MESSENGER RNA AND PROTEIN PROFILES IN FAMILIAL CHRONIC LYMPHOCYTIC LEUKAEMIA

Abdullah Mobarak Saeed ALSHAHRANI

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Sydney Medical School
Nepean

The University of Sydney
Australia

March 2018

DECLARATION

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at The University of Sydney or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at The University of Sydney or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

The human transcriptome arrays in chapter 3 was carried out in the Ramaciotti centre at the University of New South Wales, Australia.

Associate Professor Stephen Fuller assisted with editing of the thesis and (Table 1-1, Table 1-2, Table 9-1 and Table 9-2).

Ms. Kristen Skarratt, Sydney Medical School Nepean, University of Sydney, assisted with qRT-PCR and IGH mutation assays, and editing of the thesis.

Doctor Kristy Robledo, NHMRC Clinical Trials Centre, University of Sydney, assisted with bioinformatic and statistical analyses.

Ms. Maryam Hassanvand, Sydney Medical School Nepean, University of Sydney, assisted with DNA extractions and IGH mutation assays.

Ms. Angela Connolly, The Hub, Charles Perkins Centre, University of Sydney, assisted with processing of protein samples.

Signed

Date 29-03-2018

ABSTRACT

Chronic lymphocytic leukaemia (CLL) is a common adult leukaemia, characterised by the accumulation of clonal B lymphocytes in peripheral blood, bone marrow and lymphoid tissues. CLL accounts for approximately 16 to 30% of leukaemia cases in western countries, which contrasts with a low incidence of CLL in Asian countries and in Asian immigrants to western countries.

An inherited risk for developing chronic lymphocytic leukaemia (CLL) is well documented in genetic studies, and familial aggregation of CLL cases has consistently been demonstrated in large registry-based studies. However, genetic linkage studies of families segregating CLL have not detected any high-risk susceptibility genes against a background of numerous low-risk genes. To detect patterns of multiple low-risk loci, genome-wide association studies (GWAS) have used large numbers of cases and controls and dense-coverage single nucleotide polymorphism (SNP) arrays. These studies have identified risk loci that account for ≈19% of the heritability of CLL, suggesting that some of the remaining CLL risk may be associated with non-DNA sequence modifications, including inherited epigenetic changes, which regulate oncogenes and tumour suppressor genes.

In order to identify potential inherited changes in gene expression, high-resolution DNA microarrays and mass spectrometry (MS) were used to identify differentially abundant mRNA and proteins in cases of familial CLL (F-CLL) and monoclonal B lymphocytosis (F-MBL), and compared to unaffected relatives, sporadic CLL (S-CLL) and controls. In addition, mRNA and protein levels were studied in familial and sporadic CLL patients with mutated and unmutated immunoglobulin heavy chain variable genes (*IGH*). An advantage of family studies compared to association studies of unrelated subjects is that background genetic variation is

to some extent controlled as a function of the degree of relationship, increasing the contribution of epigenetic and/or environmental modifiers to variation in phenotype.

Key findings were that mRNA and protein profiles clearly segregated clonal B lymphocytes in S-CLL from clonal B lymphocytes in F-MBL and F-CLL (combined as familial-lymphoproliferative disease; F-LPD). These profiles were distinct from those found in normal B lymphocytes in unaffected family members and unrelated controls. Furthermore, increasing upregulation or downregulation of both F-LPD specific genes and genes common to S-CLL occurred in association with progression from normal familial B lymphocytes through F-MBL to F-CLL.

Using the GeneChip® Human Transcriptome 2.0 Array, 1893 mRNAs were identified that segregated F-LPD from S-CLL and healthy controls. Based on false discovery rate (FDR) p-values, the highest ranked upregulated genes in F-LPD were *LEF1* (p=9.69E-09), *ROR1* (p=2.54E-08), *ABCA6* (p=2.54E-08), and *MIR4524A* (p=4.80E-07); and downregulated genes, *SH3RF1* (p<0.0001), *PLD4* (p<0.0001), *FAM135A* (p<0.0001) and *SNX22* (p<0.0001).

Analysis of protein levels using a combination of quadrupole, ion trap and Orbitrap mass spectrometer analysis identified 4672 proteins, that after normalisation to 6 control samples, segregated F-LPD and S-CLL using unsupervised hierarchical clustering based on protein level patterns. Differential abundance analysis of grouped data for F-LPD and S-CLL proteins was used to select proteins for further analysis by semi-supervised hierarchical clustering. Thirty proteins were differentially expressed between F-LPD and S-CLL, with the highest significance for upregulation, HACD3 (fold change 2.64; t-test p=0.001) and MIF (fold change 3.9; p=0.003), and downregulation, SERPINH1 (fold change -2.6; p=0.0003). CYBB protein expression (fold change 3.2; p=0.04) correlated with upregulated gene expression

(FDR p=0.02).

The progression from normal B lymphocytes to MBL and CLL, showed progressive upregulation of the mRNAs *LEF1*, *C11orf80*, *ROR1*, *METTL8*, *PARP3*, *INPP5F* and *DFNB31;* while *SMAD3*, *GRASP* and *RASGEF1B* were progressively downregulated. The mRNA for *TBC1D10C* was downregulated in F-MBL before becoming upregulated in F-CLL. Of particular interest were 3 genes (*LEF1*, *ROR1* and *GRASP*) that were differentially expressed in both the F-LPD versus S-CLL comparison, and in association with progression from normal B cells to F-MBL and F-CLL. These results are consistent with the proposal that inherited dysregulation of these genes contributes to driving malignant progression of F-LPD. These differential profiles of mRNAs and proteins between categories of F-LPD should be useful for rapid diagnosis, and provide a basis for understanding the mechanisms that drive F-MBL and F-CLL.

In addition to predicting prognosis in UM-CLL compared to M-CLL, studying biological differences between the 2 subtypes may provide insights into the pathogenesis of CLL, and identify genes and proteins for targeted therapies. Using unsupervised hierarchical clustering, 582 mRNAs were differentially expressed between controls, and M-CLL and UM-CLL cases containing F-CLL and S-CLL. Although controls and M-CLL cases clustered, UM-CLL did not completely segregate. To identify mRNAs that were differentially expressed based on *IGH* mutation status and normalise for genes differentially expressed as a result of inter-individual B lymphocyte variation, mRNAs in M-CLL and UM-CLL cases were compared after both groups were normalised to controls (t-test p-value < 0.05). Nineteen genes were differentially expressed between M-CLL and UM-CLL and some of these genes have been functionally characterized. The majority of the identified genes in this study were upregulated in UM-CLL versus M-CLL. Of these mRNAs, *NRIP1* and *CERS6* segregated individual M-CLL, UM-CLL cases and controls (FDR $p < 0.05$)

To identify mRNAs differentially expressed based on *IGH* mutation status in family cases only and control for mRNAs that were differentially expressed as a result of inter-individual variation, mRNAs were compared between (A) controls and M-LPD, and (B) M-LPD and UM-CLL. There were 84 mRNAs differentially expressed between M-LPD and UM-CLL (t-test $p < 0.05$; fold change > 2), that did not differentiate both groups from controls. Semi-supervised hierarchical clustering using these genes segregated controls, M-LPD and UM-CLL.

Compared to mRNA studies, protein expression studies have found higher percentages of differentially expressed proteins in M-CLL and UM-CLL, however these studies have not normalised protein abundance in CLL cases to control samples or used datasets enriched with familial cases. Mass spectrometer analysis in the current study, identified 5100 proteins which were normalised to the 6 control samples. Twelve proteins were differentially expressed between M-CLL and UM-CLL cases normalised to controls, however at an individual case level, these proteins did not completely segregate M-CLL from UM-CLL, with one UM-CLL case segregating with M-CLL cases.

Although family studies of mRNA and protein levels in M-CLL and UM-CLL reported in Chapter 6 were limited by small sample sizes, the results have validated findings from previous studies and identified a number of new genes and proteins that may be used for prognostication or as targets for novel therapies.

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank my principal supervisor Associate Professor Dr Stephen James Fuller for his continuous support, assistance and patience, through my studies. I am extremely appreciative of the opportunities and challenges that this research project has presented me with, under his knowledgeable guidance.

I would like to express sincere thanks to Professor Richard Christopherson for his valuable guidance, support and suggestions. He always inspired me to do a number of interesting research ideas in the proteomics part, and without his guidance, this part would not have been possible.

I also would like to express my deepest gratitude to Kristy Skratte for training me in the lab and for always having time to explain, discuss and encourage me to pursue this exciting project. Also, I am grateful to my auxiliary supervisor Dr. Hooshang lahooti for his guidance, advice and help at all stages of this study.

The Mass spectrometry core facility (MSCF) at the Charles Perkins Centre deserve my sincere thanks for their help in the proteomics part of this project. I am grateful to Dr Ben Crossett and Dr. Angela Connolly for helping me with mass spectrometry for peptide separation, protein identification and quantification. I would also like to thank Dr. Swetlana Mactier and Dr. Munther Alomari for excellent advice in protein separation and mass spectrometry analysis.

Special thanks to Dr. Kristy Robledo, NHMRC Clinical Trials Centre, University of Sydney, for her assistance in bioinformatics and statistical analyses.

Furthermore, my special thanks go to all current and previous members of Sydney Medical School Nepean for their constant support, friendship and positive suggestions about my research project. Also, I am grateful to Maryam Hassanvand, a medical research student for helping with the IgVH analysis for this project.

I am also grateful to patients and members of the family for their participation in this study.

I would also like to acknowledge the King Khalid University and the Cultural Mission of Saudi Arabian in Canberra for funding my scholarship.

Finally, I am deeply indebted to my parents, siblings and also to my beautiful family (Amani, Ruyuf, Feras & Rafeef) for their continuous love, understanding and encouragement through my PhD study. This achievement would not have been possible without all of your excellent support, guidance and continuous encouragement.

FINANCIAL SUPPORT

The Cultural Mission of Saudi Arabia provided principal funding for the study.

A gift from Gary and Marion Krelle.

The Nepean Medical Research Fund.

