
<i>ZNF219</i>	31	616	-4.27
<i>CECR2</i>	30	590	-4.26
<i>CTTN</i>	8	160	-4.22
<i>DAB2IP</i>	10	196	-4.21
<i>LINC00173</i>	26	493	-4.20
<i>SCML1</i>	61	1114	-4.19
<i>AP1S1</i>	42	741	-4.14
<i>LGR6</i>	21	374	-4.11
<i>CIDEB</i>	26	458	-4.10
<i>EMR4P</i>	799	13215	-4.05
<i>NRCAM</i>	22	371	-4.03
<i>TRPM4</i>	8	138	-4.01
<i>COL1A2</i>	15	251	-4.00
<i>EPPK1</i>	7	121	-4.00
<i>FAM84B</i>	11	185	-3.99
<i>HSPA12A</i>	7	119	-3.97
<i>TNFRSF13B</i>	133	2039	-3.94
<i>CYB5R2</i>	20	301	-3.86
<i>COL6A2</i>	11	164	-3.82
<i>PROX2</i>	7	101	-3.74
<i>FBXO27</i>	9	122	-3.68
<i>SSH3</i>	20	254	-3.62
<i>CACNA1D</i>	20	252	-3.60
<i>PRKG2</i>	10	128	-3.60
<i>CCL28</i>	10	127	-3.59
<i>GATA3</i>	18	222	-3.57
<i>PTMS</i>	32	379	-3.53
<i>PDGFD</i>	190	2182	-3.52
<i>MLLT3</i>	29	337	-3.50
<i>DTX3</i>	170	1867	-3.45
<i>KCTD15</i>	10	115	-3.44
<i>MTMR9LP</i>	47	489	-3.38
<i>KCNN3</i>	37	392	-3.37

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>RHOBTB2</i>	546	5596	-3.36
<i>FAM83H</i>	12	128	-3.35

Table 9-13. Genes upregulated 10 or more-fold (with absolute reads of >100) in the CD49d+ clone in patient 6

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>IL8</i>	1173	3	8.40
<i>HIST1H2BG</i>	160	0	8.33
<i>PTGS2</i>	388	1	8.03
<i>LILRA6</i>	101	0	7.67
<i>IL1B</i>	294	1	7.63
<i>GOS2</i>	425	2	7.43
<i>TNFAIP2</i>	1207	7	7.35
<i>PLAUR</i>	237	1	7.32
<i>CSTA</i>	226	1	7.25
<i>GAS7</i>	374	2	7.24
<i>S100A8</i>	984	8	6.87
<i>THBS1</i>	167	1	6.82
<i>DENND3</i>	1173	11	6.69
<i>ANXA1</i>	1111	11	6.61
<i>ITGA4</i>	6691	70	6.57
<i>CLEC2B</i>	132	1	6.47
<i>LARGE</i>	128	1	6.43
<i>CCR1</i>	283	3	6.35
<i>EPHA4</i>	2215	28	6.30
<i>HK3</i>	112	1	6.25
<i>LILRA2</i>	111	1	6.23
<i>CD4</i>	330	4	6.21
<i>CD36</i>	183	2	6.21
<i>EMP1</i>	178	2	6.17
<i>EMR2</i>	175	2	6.15
<i>PAG1</i>	103	1	6.12
<i>LYZ</i>	7718	111	6.12
<i>AHR</i>	442	6	6.11
<i>KCTD12</i>	232	3	6.07
<i>FCN1</i>	1922	30	6.00
<i>HCAR3</i>	153	2	5.95
<i>EBF1</i>	3437	57	5.89
<i>CLEC7A</i>	361	6	5.82
<i>CD300E</i>	415	7	5.81
<i>S100A9</i>	2299	41	5.77
<i>VCAN</i>	890	16	5.77
<i>HBEGF</i>	224	4	5.66
<i>ZNF608</i>	615	13	5.53
<i>SERPINA1</i>	1286	28	5.51
<i>CSF3R</i>	370	8	5.46
<i>NLRP3</i>	233	5	5.42
<i>SLC11A1</i>	219	5	5.33