TABLE OF CONTENTS

DECLARATION I
ABSTRACT II
ACKNOWLEDGEMENTS VI
FINANCIAL SUPPORT VII
LIST OF ABBREVIATIONS XIII
LIST OF FIGURES XVI
LIST OF TABLES XIX
MANSCRIPTS PREPARED FOR SUBMISSION DURING THE COURSE OF THIS THESIS XXI
ACCEPTED ABSTRACT FOR ORAL PRESENTATION XXII
ORAL PRESENTATIONS XXII
THESIS STRUCTURE XXIII
CHAPTER-1: INTRODUCTION 1
1.1. Leukaemia 2
1.2. Lymphoid Neoplasms 2
1.3. Chronic Lymphocytic Leukaemia 4
1.3.1. Epidemiology 4
1.3.2. Evidence for Familial Predisposition 4
1.3.3. Monoclonal B-Cell Lymphocytosis (MBL) in S-CLL and F-CLL 7
1.3.4. Genetic Risk Factors 8
1.3.5. Cytogenetic Abnormalities 18
1.3.6. Proteomic Profiling in B-CLL 34
1.3.7. *IGH* Gene Usage and Mutation Status in F-CLL 35
1.4. SUMMARY 37
AIMS 38
CHAPTER-2: MATERIALS AND METHODS 40
2.1. SAMPLE SETS 41
2.1.1. Healthy Control Subjects 41
2.1.2. Sporadic Chronic Lymphocytic Leukaemia (S-CLL) Subjects 41
2.1.3. Familial Chronic Lymphocytic Leukaemia (F-CLL) subjects 42
2.2. CELL PURIFICATION 42
2.2.1. MATERIALS 42
2.2.2. METHOD 42

2.3. FLOW CYTOMETRY 44
2.3.1. MATERIALS 44
2.3.2. METHODS FOR FLOW CYTOMETRIC ANALYSIS 45
2.4. RNA EXTRACTION 48
2.4.1. MATERIALS 48
2.4.2. METHODS 48
2.5. INTERPHASE FLUORESCENCE IN SITU HYBRIDIZATION (FISH) 50
2.5.1. MATERIALS AND SOLUTIONS 50
2.5.2. METHOD 51
2.6. *IGH* GENE CLONALITY ASSAY 54
2.6.1. MATERIALS 55
2.6.2. METHOD 55
2.7. Quantitative reverse-transcriptase PCR (qRT-PCR) 59
2.7.1. MATERIALS 59
2.7.2. METHOD 60
CHAPTER-3: A COMPARISON OF MESSENGER RNA PROFILES IN FAMILIAL MBL, FAMILIAL B-CLL AND SPORADIC B-CLL 61
3.1. INTRODUCTION 62
HYPOTHESIS 65
AIMS 65
3.2. MATERIALS 66
3.3. METHODS 66
3.3.1. Patients and Samples 66
3.3.2. *IGH* Usage and Mutation Analysis 67
3.3.3. Interphase Fluorescence in Situ Hybridization (FISH) 68
3.3.4. RNA Extraction 68
3.3.5. Transcriptome Profiling 70
3.3.6. Affymetrix Expression Console (EC) Software 71
3.3.7. Monitoring Sample Quality: 71
3.3.8. Monitoring Hybridisation and Labelling Quality: 71
3.3.9. Validation of Gene Microarrays Using qRT-PCR 72
3.3.10. Ingenuity Pathway Analysis 73
3.3.11. Statistical and Bioinformatics Analyses 73
3.4. RESULTS 74

3.4.1. Patients and Samples 74
3.4.2. Interphase Fluorescence in Situ Hybridization (FISH) Results 74
3.4.3. *IGH* Gene Clonality Assay Results 75
3.4.4. Unsupervised Hierarchical Clustering of Gene Expression Data for Family Members with MBL or CLL and Sporadic CLL Cases 77
3.4.5. Ingenuity Pathway Analysis 79
3.4.6. Semi-Supervised Hierarchical Clustering of Gene Expression Data for S-CLL Cases and F-LPD 81
3.4.7. Validation of Gene Microarrays Using qRT-PCR 86
3.5. DISCUSSION 87
CHAPTER- 4: A COMPARISON OF PROTEIN PROFILES IN FAMILIAL MBL, FAMILIAL B-CLL AND SPORADIC B-CLL 92
4.1. Proteomic Studies in CLL 93
4.2. Advances in Mass Spectrometry 96
4.3. Bioinformatics Tools for Database Searching and Analysis 97
HYPOTHESES 100
AIMS 100
4.4. MATERIALS 101
4.5. METHOD 102
4.5.1. Negative Selection of B-CLL Cells from Blood Samples 102
4.5.2. Preparing Samples for Protein Profiles at the MSCF 103
4.5.3. Protein Precipitation and Clean-Up 103
4.5.4. Protein Assay 104
4.5.5. Protein Reduction and Alkylation 104
4.5.6. Protein Digestion 105
4.5.7. Peptides Desalting and Concentration 105
4.5.8. TMT10plex Mass Tag Labelling 106
4.5.9. Offline HILIC Separation: 107
4.5.10. Liquid Chromatography Tandem Mass Spectrometry Analysis Using an Orbitrap Fusion Tribrid™ Mass Spectrometer 108
4.5.11. Bioinformatic Tools for Database Searching and Analyses 109
4.5.12. Statistical Analyses 111
4.5.13. Western Blot validation of Mass Spectrometry 111
4.6. RESULTS 113
4.7. DISCUSSION 122

SUMMARY ... 128
CHAPTER-5: CHANGES IN MESSENGER RNA PROFILES ASSOCIATED WITH PROGRESSION FROM NORMAL B LYMPHOCYTES TO F-MBL AND F-CLL ... 130
5.1. INTRODUCTION ... 131
HYPOTHESIS ... 132
AIM ... 132
5.2. MATERIALS ... 133
5.3. METHOD ... 133
5.3.1. Statistical Analysis ... 134
5.4. RESULTS ... 134
5.4.1. Validation of Gene Microarrays Using qRT-PCR ... 139
5.5. DISCUSSION ... 141
5.6. CONCLUSION ... 144
CHAPTER-6: A COMPARISON OF MESSENGER RNA AND PROTEIN PROFILES IN IGH MUTATED AND UNMUTATED CLL ... 145
6.1. INTRODUCTION ... 146
HYPOTHESIS ... 148
AIMS ... 148
6.2. METHODS ... 149
6.2.1. M-CLL and UM-CLL mRNA Profiles ... 149
6.2.2. Protein Profiles in M-CLL and UM-CLL ... 150
6.3. RESULTS ... 152
6.3.1. M-CLL and UM-CLL mRNA Profiles ... 152
6.3.2. mRNA Profiles in *IGH* Mutated and Unmutated CLL Family Cases ... 158
6.3.3. Protein profiles in *IGH* mutated and unmutated CLL ... 163
6.4. DISCUSSION ... 166
CHAPTER-7: DISCUSSION AND FINAL COMMENTS ... 173
IN SUMMARY ... 176
FUTURE DIRECTIONS ... 177
CHAPTER-8: REFERENCES ... 179
CHAPTER-9: APPENDICES ... 216
9.1. List of CLL candidate gene association studies ... 217
9.2. List of Disease pathway and multiple gene association studies in CLL and NHL ... 223
9.3. FLOW CYTOMETRY RESULTS ... 226
9.3.1. Purity check result ... 226

9.3.2. Screening for B cell lymphocytosis 228
9.4. FISH analysis in CLL and MBL cells 230
9.5. Purity of total RNA in PBMCs 231
9.6. RNA concentrations (ng/µL) 232
9.7. A260/A280 ratio Results 232
9.8. RNA quality Assessment 233
9.9. Electropherogram summary results of total RNA 234
9.10. Monitoring sample quality 235
9.11. Monitoring signal intensity distributions 236
9.12. The distributions of probe set signals before and after normalization 237
9.13. The relative log expression values before and after normalization 238
9.14. Monitoring hybridization efficiency 239
9.15. Monitoring labelling efficiency 239

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANDED TERM
°C	degrees Celsius
2D-DIGE	2D-fluorescence difference gel electrophoresis
ACN	acetonitrile
AGC	automatic gain control
AGRF	Australian Genome Research Facility
ANOVA	analysis of variance
ATM	ataxia-telangiectasia mutated gene
BCL-2	B-cell CLL/lymphoma 2
BIRC3	baculoviral IAP repeat-containing protein 3
ß-ME	β-mercaptoethanol
CD	cluster of differentiation
CID	collision induced dissociation
CLL	chronic lymphocytic leukaemia
CPC	Charles Perkins Centre
DAPI	4,6-diamidino-2-phenylindole
DAPK1	death-associated protein kinase 1
DBD	DNA binding domain
DEL	deletion
DLEU	deleted in lymphocytic leukaemia
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DZ	dizygotic
EBV	Epstein- Barr virus
EDTA	ethylenediaminetetraacetic acid
FBXW7	F-box and WD40 domain protein 7
F-CLL	familial chronic lymphocytic leukaemia
FCR	fludarabine, cyclophosphamide and rituximab
FCS	fetal calf serum
FDR	false discovery rate
FISH	fluorescent in situ hybridization
F-LPD	familial lymphoproliferative disease
F-MBL	familial monoclonal B-cell lymphocytosis
FSC	forward light scatter
FWHM	full-width-at-half-maximum
GWAS	genome-wide association study
HCD	higher-energy C-trap dissociation
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HILIC	hydrophilic interaction liquid chromatography
HLB	hydrophilic-hydrophobic balanced
HPLC	high-pressure liquid chromatography
IAA	iodoacetamide
IFN-G	interferon gamma

IGH	immunoglobulin heavy chain
IPA	Ingenuity Pathway Analysis
IRF4	interferon regulatory factor 4
ITRAQ	isobaric tags for relative and absolute quantitation
LC-MS	liquid chromatography mass spectrometry
LC-MS/MS	liquid chromatography–tandem mass spectrometry
LFS	Li Fraumeni syndrome
LOD SCORES	logarithm of odds scores
M	molar
m/z	mass/charge ratio
MALDI-TOF/TOF-MS	matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry
MAPK	mitogen-activated protein kinase
MBL	monoclonal B-cell lymphocytosis
M-CLL	CLL with mutated IGHV genes
MDM2	mouse double minute homologue 2
MDR	minimal deleted region
µg	microgram
µl	microliter
min	minutes
ml	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MSCF	mass spectrometry core facility
mTOR	mammalian/mechanistic target of rapamycin
mTORC1/2	mammalian target of rapamycin complex 1/2
MW	molecular weight
MZ	monozygotic
NCI	National Cancer Institute
NF-KAPPA-B	nuclear factor kappa-light-chain-enhancer of activated B cells
ng	nanogram
NGS	next generation sequencing
NHL	non-Hodgkin lymphoma
NICD	notch1 intracellular domain
NK	natural killer cell
NPL	nonparametric linkage
NP-LC	normal phase liquid chromatography
OR	odds ratio
OS	overall survival
PBS	phosphate buffered saline
PCA	principal components analysis
PCR	polymerase chain reaction
PI3K	phosphatidylinositol 3-kinase
PRE-MRNAS	precursor messenger RNAs
PSM	peptide mass spectra
RB1	retinoblastoma

RIN	RNA integrity number
RLE	relative log expression
RLY	RNA lysis buffer
RP-LC	reversed-phase liquid chromatography
RT	room temperature
QRT-PCR	quantitative reverse transcriptase PCR
s	second
S/N	signal to noise
S-CLL	sporadic chronic lymphocytic leukaemia
SDS	sodium dodecyl sulfate
SEER	surveillance, epidemiology, and end Results
SF3B1	splicing factor 3B, subunit 1
SLL	small lymphocytic lymphoma
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleic-protein
SPS	synchronous precursor selection
SRM	selected reaction monitoring
SSC	side scatter
ssDNA	single stranded DNA
SVM	support vector machine
TAC	transcriptome analysis console
T-ALL	T cell acute lymphoblastic leukaemia
TEAB	tetraethylammonium bicarbonate
TFA	trifluoroacetic acid
TMT	tandem mass tag
TNF	tumour necrosis factor
UM-CLL	CLL with un-mutated IGH genes
WHO	The World Health Organisation
WT	whole transcript

LIST OF FIGURES

FIGURE 1-1. PEDIGREE OF THE FAMILY IN ABBREVIATED FORM 33

FIGURE 1-2. THE OVERVIEW OF THE EXPERIMENTAL DESIGN USED IN THIS THESIS 39

FIGURE 3-1 PEDIGREE OF THE FAMILY. 76

FIGURE 3-2 UNSUPERVISED HIERARCHICAL CLUSTERING OF B LYMPHOCYTE GENE EXPRESSION 78

FIGURE 3-3. VENN DIAGRAM ILLUSTRATING OVERLAP BETWEEN F-LPD VS S-CLL, AND F-LPD VS CONTROLS 81

FIGURE 3-4. SEMI-SUPERVISED HIERARCHICAL CLUSTERING OF B LYMPHOCYTE MRNA LEVELS FOLLOWING CORRECTION FOR RELATEDNESS. 82

FIGURE 3-5. SEMI-SUPERVISED HIERARCHICAL CLUSTERING OF B LYMPHOCYTE MRNA LEVELS FOLLOWING REMOVAL OF PAG1 AND ITGA4 84

FIGURE 3-6. VALIDATION OF THE EXPRESSION OF GRASP AND SMAD3 RELATIVE TO GAPDH (DELTA CT) IN B-LYMPHOCYTES FROM S-CLL AND F-LPD SUBJECTS 87

FIGURE 4-1. EXPERIMENTAL PROCEDURE FOR QUANTITATIVE DIFFERENTIAL PROTEOMICS USING TMT10PLEX LABELLING 99

FIGURE 4-2 TMT10PLEX MASS TAG LABELLING DESIGN FOR BOTH EXPERIMENTS 107

FIGURE 4-3 UNSUPERVISED HIERARCHICAL CLUSTER IMAGE OF PROTEIN EXPRESSION IN S-CLL AND F-LPD (COMBINED F-MBL AND F-CLL) 114

FIGURE 4-4 VOLCANO PLOT OF PROTEIN EXPRESSION FOR F-LPD VERSUS S-CLL 115

FIGURE 4-5 CLUSTERING ANALYSIS OF PROTEINS DIFFERENTIALLY EXPRESSED BETWEEN F-LPD AND S-CLL 117

FIGURE 4-6. EXPRESSION OF CYBB PROTEIN IN F-MBL (IV-17) COMPARED TO S-CLL57 118

FIGURE 4-7 PROTEINS IMPLICATED IN EARLY AND ADVANCED CLL AND B-CELL PROLIFERATION. 120

FIGURE 5-1 GENES DIFFERENTIALLY EXPRESSED USING UNSUPERVISED HIERARCHICAL CLUSTERING IN CONTROL FAMILY SUBJECTS, F-MBL AND F-CLL CASES 135

FIGURE 5-2. VENN DIAGRAM ILLUSTRATING OVERLAP BETWEEN F-MBL VS F-CLL, AND F-MBL VS CONTROLS 137

FIGURE 5-3. GENES DIFFERENTIALLY EXPRESSED BY SEMI-SUPERVISED HIERARCHICAL CLUSTERING OF B LYMPHOCYTE MRNA LEVELS WITH PROGRESSION FROM NORMAL B FAMILY SUBJECTS, PREMALIGNANT F-MBL TO MALIGNANT F-CLL CASES 138

FIGURE 5-4. QUANTITATIVE RT-PCR ANALYSES FOR 6 MRNAS IN B-LYMPHOCYTES FROM CONTROL, F-MBL AND F-CLL SUBJECTS 140
FIGURE 6-1 CLUSTER IMAGE OF MRNA EXPRESSION IN NORMAL, M-CLL AND UM-CLL 153
FIGURE 6-2. VENN DIAGRAM ILLUSTRATING OVERLAP BETWEEN M-CLL VS UM-CLL, AND M-CLL VS CONTROLS 154
FIGURE 6-3.CLUSTER IMAGE OF NORMAL, MUTATED AND UNMUTATED CLL GENE EXPRESSION FOLLOWING CORRECTION FOR RELATEDNESS. 155
FIGURE 6-4. CLUSTER IMAGE OF 2 MRNAS BETWEEN M-CLL, UM-CLL AND CONTROLS 156
FIGURE 6-5. CLUSTER IMAGE OF GENE EXPRESSION IN FAMILIAL M-CLL, UM-CLL, AND FAMILIAL CONTROLS 159
FIGURE 6-6. VENN DIAGRAM ILLUSTRATING OVERLAP BETWEEN FAMILIAL M-CLL, UM-CLL, AND FAMILIAL CONTROLS 160
FIGURE 6-7. CLUSTER IMAGE OF FAMILIAL M-CLL AND UM-CLL CASES AND FAMILY CONTROLS GENE EXPRESSION 161
FIGURE 6-8. CLUSTER IMAGE OF 18 GENES BETWEEN FAMILIAL M-CLL, UM-CLL AND FAMILIAL CONTROLS (FDR < 0.05) 162
FIGURE 6-9. VOLCANO PLOT PROTEIN EXPRESSION IN M-CLL VERSUS UM-CLL 164
FIGURE 6-10. CLUSTERING ANALYSIS OF PROTEINS DIFFERENTIALLY ABUNDANT BETWEEN M-CLL AND UM-CLL 166
FIGURE 9-1. FLOW CYTOMETER ANALYSIS SHOWING PERCENTAGE B-CLL CELLS FOLLOWING ROSETTESEP PURIFICATION 227
FIGURE 9-2. FLOW CYTOMETRY GATING STRATEGY FOR DETECTION OF MONOCLONAL B LYMPHOCYTOSIS 229
FIGURE 9-3. FISH ANALYSIS IN CLL AND MBL CELLS 230
FIGURE 9-4. PURITY OF TOTAL RNA IN PBMCS 231
FIGURE 9-5. RNA CONCENTRATIONS (NG/μL) 232
FIGURE 9-6. A260/A280 RATIO RESULTS 232
FIGURE 9-7. RNA QUALITY ASSESSMENT 233
FIGURE 9-8. ELECTROPHEROGRAM SUMMARY RESULTS OF TOTAL RNA FROM ALL SAMPLES INCLUDED IN MICROARRAY ASSAY 234
FIGURE 9-9. MONITORING SAMPLE QUALITY 235
FIGURE 9-10. SIGNAL INTENSITY HISTOGRAM DISTRIBUTIONS 236

FIGURE 9-11. BOX PLOTS FOR PROBE SET SIGNAL VALUES DISTRIBUTIONS BEFORE AND AFTER NORMALIZATION 237

FIGURE 9-12. BOX PLOTS FOR THE RELATIVE LOG EXPRESSION VALUES BEFORE AND AFTER NORMALIZATION 238

FIGURE 9-13. MONITORING HYBRIDIZATION EFFICIENCY... 239

FIGURE 9-14. MONITORING LABELLING EFFICIENCY 239

LIST OF TABLES

Table 1-1 Loci of maximum linkage in CLL genetic linkage studies ... 11

Table 1-2. Summary of genome wide association studies (GWAS) in CLL ... 17

Table 1-3 A summary of the most frequent mutations in CLL ... 20

Table 2-1 List of materials used for cell purification ... 42

Table 2-2 List of materials used for Flow cytometry ... 44

Table 2-3 Antibody cocktails for MBL screening ... 45

Table 2-4 Cell staining for assessing B cell enrichment ... 46

Table 2-5 List of materials used for RNA Extraction ... 48

Table 2-6 List of materials used for FISH Analysis ... 50

Table 2-7 List of solutions used for FISH Analysis ... 51

Table 2-8 List of materials used for IGH gene clonality assay ... 55

Table 2-9 List of primers used for qRT-PCR validation of gene expression ... 59

Table 3-1 List of materials used for Affymetrix GeneChip Human Transcriptome Array 2.0 ... 66

Table 3-2 Summary of sample characteristics ... 76

Table 3-3 A summary of the IPA results for 1893 cDNA elements differentially abundant (FDR < 0.05) between F-LPD, S-CLL and normal controls ... 80

Table 3-4 Differential levels of mRNAs from grouped data for F-LPD and S-CLL ... 85

Table 4-1 List of materials used for proteomics ... 101

Table 4-2. List of materials used for western blot ... 101

Table 4-3. List of reagents required for western blot ... 102

Table 4-4. Gel composition to make a 3–15% gradient gel ... 102

Table 4-5. List of primary antibodies used for western blot ... 102

Table 4-6. Proteins differentially expressed between F-LPD vs S-CLL with > 2-fold change and p-value < 0.05 ... 116

Table 4-7 Integrated pathway analysis of proteins differentially abundant between F-LPD & S-CLL ... 121

Table 5-1 List of materials used for Affymetrix GeneChip Human Transcriptome Array 2.0 ... 133

Table 5-2 Genes differentially expressed using unsupervised hierarchical clustering in control family subjects, F-MBL and F-CLL cases ... 136

Table 5-3 Genes differentially expressed between F-MBL and F-CLL cases that were also differentially expressed for control versus F-MBL 137
Table 6-1. List of the selected samples applied in mRNA and Protein profiles in IGH M-CLL and UM-CLL........ 149
Table 6-2 Genes differentially expressed between M-CLL and UM-CLL normalised for controls groups 157
Table 6-3. Genes differentially expressed between normal, M-CLL and UM-CLL groups (FDR < 0.05) 158
Table 6-4. Genes differentially expressed between familial M-CLL, UM-CLL and familial normal groups (FDR < 0.05) 163
Table 6-5 Proteins differentially expressed between M-CLL and UM-CLL groups 165
Table 9-1. List of CLL candidate gene association studies 217
Table 9-2. Disease pathway and multiple gene association studies in CLL and NHL 223

MANSCRIPTS PREPARED FOR SUBMISSION DURING THE COURSE OF THIS THESIS

Two manuscript baseds on the results presented in this thesis are currently in preparation

"Messenger RNA and protein profiles identify differentially abundant markers in clonal B lymphocytes associated with familial chronic lymphocytic leukemia"

Abdullah Alshahrani[1], Kristen K Skarratt[1], Kristy Robledo[2], Angela Connolly[3], Maryam Hassanvand[1], Benjamin Tang[1], Benjamin Crossett[3], James S Wiley[4], Richard I Christopherson[5], Stephen J Fuller[1].