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>GNB4</i>	250	6	5.29
<i>PTCH1</i>	706	18	5.27
<i>PTH2R</i>	610	16	5.23
<i>MAFB</i>	860	23	5.21
<i>TDRD1</i>	301	8	5.16
<i>SIRPA</i>	124	3	5.16
<i>ID2</i>	502	14	5.13
<i>POF1B</i>	1294	37	5.13
<i>LOC100507377</i>	223	6	5.12
<i>NFAM1</i>	118	3	5.10
<i>IGJ</i>	3708	108	5.10
<i>PCDH9</i>	4962	145	5.09
<i>S100A12</i>	116	3	5.08
<i>CLIP4</i>	614	18	5.07
<i>TLR4</i>	347	10	5.07
<i>APOBEC3A</i>	114	3	5.05
<i>SLC4A7</i>	2201	66	5.05
<i>PTGER2</i>	112	3	5.03
<i>CST3</i>	587	18	5.01
<i>IRAK3</i>	109	3	4.99
<i>HCK</i>	1228	39	4.98
<i>CPVL</i>	511	16	4.97
<i>VPS37B</i>	4726	155	4.92
<i>ETS2</i>	124	4	4.80
<i>HIST1H2BI</i>	175	6	4.77
<i>PHYHD1</i>	629	23	4.76
<i>BCAT1</i>	1002	37	4.76
<i>GPAT2</i>	462	17	4.74
<i>ZNF711</i>	3445	134	4.68
<i>FSIP2</i>	314	12	4.67
<i>GEM</i>	135	5	4.64
<i>TLR8</i>	231	9	4.62
<i>TRPS1</i>	1378	57	4.58
<i>PDPN</i>	497	21	4.55
<i>CYBB</i>	11255	484	4.54
<i>CEBPD</i>	126	5	4.53
<i>GPR146</i>	537	23	4.53
<i>EPB41L2</i>	442	19	4.52
<i>IL10</i>	125	5	4.52
<i>CSF1R</i>	101	4	4.51
<i>CORO2B</i>	1593	70	4.50
<i>DMXL2</i>	322	14	4.49

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>HES1</i>	241	11	4.41
<i>FCER1G</i>	383	18	4.39
<i>RGPD1</i>	900	43	4.36
<i>PRKCH</i>	207	10	4.32
<i>TBXAS1</i>	626	32	4.29
<i>CD14</i>	125	6	4.28
<i>PLXNB2</i>	589	31	4.24
<i>TSPAN13</i>	1406	74	4.24
<i>EVC</i>	341	18	4.22
<i>FGL2</i>	247	13	4.21
<i>NPAS2</i>	100	5	4.21
<i>KIAA1324L</i>	1682	91	4.20
<i>HCG11</i>	1598	90	4.14
<i>VASH1</i>	6444	368	4.13
<i>FAM49A</i>	1923	110	4.13
<i>LMO2</i>	163	9	4.12
<i>MOXD1</i>	258	15	4.08
<i>FYB</i>	173	10	4.06
<i>TRIB2</i>	1523	92	4.04
<i>EVC2</i>	739	44	4.04
<i>CCL3</i>	117	7	3.99
<i>USP32P1</i>	274	17	3.99
<i>LDLR</i>	147	9	3.97
<i>PAM</i>	208	13	3.96
<i>HIC1</i>	113	7	3.94
<i>GRASP</i>	5350	375	3.83
<i>SORL1</i>	159	11	3.81
<i>CMTM7</i>	186	13	3.81
<i>TREM1</i>	101	7	3.78
<i>LGSN</i>	101	7	3.78
<i>HLX</i>	675	50	3.73
<i>FCGR3A</i>	163	12	3.73
<i>LGALS3</i>	345	26	3.72
<i>ZNF280B</i>	136	10	3.71
<i>SMAD3</i>	1256	96	3.71
<i>SYNPO</i>	611	46	3.70
<i>PSTPIP2</i>	181	14	3.66
<i>MS4A6A</i>	228	18	3.64
<i>NRIP1</i>	1626	132	3.61
<i>ANPEP</i>	100	8	3.58
<i>SIK1</i>	22590	1895	3.57
<i>DOK2</i>	1434	121	3.57
<i>CEBPA</i>	119	10	3.53
<i>PXDC1</i>	106	9	3.51

Gene	Normalised reads (CD49d+ clone)	Normalised reads (CD49d- clone)	Log ₂ fold change
<i>NXP4</i>	1281	114	3.49
<i>C5AR1</i>	448	40	3.49
<i>FLJ13197</i>	424	38	3.48
<i>NR4A3</i>	1083	97	3.48
<i>NIPAL4</i>	251	23	3.44
<i>ARRB1</i>	159	15	3.38
<i>C2orf88</i>	426	40	3.38
<i>LOC100505746</i>	455	44	3.34
<i>HHAT</i>	571	56	3.33
<i>TBC1D8</i>	370	37	3.32