1. Department of Medicine, Sydney Medical School Nepean, University of Sydney, Nepean Hospital, Penrith, NSW, 2750, Australia

2. NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW, 2006, Australia

3 The Hub, Charles Perkins Centre, Building D17, Orphans School Creek Lane, University of Sydney Camperdown NSW, 2006, Australia

4. The Florey Institute of Neuroscience and Mental Health, Level 1, HFI Building, University of Melbourne, Parkville, Victoria, 3052, Australia

5. School of Life and Environmental Sciences, University of Sydney, Camperdown, NSW, 2006, Australia

"Messenger RNA profiles differentiate mutated and unmutated CLL cases and progression of MBL to CLL in familial CLL"

Abdullah Alshahrani[1], Kristen K Skarratt[1], Kristy Robledo[2], Angela Connolly[3], Maryam Hassanvand[1], Benjamin Tang[1], Benjamin Crossett[3], James S Wiley[4], Richard I Christopherson[5], Stephen J Fuller[1]

1. Department of Medicine, Sydney Medical School Nepean, University of Sydney, Nepean Hospital, Penrith, NSW, 2750, Australia

2. NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW, 2006, Australia

3 The Hub, Charles Perkins Centre, Building D17, Orphans School Creek Lane, University of Sydney Camperdown NSW, 2006, Australia

4. The Florey Institute of Neuroscience and Mental Health, Level 1, HFI Building, University of Melbourne, Parkville, Victoria, 3052, Australia

5. School of Life and Environmental Sciences, University of Sydney, Camperdown, NSW, 2006, Australia

ACCEPTED ABSTRACT FOR ORAL PRESENTATION

<u>A.ALSHAHRANI</u>, M. HASSANVAND, K. SKARRATT, K. ROBLEDO, S. FULLER. (2017).*The Molecular Mechanisms of Familial and Sporadic Chronic Lymphocytic Leukaemia (CLL)***.** The 2017 Annual Nepean Research Day.

ORAL PRESENTATIONS

- A talk entitled “Proteome analysis of familial and progressive CLL” was given to the department of haematology at the Sydney Medical School Nepean, the University of Sydney.

- A talk entitled “Transcriptome analysis of familial and progressive CLL” was given to the Professor Richard Christopherson research group at the School of Life and Environmental Sciences, the University of Sydney.

THESIS STRUCTURE

This thesis consists of seven chapters: Chapter 1 reviews the evidence for familial predisposition to CLL, the genetic risk factors for neoplastic transformation, mutations associated with CLL using genome-wide association studies, and proteomic studies; Chapter 2 details the materials and methods used in this thesis; Chapter 3 reports the results of a study to determine if F-LPD B lymphocytes contained unique mRNA signatures compared to B lymphocytes from unaffected family members and sporadic CLL (S-CLL) cases; Chapter 4 reports the results of a study to determine if F-LPD B lymphocytes contained unique protein signatures compared to B lymphocytes from unaffected family members and S-CLL cases; Chapter 5 reports the results of a study to identify changes in mRNAs associated with progression of normal B lymphocytes through pre-malignant MBL cells to malignant CLL; Chapter 6 reports the results of a study to determine whether UM-CLL B lymphocytes contain unique mRNA and/or protein signatures compared to M-CLL, and the results of a study to identify mRNAs differentially expressed based on *IGH* mutation status and partially controlled for germline genetic factors, between controls, M-LPD, and UM-CLL in familial cases; Chapter 7 concludes the thesis by summarising the main results and discussing future directions for identifying additional candidate genes and proteins associated with the development and progression of F-CLL and S-CLL using a combination of mRNA and protein profiling.

CHAPTER-1: INTRODUCTION

1.1. Leukaemia

Leukaemias are a heterogenous group of haematologicial neoplasms, classified as either lymphoid or myeloid based on the origin of the malignant cell clone. Lymphoid leukaemias are subclassified into T-cell, B-cell and natural killer- (NK) cell leukaemias. Myeloid leukaemias, which originate in a single myeloid lineage or a pluripotent progenitor cell, can affect single or multiple granulocyte cell types, including neutrophil, basophil and eosinophil, monocyte/ macrophage, mast cell, erythrocyte, and megakaryocytic lines (Vardiman et al., 2009). In addition to cell of origin, leukaemias are classified according to rate of disease progression, broadly divided into acute and chronic. In acute leukaemia, the bone marrow fails to produce mature blood cells due to a combination of failure of cell differentiation, and rapid accumulation of neoplastic cells. Chronic leukaemias progress over a longer period, leading to increased numbers of differentiated cells in the peripheral blood. Chronic leukaemia is the most prevalent form of adult leukaemia, however, treatment is commonly delayed and cure remains challenging (Mughal et al., 2006).

1.2. Lymphoid Neoplasms

There is significant overlap between the malignant cell of origin in each of the lymphoid neoplasms, which are broadly classified as lymphoma and leukaemia (Swerdlow et al., 2016), however, the two conditions are distinguished clinically by the principle site/s where malignant cells accumulate. Leukaemia is characterised by accumulation of malignant cells in the peripheral blood and bone marrow, whereas in lymphomas, malignant lymphoid cells accumulate in lymph nodes and other lymphoid tissues. The mature lymphoid neoplasms are a heterogeneous group of neoplasms, which are subdivided into mature B-cell neoplasms,

mature T-cell and NK-cell neoplasms, Hodgkin lymphoma and post-transplant lymphoproliferative disorders (Swerdlow et al., 2016) (Swerdlow et al., 2017).The aggressive B-cell lymphoma group contains precursor B-cell lymphoblastic leukaemia/lymphoma, and lymphomas categorized as mature B cell neoplasms, including diffuse large B-cell lymphoma, blastic (blastoid/pleomorphic) mantle cell lymphoma, Burkitt lymphoma, and high-grade B cell lymphoma (Arber et al., 2016, Ott, 2017). The distinction between B-cell acute lymphoblastic leukaemia and lymphoblastic lymphoma is based on the presence of ≥ 20% blasts in the bone marrow or peripheral blood (Mughal et al., 2006) (Vardiman et al., 2009).

Chronic lymphocytic leukaemia is classified as a mature B-cell lymphoid neoplasm (Swerdlow et al., 2016). The diagnosis of CLL requires an absolute malignant B lymphocyte count of ≥ 5 x 10^9/L which co-express CD5 and CD23 on immunophenotyping (Hallek et al., 2008). Patients who present with lymphadenopathy and malignant B lymphocytes immunophenotypically identical to CLL, but with a peripheral blood lymphocyte count < 5 x 10^9/L, had been considered by the National Cancer Institute (NCI) criteria to have small lymphocytic lymphoma (SLL) (Hallek et al., 2008) . However, the World Health Organization classifies CLL together with small lymphocytic lymphoma (SLL) as CLL/SLL (Swerdlow et al., 2016). Patients with CLL and SLL share similar genetics, history and complications, and the clinical management of both are similar. Virtually all CLL cases are preceded by a monoclonal B-cell lymphocytosis (MBL) in the peripheral blood of < 5 x 10^9/L with the phenotype of CLL or atypical CLL B cells in the absence of other features of lymphoma (Swerdlow et al., 2016).

Together, B-cell lymphomas and CLL/SLL are the most common haematological malignancies in western countries (Siegel et al., 2015), and are considered to have a significant inherited component distinct from syndromes such as Fanconi anaemia, which have well-defined predisposition genes (Segel and Lichtman, 2004). There are likely to be similarities and

differences in genetic risk factors between lymphoma subtypes and CLL/SLL, and the frequency of familial CLL has generated focus on this subtype to identify susceptibility genes.

1.3. Chronic Lymphocytic Leukaemia

1.3.1. Epidemiology

Chronic lymphocytic leukaemia is the most common leukaemia in western countries, representing ≈ 16 to 30% of all cases (Howlader N, 2017). A disease of older adults, CLL has a median age at diagnosis of 72 years and is more common in males (Inamdar and Bueso-Ramos, 2007) (Ruchlemer and Polliack, 2013). A genetic predisposition to the disease is further suggested by a low incidence of CLL in Asian countries, and in Asian immigrants to western countries (Dighiero and Hamblin, 2008) (Crowther-Swanepoel and Houlston, 2010). In a comparison population study of leukaemia incidence rates among Asian-American immigrants, particularly from Japan and China, compared to US white residents, Asian-Americans had lower incidence rates of CLL (Pang et al., 2002). Lower incidence has also been reported among African-American, Hispanic and Middle Eastern countries, except Israel, which has a high incidence rate of CLL, similar to North America and Europe. The highest incidence of B-CLL has been reported in Russian populations, specifically Latvian (Ruchlemer and Polliack, 2013).

1.3.2. Evidence for Familial Predisposition

Several lines of evidence suggest that a family history of CLL is associated with an increased risk that is not confounded by non-genetic risk factors. Studies that have found a familial predisposition to CLL are summarized below.

1.3.2.1. Twin Studies

Comparisons between monozygotic (MZ) twins and dizygotic (DZ) twins are used to determine the degree of genetic and environmental influence on a specific trait, and the concordance of CLL between MZ and DZ pairs of twins provides information on whether a familial pattern is due to hereditary or environmental influences. Although there have been case reports of only 5 MZ twins (Brok-Simoni et al., 1987, Cuttner, 1992, Eriksson and Bergstrom, 1987, Guasch, 1954, Lynch et al., 2002), in a study of 44 788 pairs of twins listed in the Swedish, Danish, and Finnish twin registries there was an excess of concordant MZ twins compared with DZ twins for leukaemia, and heritability was estimated to be 21% (Lichtenstein et al., 2000). Although this study had not examined the risk of leukaemia by specific subtype, it is likely that the familial risk of leukaemia reflected an increased risk of CLL since acute lymphoblastic and myeloid leukaemias are the primary potential confounding diagnoses, but do not display increased sibling risks (Albright et al., 2012).

1.3.2.2. Familial Aggregation

Multiple-case CLL families provide some of the strongest evidence for inherited susceptibility. The majority of reported families have been nuclear clusters with inherited disease compatible with the full range of genetic models of predisposition (Goldin et al., 2003, Sellick et al., 2005b, Ng et al., 2006), and large multi-generational families, such as the family studied in this current study, can potentially provide insight into the most likely model of inherited predisposition (Fuller et al., 2008).

CLL families are defined as having ≥ 2 related family members with CLL, and a familial CLL case is an affected subject who has ≥ 1 relative/s with CLL (Slager and Zent, 2014). The designs used to study these families have included case-control, cohort, and registry-based

studies. Although these studies cannot separate the roles of shared genetics from a shared environment; there is a lack of a clear relationship between CLL and exposure to environmental factors (Hatch and Cardis, 2017, Chang et al., 2005).

1.3.2.3. Case-Control Studies

The largest case-control study to date was a pooled analysis of 17 471 NHL cases, which included 2 440 CLL/SLL, and 23 096 controls from 20 studies in the International Lymphoma Epidemiology Consortium (Morton et al., 2014). Among family history variables, the greatest heterogeneity among NHL subtypes was observed for a family history of CLL/SLL, which increased risk 2.41-fold.

A large population-based case–control study of NHL, which linked the Swedish Multi-Generation Register and the Swedish Cancer Register, found an elevated risk of CLL with a family history of CLL [odds ratio (OR) = 6.3] (Chang et al., 2005). This study used validated, registry-based family data, and identified no associations between NHL risk and environmental exposures.

These studies support a familial predisposition to CLL; however, the case-control study design is susceptible to several types of bias (Guyatt et al., 2011), including under- or over-matching of participants, and recall bias; in particular, that cases and controls may differentially report a family history of CLL.

1.3.2.4. Cohort Studies

Cohort studies overcome some of the limitations of a case-control design (Grimes and Schulz, 2002). The 3 cohort studies identified for CLL were all retrospective in design. In 1975 Gunz *et al*. first reported a familial relative risk of 2.4 for CLL in an Australia population, which was based on a survey of 909 families ascertained through leukaemia cases (Gunz et al., 1975).

A second study in a Tasmanian population analysed the family histories of cases diagnosed over a 9 year period, however LPD was not subclassified into CLL (Giles et al., 1984). In a third study, the familial relative risk of CLL was studied using the Utah Population Database by identifying all cases of CLL in first-degree relatives (Goldgar et al., 1994). Observed values were compared with expected based on cohort-specific internal rates calculated from 399 786 relatives of all individuals in the database known to have died in Utah. The relative risk of CLL in first-degree relatives reported in this study was 5.7 (Goldgar et al., 1994).

1.3.2.5. Registry Studies of Familial Risk

The most comprehensive registry statistics available for familial aggregation is from a study of 9,717 CLL cases and 38,159 controls ascertained through the Swedish Cancer Registry (Goldin et al., 2009b). This study compared CLL risk in first-degree relatives of lymphoma patients with risk in relatives of matched population controls. In this study, first degree relatives of affected individuals showed an 8.5-fold increased relative risk of CLL (Goldin et al., 2009a).

1.3.3. Monoclonal B-Cell Lymphocytosis (MBL) in S-CLL and F-CLL

Following the introduction of sensitive monoclonal antibody panels for use in flow cytometry, small populations of B cells expressing the same immunophenotype as B-CLL cells were found in a proportion of normal subjects (Rawstron, 2002). Furthermore, monoclonality was demonstrated using immunoglobulin light chain restriction and, in a proportion of cases, *IGH* restriction could be established by gene rearrangement studies (Rawstron, 2002). A monoclonal B lymphocytosis (MBL) was found in 3.5% of 910 normal individuals > 40 years of age using 4 colour flow cytometry analysis of CD19/CD5/CD79b/CD20 expression followed by analysis of kappa and light chain expression. *IGH* PCR performed on 20 of 32 cases detected a

monoclonal rearrangement in 8 of 12 amplifiable samples, which was similar to the detection rate in CLL (Rawstron, 2002).

Monoclonal B lymphocytosis precedes almost all cases of CLL/ SLL (Landgren et al., 2009), and is subclassified into "low-count" MBL, defined as a peripheral blood monoclonal population < 0.5 × 10^9/L, and "high-count" MBL (Swerdlow et al., 2016). Low count MBL has a low risk of progression to CLL and, based on current evidence, does not require routine follow-up outside of standard medical care (Vardi et al., 2013). However, high count MBL has similar phenotypic and molecular features to early stage CLL and requires yearly follow-up (Vardi et al., 2013).

Monoclonal B lymphocytosis is reported in 13-18% of first degree relatives of F-CLL patients compared to 3-5% in the general population (Marti et al., 2003, Rawstron, 2002, Rawstron et al., 2002). These findings suggest that MBL is a marker of inherited predisposition to CLL. In this present study, it is proposed that germ line variation in genes or gene expression are acting early in neoplastic transformation of B lymphocytes, resulting in an MBL, and subsequent somatic oncogenic events are required before overt CLL develops.

1.3.4. Genetic Risk Factors

Molecular studies have shown that multiple genetic mutations are required for neoplastic transformation. Genetic mutations can either be inherited in the germline or arise somatically. A mutation within a gene that confers a selective growth advantage, promoting carcinogenesis, is termed a "driver mutation", while those that do not provide a growth advantage are "passenger mutations" (Vogelstein et al., 2013). Driver mutations are often somatic in origin, however, there are clear examples of germline driver mutations, including *BRCA1* and *BRCA2* in familial breast and ovarian cancer and *TP53* mutations in Li-Fraumeni

syndrome (Berchuck et al., 1998, Malkin et al., 1990). Driver mutations often occur in protein-coding regions of genes however it is increasingly recognised that non-coding mutations, such as splice-site or promoter mutations, can function as driver mutations (Sveen et al., 2016).

The development of CLL is likely to be similar to other cancers, in that a subset of cases occurs in individuals with germline mutations required for neoplastic transformation. In the majority of cancers, these genes are altered at the cellular level by random mutations and malignant transformation depends upon multiple genetic alterations (Knudson, 1971). This present study explores the proposal that alterations involved in the transformation of normal B lymphocytes to malignant CLL cells affect mRNA and protein expression, which in addition to DNA sequence variation, may result from epigenetic imprinting transmitted from generation to generation, or mutated transcription factors or promoters. The genetic contribution of DNA sequence variation to CLL risk has been studied using genetic linkage studies and genetic association studies, which includes candidate gene studies and GWAS, and it is increasingly recognised that mutations affecting gene regulation, including transcription factor binding and epigenetic modification, are involved in acquired and inherited susceptibility to neoplastic transformation of cells (Gazzoli et al., 2002, Esteller et al., 2001, Hedenfalk et al., 2001).

1.3.4.1. Genetic Linkage Studies

Clusters of cases in families have provided an indication of the inherited basis of CLL, and genetic linkage-based analyses have been performed to identify susceptibility genes for further study (Table 1-1). However, genetic linkage studies have been unable to detect any significant loci (Fuller et al., 2008, Raval et al., 2007), and the presence of many collaborating, low risk genes is likely to have limited the usefulness of this approach.

To address the rarity of very large multigenerational families, researchers have combined data from 2-4 affected subjects in multiple families (Goldin et al., 2003, Sellick et al., 2005b, Sellick et al., 2007, Ng et al., 2006, Ng et al., 2007b). However, since it is unlikely that each family have had the same group of susceptibility genes which map to the same disease loci, linkage signals are likely to have conflicted, resulting in an overall reduction in LOD scores. Consequently, most studies have been limited by weak evidence for linkage, and the identification of many loci has not shown concordance between studies (Table 1-1). Two loci at chromosome bands 11p11 and 13q21 have been supported by significant linkage scores; however no genes at either position have been implicated in the pathogenesis of CLL (Sellick et al., 2005b). (Table 1-1)

Table 1-1 Loci of maximum linkage in CLL genetic linkage studies

No. Families	NHL (n)	CLL (n)	Control (n)	Max. LOD	Chromosome/ Gene	Max. NPL	Chromosome/ Gene	Reference
28	0	63	92	1.32	1p22	-	-	(Ng et al., 2006)
206	44	304	0	-	-	2.84	2q22	(Sellick et al., 2007)
28	0	63	0	-	-	1.63	3q22	(Ng et al., 2006)
206	44	304	0	-	-	2.37	5q21	(Sellick et al., 2007)
206	44	304	0	-	-	2.44	6p22	(Sellick et al., 2007)
206	44	304	0	-	-	2.00	7q32	(Sellick et al., 2007)
1	0	6	0	-	-	-	9q21	(Raval et al., 2007)
115	16	228	0	2.78	11p11	3.14	11p11	(Sellick et al., 2005b)
28	0	63	92	-	-	2.81	12q24	(Ng et al., 2006)
28	0	63	92	-	-	1.78	13q21	(Ng et al., 2006)
6	0	19	44	1	13q22	-	-	(Ng et al., 2007b)
28	0	63	92	-	-	2.78	17p13	(Ng et al., 2006)
206	44	304	0	-	-	2.21	18q21	(Sellick et al., 2007)
1	0	4	18	0.8	2q37	2.24	14q24-31	(Fuller et al., 2008)

NHL; non-Hodgkin lymphoma, CLL; chronic lymphocytic leukaemia, LOD; logarithm of odds scores, NPL; nonparametric linkage

1.3.4.2. Genetic Association Studies

Following the development of high-throughput and relatively inexpensive genotyping technologies, case-control studies became the most widely used form of genetic association study, comparing sequence variation between germline DNA in a healthy control group and a CLL group.

The two methods used in genetic association are candidate gene studies and GWAS. Candidate-gene studies focus on a gene or set of genes that have biological plausibility in the pathophysiology of the CLL. In contrast, GWAS have no *a priori* assumptions and investigate for associations across the entire genome.

1.3.4.2.1. Candidate Gene Studies

Candidate genes for association studies have predominantly been selected for evaluation in association studies based on B lymphocyte and CLL biology, or because of roles in immune function, cell cycle, apoptosis, or DNA repair (Table 9-1). The majority of studies have used a limited number of genetic markers in a small number of genes and few associations have been replicated. *TNF* was one of the first genes to be studied following identification of an association between CLL and the TNF-α –308 promoter SNP, which was associated with increased TNF-α levels (Demeter et al., 1997). However, this association was not replicated in subsequent studies. A number of other candidate genes have similarly failed to show significant associations in replication studies, including *P2RX7, MTHFR, ARLTS1, BAX* and xenobiotic-metabolizing phenotypes. (Table 9-1)

SNPs in the co-stimulatory molecules *CTLA-4*, CD28 and *ICOS* have been found to be associated with CLL (Suwalska et al., 2008). An increased frequency of the *CTLA-4* promoter SNP, –319C>T, which upregulates the expression of CTLA-4, was reported however not

replicated. Similarly, an association has been demonstrated but not replicated between CLL risk and *CD28* 17+3 T>C (Suwalska et al., 2008). A significant association between CLL and *CD38* SNPs that increase B cell CD38 mRNA and percentage of CD38 positive B cells has been found, however requires validation (Jamroziak et al., 2009). Moreover, it should be noted that the majority of CLL candidate gene studies listed in (Table 9-1) have not been validated in independent studies.

Low sample sizes have limited the power to reliably identify genes that confer small risks for development of CLL. Sample sizes in the majority of CLL candidate gene association studies have usually not included greater than one to two hundred cases and controls, with the exception of Rudd *et al*.,(Rudd et al., 2006) Broderick *et al*.,(Broderick et al., 2008) Crowther *et al*.,(Crowther-Swanepoel et al., 2009) and Sellick *et al.*,(Sellick et al., 2008c).

As genotyping technologies increased in throughput and decreased in cost, disease pathway and multiple gene association studies were performed (De Roos et al., 2006, Hill et al., 2006, Shen et al., 2006, Nieters et al., 2006, Wang et al., 2006c, Wang et al., 2006b, Wang et al., 2006a, Lan et al., 2007, Lim et al., 2007, Cerhan et al., 2007, Enjuanes et al., 2008, Ennas et al., 2008, Gra et al., 2008, Ganster et al., 2009, Liang et al., 2009, Rudd et al., 2006) (Table 9-2), and promising candidate genes were identified in several pathways including apoptosis, DNA repair, immune response, oxidative stress and xenobiotic-metabolizing enzyme pathways. This focused genome-wide approach was used in an analysis of 865 candidate genes in 992 patients and 2707 controls to identify associations between susceptibility to CLL and DNA damage-response and cell-cycle pathway genes including *ATM*, *CHEK2*, *BRCA2* and *BUB1B* (Rudd et al., 2006). The most significant finding was an OR of 2.28 using a dominant model of inheritance for variants in *ATM* (Rudd et al., 2006). In another large study using SNPs selected from genes involved in cancer biology, associations were found for gene variants in

CCNH, *APAF1*, *IL16*, *CASP8*, *NOS2A*, and *CCR7* (Enjuanes et al., 2008).