9.10 FLOW CYTOMETRY DATA

Table 9-14. Flow cytometry data from Chapters 3 and 6. Mean fluorescence intensity (MFI) and percentage positive cells of CD49d, TLR4, CD14, CD38, CD11b and CCR7 from a cohort of trisomy 12 and disomy 12 CLL. Asterisked samples* have a hypermutated IGHV. Grey boxes = data not collected. MFI numbers refer to the MFI of the CLL population of the stained cell aliquot minus the MFI of the lymphoid gate of a concurrent unstained sample. CD49d, TLR4 and CD14 were measured in 2020. CD38, CD11b and CCR7 were measured in 2017.

Patient ID	MFI (stained-unstained)						% positive cells			
	CD49d	TLR4	CD14	CD38	CD11b	CCR7	CD49d	TLR4	CD14	CD38
2	1214	1739	40	278	553		61	69	1	1
3				420	672	297				1
5*	7563	1255	22	389	1666	172	81	59	0	0
6	14180	196	107	520	174	204	81			10
6 (CD49d+)*	17763	189	106				N/A			
6 (CD49d-)	563	227	113				N/A			
7	9438	1285	-62	178	236	354	79	48	0	0
8				449	422	445				0
9				313	650	751				0
10	3576	1749	33	289	604	1051	87	81	0	0
11	4468	2310	40	469	481	322	81	80	0	6
12*	4219	2183	52	630	1575	601	94	96	0	1
14*	3343	1393	28	1206	2140	430	94	75	0	29
15	1997	1555	44	455	467	431	70	74	0	3
17				271	449	435				0
18	5979	949	29	496	637	275	89	59	2	2
19				333	451	374				0
20	1934	1168	70	333	503	404	61	55	1	0
21		1213	42	262	439	380		56	2	0
21 (CD49d+)	5707	1356	216				N/A			
21 (CD49d-)	155	1161	197				N/A			
22	3679	432	23	269	480	922	97	95	0	0
68	5515	480	26				100	91	0	
70*	5348	453	8				100	96	0	
71	580	1322	49				50	66	1	
72		1989	48					98	0	
73		1820	46					97	1	
DIS1	72	1131	37	110	193	531	25	72	0	0
DIS3*	141	1494	32	148	231	589	27	78	0	0
DIS4*	104	1365	49	317	274	223	26	63	1	1
DIS5	5341	1248	29	206	144	250	99	73	0	0
DIS7	-89	172	3				12	67	0	
DIS9	-222	217	4				1	72	0	
DIS10	1515	293	-82				20	42	0	
DIS11	102	266	-41				11	34	0	

Patient ID	MFI (stained-unstained)						% positive cells			
	CD49d	TLR4	CD14	CD38	CD11b	CCR7	CD49d	TLR4	CD14	CD38
DIS12*	3528	635	-75				43	43	0	
DIS13	72	308	-59				5	40	0	
DIS14	9785	579	-42				86	41	0	
HEA1	13366	2378	-12				85	100	1	
HEA2	10342	1843	-102				84	100	0	
HEA3	11772	1516	-3				80	100	0	

9.11 REAL-TIME PCR EXPRESSION DATA

Table 9-15. Relative mRNA expression of ITGA4, IRF4, EZH2, KMT2D and TNFAIP3 from a cohort of trisomy 12 and disomy 12 CLL (chapter 5). CT = cycle threshold in qRT-PCR and refers to the mean of 3 replicates (that are within 0.5 cycles of each other). GUSB is the housekeeping gene. Relative expression to GUSB follows the formula: $2^{-(\text{mean CT}_{\text{target}} - \text{mean CT}_{\text{GUSB}})}$ and is rounded to two decimal places. Asterisked samples* have a hypermutated IGHV and samples in italics are CD49d-.