1.3.4.3. Genome Wide Association Studies

Common genetic variations are estimated to comprise 46 - 59% of CLL heritability (Berndt et al., 2013, Di Bernardo et al., 2013). To detect low-risk loci located across the genome, genome-wide association studies (GWAS) of CLL have used large numbers of cases and controls and a dense coverage of SNPs (Di Bernardo et al., 2008, Speedy et al., 2014, Slager et al., 2010, Slager et al., 2011, Berndt et al., 2013, Law et al., 2017, Crowther-Swanepoel et al., 2010). (Table 1-2).

The first CLL GWAS was conducted in 505 CLL cases, 155 of which were F-CLL cases, potentially enriching the dataset for genetic susceptibility, and 1438 controls from the British 1958 Birth Cohort (Di Bernardo et al., 2008). This study provided the first evidence that multiple low-risk variants predisposed to developing CLL. Six loci were identified and validated on chromosomal bands 6p25.3, 11q24.1, 15q23, 2q37.1, 2q13, and 19q13.32 (Di Bernardo et al., 2008). Five of these loci, excluding the locus on 19q13.32, were subsequently validated in an independent study (Slager et al., 2010). A number of candidate genes are located at or close to these regions, including the interferon regulatory factor 4 (*IRF4*), GRAM domain containing 1B (*GRAMD1B)*, the nuclear body protein SP140 (*SP140*), and acyl-coenzyme A oxidase-like (*ACOXL*) (Slager et al., 2010). IRF4 is a key regulator of lymphocyte development and proliferation, and changes in its expression have been previously linked to the development of CLL (Shukla et al., 2013). Furthermore, a dose relationship was found between *IRF4* mRNA in EBV-transformed lymphocytes and its SNP genotype, consistent with a model in which the causal variant contributes to risk by preventing transition of memory B cells through decreased IRF4 expression (Di Bernardo et al., 2008).

A second GWAS, with validation in 4 additional series totalling 2,503 cases and 5,789 controls, identified 4 additional risk loci at 2q37.3 (*FARP2*), 8q24.21, 15q21.3, and 16q24.1 (Crowther-Swanepoel et al., 2011). In addition to these 4 loci, there was suggestive evidence for disease loci at 15q25.2 and 18q21.1, and a validation study using 1428 cases and 1920 controls found an association between these 2 loci and CLL risk (Crowther-Swanepoel et al., 2011).

To identify risk loci specific to familial CLL, a third GWAS used a case group enriched with F-CLL and F-MBL cases (Slager et al., 2011). This study of 407 CLL cases, included 102 with a family history of CLL, and 296 controls. Four SNPs were identified that met genome-wide statistical significance within the *IRF8* (interferon regulatory factor 8) gene, located at the previously identified 16q24.1 locus. Within F-CLL cases, a susceptibility locus was identified at 6p21.3 (Slager et al., 2011), which includes the *HLA-DQA1* and *HLA-DRB5* genes. Within the CLL families, 60 F-MBL cases were evaluated for associations with the initially reported loci, and significant associations were found for 2q37.1 and 6p21.3 (Slager et al., 2011).

Following 3 GWAS and 1 case control study (Crowther-Swanepoel et al., 2010, Crowther-Swanepoel et al., 2011, Di Bernardo et al., 2008, Slager et al., 2011), 13 loci associated with risk of CLL had been identified. Using genotyping data from these studies and a 4th GWAS of CLL using cases and controls from 22 studies of non-Hodgkin lymphoma, a meta-analysis was performed on 3100 CLL cases and 7667 controls (Berndt et al., 2013). This study found associations with 9 loci located at 10q23.31 (*ACTA2/FAS*), 18q21.33 (*BCL2*), 11p15.5 (*C11orf21*), 4q25 (*LEF1*), 2q33.1 (*CASP10/CASP8*), 9p21.3 (*CDKN2B-AS1*), 18q21.32 (*PMAIP1*), 15q15.1 (*BMF*), 2p22.2 (*QPCT*), and one independent SNP in an established locus at 2q13 (*ACOXL/BCL2L11*). Several of these loci are located in or near genes involved in regulating apoptosis including *FAS*, *BCL2*, and phorbol-12-myristate-13-acetate-induced protein 1

(*PMAIP1* or *NOXA*). *NOXA* has been identified as a critical factor for B cell expansion after antigen triggering and suppression in lymph node environments has been associated with persistence of B-CLL (Wensveen et al., 2012) (Berndt et al., 2013). The B-cell lymphoma-2 modifying factor (*BMF*) at 15q15.1 has also been implicated in apoptosis, binding BCL-2 protein, and to have a role in regulating growth and survival in normal B cells and B-CLL cells (Morales et al., 2004).

The most recently published GWAS and meta-analysis, comprised 1,739 CLL cases and 5,199 controls with validation in 1,144 CLL cases and 3,151 controls (Speedy et al., 2014). Associations were found for 3q26.2 (*MYNN*), 4q26 (*CAMK2D*), 6q25.2 (*IPCEF1*) and 7q31.33 (*POT1*), and two previously established loci within 5p15.33 (*CLPTM1L* and *TERT*) and 8q22.3 (*ODF1*) (Speedy et al., 2014). Telomerase (*TERT1*), which synthesises the telomere ends of linear chromosomes (Morin, 1989), has been implicated in cancer cell pathogenesis (Hahn et al., 1999), and the protection of telomeres protein 1 (*POT1*) functions to maintain telomeres, regulate telomere lengths and stabilize chromosome ends. *POT1* somatic point mutations have been demonstrated in 3.5% of CLL cases and 9% of UM-CLL (Ramsay et al., 2013).

To date, GWAS have identified 30 risk variants at 26 different regions of the genome, however these account for only an estimated 19% of CLL heritability (Speedy et al., 2014). This suggests that a proportion of the remaining inherited CLL risk may be associated with non-DNA sequence-linked heritable information, including epigenetic modifications, which have been found to regulate oncogenes and tumour suppressor genes in a number of hereditary cancers (Gazzoli et al., 2002, Esteller et al., 2001).

Table 1-2. Summary of genome wide association studies (GWAS) in CLL

CLL (n)	Controls (n)	Markers	Loci	Gene/Notes	Reference
1529	3115	299 983	2q13 (p=2.36x10^{-10})	*ACOXL*	(Di Bernardo et al., 2008)
			2q37.1(p=5.40x10^{-10})	*SP140*	
			6p25.3 (p=1.91x10^{-20})	*IRF4*	
			11q24.1 (p=3.78x10^{-12})	50kb telomeric to *GRAMD1B*	
			15q23 (p=4.54x10^{-12})	*PRKD2*	
			19q13.32 (p=3.96x10^{-9})		
2503	5789	299 983	2q37.3 (OR=1.39; p=2.11x10^{-9})	*FARP2*	(Crowther - Swanepoel et al.)
			8q24.21 (OR=1.26 ; p=7.84x10^{-10})	Locus with no genes	
			15q21.3 (0R=1.36 ; p=4.74x10^{-7})	*NEDD4* and *RFX7*	
			16q24.1 (OR=1.22 ; p=3.60x10^{-7})	*IRF8*	
407	296	934 968	16q24.1 (OR= 1.3-1.8, combined P values < 3.37 x 10^{-8})	*IRF8*	(Slager et al., 2011)
			6p21.3 (OR= 1.6, combined P value= 6.92 x 10^{-9}).	*HLA-DQA1 and HLA-DRB5*	
1,739	5,199	450,000	3q26.2 (P = 1.74 × 10^{-9})	*MYNN*	(Speedy et al., 2014)
1,144 (validation)	3,151 (validation)		4q26 (P = 3.07 × 10^{-9})	*CAMK2D*	
			6q25.2 (= 1.50 × 10−10)	*IPCEF1*	
			7q31.33 (P = 3.40 × 10^{-8})	*POT1*	
			5p15.33 (P = 1.72 × 10^{-7})		
			8q22.3 (P = 2.90 × 10^{-9}).	*CLPTM1L and TERT*	
				ODF1	

1.3.4.4. Combined Molecular Studies

Another strategy used to identify susceptibility loci in CLL families has been to combine epigenetic, genetic and proteomic studies. This approach identified downregulation of death-associated protein kinase 1 (*DAPK1*), as a susceptibility gene in a CLL family and in sporadic CLL patients (Raval et al., 2007).

DAPK1 was initially identified as an inducer of apoptosis after activation by IFN-γ (Deiss et al., 1995). Using a quantitative high-throughput analysis, DNA methylation levels of *DAPK1* in sporadic CLL samples were found to be significantly different to normal B lymphocytes. In parallel, a genetic linkage study in a family with 7 F-CLL cases found the highest nonparametric linkage (NPL) score of 0.96 at a locus on chromosome 9 and identified a common haplotype in all affected family members that included *DAPK1* (Raval et al., 2007, Lynch et al., 2002). RT-PCR analysis showed *DAPK1* mRNA levels were lower in F-CLL family members compared to unaffected, and lymphoblastoid cells from 1 F-CLL case were used to generate monoallelic clones in which *DAPK1* gene and protein expression were reduced. Germline DNA variants were detected in *DAPK1*; however, none were unique for the CLL phenotype or identified in additional families. Nevertheless, these *DAPK1* variants showed reduced gene and protein expression in mono-chromosomal hybrid clones and *DAPK*1 was found to be epigenetically silenced in most (89%) sporadic CLL cases suggesting a role in CLL pathogenesis, and highlighting a putative pathway for development of targeted therapies (Raval et al., 2007).

1.3.5. Cytogenetic Abnormalities

For many years, chromosome breakage syndromes have been known to be associated with an increased risk of leukemia (German, 1980). Acquired chromosomal abnormalities are found in over 80% of CLL cases and are independent predictors of disease progression and

survival (Dohner et al., 2000). The commonest somatic alteration is an interstitial deletion in 13q14.3, which is found in over 50% of CLL cases and is associated with a favourable prognosis. The next most common chromosomal aberration, found in 20% of cases and associated with a progressive course, is deletion 11q22-q23 with consequent loss of the *ATM* tumour suppressor gene. In frequency, these chromosomal aberrations are followed by trisomy 12 (15%) and deletion 17p13.1 [del (17p13.1)] (5-10%), with deletion of *TP53* predicting a very poor outcome (Dohner et al., 2000). The genes affected by, and associated with, these chromosomal aberrations are summarized in (Table 1-3), and described in the following sections. In addition to somatic mutations, in the last several years, population and family studies have identified a number of germ-line genetic mutations that increase the risk of leukaemia in carriers. Leukaemia susceptibility had been primarily associated with Li Fraumeni syndrome, Fanconi anaemia, dyskeratosis congenita, and trisomy 21 (Garriga and Crosby, 1959, Li and Fraumeni, 1969), however an increasing number of germ-line mutations have been shown to be associated with leukaemia predisposition (Porter, 2016). CLL has been reported in families with ataxia- telangiectasia (A-T) (Swift et al., 1987), suggesting that heterozygotes for mutations in the *ATM* gene may be at an increased risk.

Table 1-3 A summary of the most frequent mutations in CLL

Gene Mutation	Association with	Functional Role	Type of mutation	References
IGH		Mutated IGH (M-CLL) typically has longer survival and better prognosis compared to unmutated (UM-CLL).		(Sagatys and Zhang, 2012)
TP53	Del (17p13)	Critical role in the cell cycle regulatory networks including DNA repair and apoptosis	inactivation	(Shahjahani et al., 2015)
ATM	Del (11q23)	Activates cell cycle checkpoints, and initiates apoptosis in response to DNA double-strand breaks	inactivation	(Inamdar and Bueso-Ramos, 2007) (Austen et al., 2005) (Austen et al., 2007)
NOTCH1	Trisomy 12 and UM- IGHV	Regulates target genes, such as MYC, TP53 and other molecules that are involved in the nuclear factor-kappa B (NF-kB) pathway	activation	(Puente et al., 2011) (Rosati et al., 2009)
		Plays an essential role in cell differentiation, proliferation, and apoptosis.		(Del Giudice et al., 2012)
FBXW7	Trisomy 12	Acts as a tumour suppressor gene, targeting NOTCH1 and other onco-proteins such as MYC and cyclin-E1	activation	(Wang et al., 2011) (Falisi et al., 2014)
SF3B1	Del (11q22-q23), ATM mutations, UM- IGHV,	A core component of the spliceosome.	inactivation	(Matera and Wang, 2014) (Wan and Wu, 2013) (Rodriguez-Vicente et al., 2013)
BIRC3	Del (11q22-q23)	Has a negative regulatory function for the nuclear factor-kappa B (NF-kappa-B) signalling cascades through its inhibition of mitogen-activated protein kinase-kinase 14 (MAP3K14).	inactivation	(Campregher and Hamerschlak, 2014). (Rossi et al., 2012).

1.3.5.1. Deletion 11q22-q23 in S-CLL and F-CLL

Deletion 11q [del (11q22-23)] can be detected in ≈20% of S-CLL cases (Dohner et al., 2000). Del (11q22-23) has been associated with advanced CLL patients, B-symptoms, extensive lymphadenopathy, rapid lymphocyte doubling times and shorter survival (Rodriguez-Vicente et al., 2013). There are 2 genes, *ATM* and *BIRC3*, at this locus which are involved in the pathogenesis of CLL (Schaffner et al., 1999, Chiaretti et al., 2014).

1.3.5.1.1. Ataxia-Telangiectasia Mutated Gene (ATM)

ATM gene mutations are a frequent event in CLL and occur as a monoallelic loss in del (11q22-23), with and without mutations in the remaining *ATM* allele. Inactivation of *ATM* prevents B-CLL from responding to DNA double-strand breaks, causing genomic instability (Schaffner et al., 1999). CLL with del (11q22-23) can be divided into two subgroups based on the presence or absence of mutations in the residual *ATM* allele. The residual *ATM* allele is mutated in 36% of CLLs with del(11q22-23) and these leukemias demonstrate impaired cellular responses to genotoxic damage *in vitro* (Austen et al., 2007). CLL patients with del (11q22-23) are at risk of developing a mutation in the remaining ATM allele, which leads to rapid clonal expansion and reduced survival (Austen et al., 2007). In these patients, use of treatments that bypass the ATM/p53 apoptotic pathway, has been proposed as an alternative to fludarabine, cyclophosphamide and rituximab (FCR) therapy (Kojima et al., 2006, Lozanski et al., 2004)

There is an association between deletion of *ATM* and p53 dysfunction (Carter et al., 2006), and inactivation of *ATM* has been reported in one-third of CLL patients with mutated *TP53* (Stankovic et al., 2002) (Pettitt et al., 2001). The frequency of p53 dysfunction observed among del(11q22-23) cases is ≈60% (Carter et al., 2006).

Del (11q22-23) has been associated with unmutated *IGH*, and newly diagnosed untreated CLL can exhibit variable clinical outcomes (Marasca et al., 2013). To evaluate the correlation between percentages of cells displaying del (11q22-23) and clinical outcome, a cut-off point of 25% positive nuclei was defined for CLL patients at risk of short time to first treatment (TTFT) (Marasca et al., 2013). CLL cases with ≥ 25% positive cells had a median TTFT of 14 months (Marasca et al., 2013). A study of mutated and deleted *ATM* cases, in the absence of 11q22-23 deletion, showed differences in gene expression profiles, and mutations in this gene were associated with an unfavourable clinical course and shorter-treatment intervals (Guarini et al., 2012).

ATM mutations occur at different stages of CLL development and can be present in the germline (Austen et al., 2005, Bullrich et al., 1995, Austen et al., 2007). Ataxia telangiectasia (A-T) patients have an increased risk of lymphoma and leukemia, and A-T heterozygotes may have an increased risk of breast cancer (Athma et al., 1996, Taylor et al., 1996). In a retrospective study of cancer incidence in 110 A-T families, the risk of hematological and lymphoid malignancies was increased in relatives of A-T patients and CLL accounted for 5 of 13 cases (Swift et al., 1987), although this did not attain statistical significance.

It has been demonstrated that some CLL patients have heterozygous *ATM* germline mutations (Schaffner et al., 1999, Stankovic et al., 1999, Starostik et al., 1998). In a series of 32 cases of B-CLL, two germline *ATM* mutations were found in 6.3% of the cases (Stankovic et al., 1999). In a linkage analysis of 28 families, the observed distribution of *ATM* shared haplotypes between affected individuals supported the proposal that *ATM* is involved in F-CLL. However, assuming ≈6.3% of CLL is caused by *ATM* (Stankovic et al., 1999), mutations in this gene should confer a sibling relative risk of 1.1 (Bevan et al., 1999). On this basis, it was concluded that the study numbers gave insufficient power to detect linkage between *ATM*

and CLL (Bevan et al., 1999).

More recently, *ATM* germ-line mutations have been found to be associated with later events in CLL pathogenesis rather than clonal initiation (Skowronska et al., 2012). To explore the presence of clonal and sub-clonal mutations of *ATM*, and mutations carrying a negative prognosis in *TP53*, *SF3B1*, *NOTCH1*, and *BIRC3* in 406 untreated CLL cases, a study was performed using next generation sequencing (NGS) to define the evolution of sub-clones at different time points, and determine their influence on outcome (Nadeu et al., 2016). In *ATM,* 126 variants were found in 95 patients, with 53 mutations classified as somatic and 73 as germ-line. The germ-line variants were classified as definitely (n = 8) or likely (n = 2) pathogenic, rare missense (n = 33), variants of unknown significance (n = 12), and polymorphisms (n = 18). Relevant to F-CLL, 4/9 cases (44%) had germ line pathogenic mutations, but only 3/53 (6%) with non-pathogenic variants had acquired 11q deletions ($P < 0.01$), which suggested a role of the germline variants in disease progression via deletion of the remaining allele (Nadeu et al., 2016).

In conclusion, *ATM* is a credible candidate predisposition locus for CLL, although there is currently insufficient evidence to unambiguously show that *ATM* mutations are involved in the development or progression of CLL.

1.3.5.1.2. Baculoviral IAP Repeat-Containing Protein 3 (BIRC3)

The baculoviral IAP repeat-containing protein 3 (*BIRC3*) gene, located at 11q22.2, is affected by del (11q22-23). *BIRC3* belongs to a family of proteins that inhibit apoptosis by binding tumour necrosis factor (TNF) receptor-associated factors TRAF1 and TRAF2 (Uren et al., 1996). It has been shown that BIRC3 has a negative regulatory function for the nuclear factor kappa-light chain enhancer of activated B cells (NF-kappa-B) signalling cascades through

inhibition of mitogen-activated protein kinase-kinase-kinase 14 (MAP3K14) (Matsuzawa et al., 2008). Therefore any abnormality affecting the BIRC3 pathway increases NF-kappa-B activity, leading to increased cell survival (Campregher and Hamerschlak, 2014).

BIRC3 mutations have been reported in UM-CLL patients (Chiaretti et al., 2014). Although *BIRC3* mutations are rare in CLL patients at diagnosis (≈4%), mutations are associated with 24% of fludarabine-refractory CLL (Rossi et al., 2012). Similar to *ATM*, a subset of missense variants detected using NGS in tumour samples are present in the germ-line (Nadeu et al., 2016).

1.3.5.1.3. Mutations Associated with del (11q22-23)

Mutations in *SF3B1*, which is located at 2q33.1, have been correlated with del (11q22-23) (Wang et al., 2011), although this correlation has not been consistently replicated (Oscier et al., 2013, Rossi et al., 2011). In a large study of 1160 patients without prior treatment, *SF3B1* mutations were associated with del (11q22-23) (20.3%), ≥ 30% CD38 expression (14.6%), unmutated *IGH* (15.3%), shorter time to treatment (median 3.8 years), and reduced 5 years overall survival (Jeromin et al., 2014).

Splicing factor 3B, subunit 1 is an essential component of the spliceosome machinery, which consists of uridine-rich small nuclear RNAs (snRNAs) and a large number of associated protein factors that are assembled to form ribonucleic-proteins (snRNPs) (Isono et al., 2005). The spliceosomal complex has a critical role in many cellular functions, including gene expression and regulation, and post-transcriptional processing of mRNA (Wahl et al., 2009).

Although *SF3B1* has been implicated in the pathogenesis of CLL (Quesada et al., 2012), its precise role remains unclear. *SF3B1* is mutated in ≈5% of CLL patients at initial diagnosis and increases to 17% in patients with fludarabine-refractory disease (Rossi et al., 2011). The

majority of somatic *SF3B1* mutations in CLL cases have been in the C-terminal regions of the gene. Mutations are predominantly missense or rarely frame-shift mutations causing insertions or deletions (Wang et al., 2011) (Rossi et al., 2011). There is also evidence of abnormal RNA splicing in CLL patients with mutated *SF3B1* when compared with wild-type *SF3B1* (Rossi et al., 2011) (Wang et al., 2011). All *SF3B1* mutations detected to date have been confirmed as somatic, and it is unlikely that there are inherited mutations in *SF3B1* predisposing to F-CLL (Nadeu et al., 2016).

1.3.5.2. Deletions at 13q14 and microRNAs

Deletion 13q14 [del (13q14)] is the most common genomic aberration in CLL (Dohner et al., 2000), and is often associated with M-CLL. Patients with del (13q14) have a better prognosis and longer survival compared to those with other cytogenetic abnormalities (Inamdar and Bueso-Ramos, 2007).