Patient ID	Gene expression														
	Mean CT (GUSB)	ITGA4		Mean CT (GUSB)	IRF4		Mean CT (GUSB)	EZH2		Mean CT (GUSB)	KMT2D		Mean CT (GUSB)	TNFAIP3	
		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression
2							25.09	25.57	0.71						
5*	26.35	25.19	2.22	26.35	21.55	27.7	27.75	27.64	1.08						
6* (CD49d+)	28.81	26.83	3.95	28.81	25.88	7.60	28.81	30.29	0.36	27.99	25.07	7.57	27.99	25.57	5.37
6 (CD49d-)							30.67	24.49	72.35	29.64	27.52	4.35	29.64	26.02	12.37
7	21.27	22.90	0.32	21.27	20.18	2.13				20.96	20.75	1.15	20.96	18.48	5.58
8				23.63	19.30	20.14	24.30	26.02	0.3	22.45	17.90	23.34	22.45	17.80	25.05
9				27.52	23.35	17.99				25.83	22.73	8.59	25.83	22.15	12.79
10				23.99	20.72	9.65				23.39	20.01	10.40	23.39	19.23	17.88
11	21.26	24.06	0.14	21.26	20.92	1.27	26.58	27.69	0.46	21.04	18.41	6.19	21.04	17.90	8.78
14	24.87	22.50	5.16	24.87	20.72	17.75				24.30	18.90	42.31	24.30	18.21	68.40
15	29.59	28.11	2.78	29.59	26.54	8.28	30.54	32.43	0.27	27.35	20.07	155.42	27.35	20.22	139.58
21	26.51	26.01	1.42	26.51	22.85	12.66				25.96	22.15	13.99			
21 (CD49d+)	27.90	26.78	2.17	27.9	26.63	2.42	27.9	30.97	0.12						
21 (CD49d-)	29.31	34.03	0.04	29.31	27.85	2.75	29.31	32.15	0.14						
22	21.96	24.30	0.20	21.96	23.44	0.36	22.86	25.85	0.13				21.49	18.75	6.67
68	23.55	21.65	3.73	23.55	21.71	3.58	23.55	25.38	0.28	23.36	18.24	34.95			
70*	25.09	24.20	1.86	25.09	21.12	15.73	26.11	27.84	0.30	24.84	19.60	37.95	24.84	20.39	21.93
<i>DIS1</i>	24.44	26.88	0.18	24.44	20.41	16.38	25.35	26.24	0.54	22.45	17.56	29.69	22.45	17.43	32.45
DIS5	23.00	22.96	1.02	22.82	20.12	6.5	24.33	26.07	0.30	22.32	18.44	14.72	22.32	16.62	52.02

Patient ID	Gene expression														
	Mean CT (GUSB)	ITGA4		Mean CT (GUSB)	IRF4		Mean CT (GUSB)	EZH2		Mean CT (GUSB)	KMT2D		Mean CT (GUSB)	TNFAIP3	
		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression		Mean CT	Relative expression
<i>DIS10</i>				23.13	22.48	1.57				24.98	21.98	8.00	24.98	19.95	32.76
<i>DIS11</i>	24.75	26.65	0.27	24.75	21.75	8.02	24.75	25.60	0.55				23.79	21.36	5.39
DIS12*	23.52	21.36	4.46							22.51	20.14	5.19			
<i>DIS13</i>	25.27	26.22	0.52	25.27	21.13	17.55	25.27	26.05	0.58				24.83	21.37	11.01

Table 9-16. Relative mRNA expression of TLR4 and IL8 from a cohort of trisomy 12 and disomy 12 CLL (chapter 6). CT = cycle threshold in qRT-PCR and refers to the mean of 3 replicates (within 0.5 cycles of each other). GUSB is the housekeeping gene. Relative expression to GUSB follows the formula $2^{-(\text{mean CT}(\text{target})-\text{mean CT}(\text{GUSB}))}$ and is rounded to two decimal places. Asterisked samples* have a hypermutated IGHV and samples in italics are CD49d-.

Patient ID	Gene expression						Patient ID	Gene expression					
	Mean CT (GUSB)	TLR4		Mean CT (GUSB)	IL8			Mean CT (GUSB)	TLR4		Mean CT (GUSB)	IL8	
		Mean CT	Relative expression		Mean CT	Relative expression			Mean CT	Relative expression		Mean CT	Relative expression
2				25.09	25.15	0.96	22				22.32	22.17	1.11
5*	27.05	25.82	2.34	27.75	22	53.63	68	23.55	25.16	0.33	23.55	25.16	0.33
7				21.27	21.0	1.13	70*	25.61	24.94	1.59			
8				23.63	24.22	0.66	<i>DIS1</i>	23.27	23.88	0.65	25.04	21.84	9.21
10	23.99	23.72	1.21	23.99	24.27	0.82	DIS5	23.00	24.59	0.33	24.33	25.56	0.43
11	21.51	25.29	0.07	26.58	27.69	0.46	<i>DIS11</i>	24.75	23.78	1.96	24.75	23.48	2.41
14*				24.22	26.72	0.18	DIS12*				23.52	23.26	1.20
15	29.20	29.81	0.66	30.54	29.23	2.47	<i>DIS13</i>	25.27	25.74	0.72	25.27	29.41	0.06
21	26.51	21.11	42.34										

9.12 STIMULATION ASSAY DATA

Table 9-17. LPS stimulation assay data: viability and immunophenotyping (Chapter 6). % viable CLL cells at 0h measured with Trypan blue exclusion assay. % viable CLL cells at 48h measured using DiLC/PI flow cytometry assay and are rounded to one decimal place. MFI = mean fluorescence intensity and refers to MFI of target on CLL population. LPS = lipopolysaccharide. Asterisked samples* have a hypermutated IGHV.

Patient ID	All (0h)	Control (48h)				LPS (48h)			
	% viable CLL cells	% viable CLL cells	MFI			% viable CLL cells	MFI		
			TLR4	CD49d	CD14		TLR4	CD49d	CD14
5*	40	26.7	932	479	761	24	1478	550	883
7	57	40.2	498	323	1707	66.4	418	355	1230
12*	31	20	1136 5	1819	1637	31.8	1668 9	1161	2449
14*	19	4.7	9741	295	2738	3.8	8653	201	2542
15	22	5.1	1423	333	3467	5.2	1227	316	2660
18	34	2.0	1350	566	4879	2.6	1695	576	4445
73	88	40.8	1764	565	997	22.7	4317	537	1732
DIS4*	39	10	1607	243	2227	9.8	1126	75	2411
DIS5	74	37.8	471	542	598	43.5	742	586	588
DIS10	70	50.8	1199	98	1884	48.6	754	107	1581
DIS15	23	52.3	9741	295	2738	57.6	8653	201	2542

Table 9-18. LPS stimulation assay data: gene expression (Chapter 6). CTL = control, LPS = lipopolysaccharide. CT = cycle threshold in qRT-PCR and refers to the mean of 3 replicates (within 0.5 cycles of each other). GUSB is the housekeeping gene. Relative expression to GUSB follows the formula $2^{-(\text{mean CT}(\text{target})-\text{mean CT}(\text{GUSB}))}$ and is rounded to two decimal places. Asterisked samples* have a hypermutated IGHV.

Patient ID	Conditions	Gene expression										
		Mean CT (GUSB)	TLR4		ITGA4		KMT2D		TNFAIP3		IL8	
			Mean CT	Relative Expression								
5*	CTL	34.20	35.15	0.52	31.86	5.06	31.13	8.39	29.23	31.28	26.96	151.59
	LPS											
7	CTL	29.25			28.76	1.40	26.40	7.20	25.56	15.79	30.69	0.37
	LPS	28.69			28.65	1.03	25.49	9.19	24.54	17.78	26.12	5.92
12*	CTL	26.69	31.08	0.38	29.93	0.85	28.80	1.86	29.08	1.53	26.97	6.58
	LPS	27.69	27.41	1.22	28.42	0.60	26.08	3.05	25.67	4.04	25.12	5.93
15	CTL	31.08	28.18	7.45	30.12	1.94	27.66	10.73	26.00	33.85	26.29	27.76
	LPS	31.17	28.08	8.52	30.05	2.18	27.57	12.18	25.66	45.66	27.00	18.09
73	CTL	31.10	32.70	0.33	30.69	1.34	29.66	2.72	28.16	7.68	26.87	18.82
	LPS	29.37	29.24	1.09	29.21	1.11	27.52	3.60	26.30	8.37		
DIS4*	CTL	32.87	31.84	2.04	31.66	2.32	28.34	23.12				
	LPS	31.52	31.26	1.20								
DIS5	CTL	28.98	29.01	0.97	27.59	2.62	27.00	3.94			26.65	5.01
	LPS	28.20	29.02	0.56	26.94	2.39	25.80	5.25	23.68	22.83	24.82	10.35
DIS15	CTL	27.64	28.44	0.57	30.49	0.14	25.88	3.37	24.82	7.02	23.06	23.77
	LPS	28.98	33.17	0.05	33.48	0.04	28.70	1.22	27.00	3.94	25.40	11.93