Improvements in DNA sequencing technologies have defined the minimal deleted region (MDR) at the 13q14 locus, which contains 8 genes: retinoblastoma (*RB1*), deleted in lymphocytic leukaemia 1 (*DLEU1*), 2 (*DLEU2*) and 5 (*DLEU5*) genes, CLL deletion 6 (*CLLD6*), 7 (*CLLD7*), and 8(*CLLD8*) genes, karyopherin α-3 (*KPNA3*), and microRNA (*miR-15* and *miR-16-1*) genes (Lia et al., 2012). Mono-allelic losses of *RB1* are seen in > 30% of B-CLL (Liu et al., 1992), however germline mutations in *RB1*, seen in the heritable genetic form of retinoblastoma, have not been reported to be associated with F-CLL.

miR-15a and *miR-16-1* are deleted or down-regulated in > 50% B CLL patients (Humplikova et al., 2013). Both are located at a cluster within intron 4 of *DLEU2* (Klein and Dalla-Favera, 2010). In mice studies, deleting the MDR, which encodes *DLEU2* and miR-15a/16-1, induced more aggressive disease compared to mice with deleted microRNA-15a/16-

1 cluster alone (Klein et al., 2010). In addition, targeted deletions are associated with the development of clonal B cell lymphoproliferative disorders including MBL, CLL/SLL, and NHL (Klein et al., 2010). The presence of germline mutations in *miR-16-1* and *miR-15a*, or *DLEU2*, and potential role in F-CLL have not been studied.

Another candidate gene found in the MDR at 13q14 is *ARLTS1*. A nonsense SNP in this gene was associated with a family history of cancer and/or multiple personal cancers, and the SNP was subsequently found in 2/17 F-CLL cases (Calin et al., 2005). However, a replication study failed to find an association between CLL and this *ARLTS* SNP or 5 other nonsynonymous SNPs in a cohort with 413 familial cases (Sellick et al., 2006b). A second replication study found this SNP in 2/31 F-CLL cases but failed to segregate with other cases in these families, and was not over-represented in LPD families compared with controls (Summersgill et al., 2002). Therefore, there is limited evidence for the involvement of inherited mutations in *ARLTS1* in the development of F-CLL.

1.3.5.3. Deletion 17p13.1

Loss of genetic material on chromosome 17p in CLL generally includes band 13.1, where the tumour suppressor gene *TP53* is located. Del (17p13) is found in ≈3-8% of treatment naïve CLL cases, but the prevalence increases to 50% in relapsed and/or refractory CLL (Wang and Wang, 2013). The presence of del (17p13) is associated with a more aggressive course and resistance to chemotherapeutic agents compared to patients with other cytogenetic mutations (Shahjahani et al., 2015). However, not all *de novo* patients with del (17p13) progress rapidly, with some having an indolent course which progresses in association with secondary acquired mutations (Tam et al., 2009). Untreated del (17p13)with low-risk factors, such as mutated *IGH* and Rai stage 0, have a significantly longer overall survival (4-5 years)

compared to those with high-risk factors, including unmutated *IGH* and Rai stage 1 or higher (median survival 1 to 1.5 years) (Tam et al., 2009).

1.3.5.3.1. TP53

The tumour suppressor protein p53, which is encoded by *TP53*, has a critical role in cell cycle regulation networks including DNA repair and apoptosis (Unger et al., 1992, Artandi and Attardi, 2005). Under normal conditions, p53 is tightly regulated and functionally inactivated by interaction with mouse double minute homologue 2 (MDM2), which transfers p53 from the nucleus to the cytoplasm where it undergoes ubiquitination and degradation in the proteasome (Fuchs et al., 1998). Disruption of these pathways increase and activate p53 (Yin et al., 2002).

The p53 activation process is complex and involves directly, or indirectly, multiple mechanisms that are essential for cell cycle arrest, DNA repair and inducing apoptosis if DNA repair fails (Vousden and Lane, 2007). Signalling to p53 occurs via separate pathways in response to genetic damage or activation of oncogenes (Efeyan and Serrano, 2007). In response to expression of oncogenes, p53 is stimulated via the p53- stabilising protein, cyclin-dependent kinase inhibitor 2A (*CDKN2*) locus alternative reading frame (p14ARF), which in turn, interacts with MDM2, inhibiting the E3 ubiquitin ligase interaction, which protects p53 from degradation. This results in accumulation of p53 in tumour cells, however, p53 activity is lost through the disruption of p53/MDM2 interactions (Efeyan and Serrano, 2007).

In response to cellular stress signals, such as double-strand DNA damage induced by radiation, cytotoxic agents, or radiotherapy, ATM/CHK2 pathways are activated. This leads to an increase in the downstream activities of p53 (Joerger and Fersht, 2008). If DNA repair fails, p53 initiates the intrinsic apoptotic cell death mechanisms via the B-cell leukemia lymphoma

2 (Bcl-2) family and the caspase cascade.

Del (17p13) and *TP53* mutations correlate with poor outcomes in most CLL patients (Strati et al., 2014). Rossi *et al.,* (Rossi et al., 2013) mentioned that the majority (≈60%) of CLL cases show mono-allelic del (17p13) and a point mutation of the second *TP53* allele, while the remaining cases have either del (17p13) with no TP53 mutation (≈10%) or TP53 mutations without del (17p13) (≈30%). Some studies report that CLL cases that have mutations in *TP53* without del (17p13) showed clinical, pathological and prognostic features similar to del (17p13) patients (Zenz et al., 2010a). However, other studies found that although the subgroup with isolated *TP53* mutations was small, it remains an independent predictor of rapid disease progression (Zenz et al., 2008) (Rossi et al., 2009).

Chromosomal aberrations can be defined by fluorescence in situ hybridization (FISH) and chromosome banding analyses. The FISH assay utilises DNA specific probes for the detection of *TP53* at 17p13.1 and *ATM* at 11q22. FISH studies increasingly provide clinically relevant prognostic information to identify patients who may be targeted for earlier treatment (Delgado et al., 2012). To predict overall survival (OS), initial studies defined the cut-off level for *TP53* deletion at 3% positive cells (Dohner et al., 1995), and 20% (Catovsky et al., 2007). However, a cut-off level of 25% positive cells was found to accurately predict rapid disease progression and guide initiation of therapy (Tam et al., 2009) (Delgado et al., 2012).

The development of dense-coverage SNP arrays and next-generation sequencing (NGS) techniques have provided the potential to understand the molecular basis of CLL, including the role of *TP53* (Chiorazzi, 2012) (Foa et al., 2013). Using NGS, *TP53* mutations in CLL have been found to be mostly missense mutations that result in amino acid substitutions in the sequence-specific DNA binding domain (DBD) of p53 (Foa et al., 2013). The majority

(74%) of these mutations are located in exons 5,6,7 and 8 (Chiorazzi, 2012). However, other less common genetics events affecting *TP53* have also been demonstrated in CLL including nonsense mutations (4%), frame-shift mutations (20%) and splice site mutations (2%) (Zenz et al., 2010b) (Lin et al., 2013).

All *TP53* mutations detected to date in CLL have been confirmed as somatic, and it is unlikely that there are inherited mutations in *TP53* that predispose to F-CLL (Nadeu et al., 2016). Leukemia accounts for 3 - 6% of Li Fraumeni syndrome (LFS) tumours (McBride et al., 2014) (Malkin et al., 1990). Leukaemias associated with LFS include acute lymphoblastic leukaemia, acute myeloid leukaemia, and therapy-related acute myeloid leukaemia (Hof et al., 2011, Salmoiraghi et al., 2016, Schulz et al., 2012). However, germ-line mutations in *TP53* have not been reported in association with CLL.

1.3.5.4. Trisomy Chromosome 12

Trisomy 12 is the third most common chromosomal aberration found in CLL, occurring in ≈10-20% of cases (Dohner et al., 2000) (Chiorazzi, 2012). In earlier studies, trisomy 12 was found to be associated with atypical lymphocyte morphology and immunophenotypic features including high CD38, CD20, and FMC7 expression (Inamdar and Bueso-Ramos, 2007). Additionally, trisomy 12 was associated with UM-CLL and an aggressive clinical course (Inamdar and Bueso-Ramos, 2007). However, recent studies have considered trisomy 12 as an intermediate risk or even low risk marker (Del Giudice et al., 2012) (Puiggros et al., 2014).

A minimal common gained region in chromosome 12 has been identified that spans bands 12q13 to 12q15 (Jimenez-Zepeda et al., 2013). This region contains *MDM2*, which plays an essential role in the regulation of cell growth and death, and has been reported to be overexpressed in B-CLL (Watanabe et al., 1996). Another band on chromosome 12q22

contains chronic lymphocytic leukemia up-regulated, 1 (*CLLU1*), which is overexpressed in patients <70 years of age at diagnosis and is associated with a poor prognosis (Rodriguez-Vicente et al., 2013). This gene may be involved in the pathogenesis of CLL, however the underlying biology of this association is unknown (Rodriguez-Vicente et al., 2013) (Inamdar and Bueso-Ramos, 2007).

Trisomy 12 can occur in isolation in ≈70% of cases, or in combination with other chromosomal abnormalities, including trisomy 18 and trisomy 19, and deletions of chromosomes 13q, 14q, 11q or 17p (Puiggros et al., 2014) (Del Giudice et al., 2012) (Nguyen-Khac et al., 2013).

1.3.5.4.1. Mutations Associated with Trisomy 12

1.3.5.4.1.1. F-box/WD Repeat Containing Protein 7

F-box and WD40 domain protein 7 (*FBXW7*) gene mutations have been identified in patients with CLL and trisomy 12. Fbw7, the protein product of *FBXW7*, is a member of the E3 ubiquitin ligase complex and functions as a tumour suppressor gene, targeting Notch1 and other oncoproteins such as Myc and cyclin-E1 (Welcker et al., 2004). Fbw7 has a negative regulatory role in the Notch signalling pathway by binding the PEST domain, leading to rapid degradation of the Notch intracellular domain. Mutations in *FBXW7*, are associated with T-ALL development, and have been reported in ≈4% of CLL patient samples with trisomy 12 (Falisi et al., 2014). These observations suggest that mutations in *FBXW7* are associated with *NOTCH1* mutations and may play an important role in the disruption of *NOTCH* signalling in patients with trisomy 12 (Falisi et al., 2014). However, the precise role of these mutations in CLL patients carrying trisomy 12 requires further investigation (Wang et al., 2011) (Falisi et al., 2014).

1.3.5.4.1.2. NOTCH1

The *NOTCH1* gene is located on chromosome 9q34.3 and encodes a single-pass transmembrane receptor that regulates cellular differentiation, proliferation, and apoptotic programs (Das et al., 2004). The Notch family includes 4 receptors, *NOTCH1*, *NOTCH2*, *NOTCH3*, and *NOTCH4*. The Notch1 signalling pathway is activated following binding of its ligands, from the Jagged or Delta families, on the extracellular membrane of the target cell (Artavanis-Tsakonas et al., 1995). Binding activates multiple proteolytic cleavages and release of the Notch1 intracellular domain (NICD), which translocates to the nucleus where it interacts with multiple transcription factors (Artavanis-Tsakonas et al., 1995). Once these interactions become active, the transcription process of Notch is stimulated in different target genes, including *MYC*, *TP53* and other molecules involved in the nuclear factor kappa B (NF-kB) pathway (Arruga et al., 2014) (Rossi et al., 2013).

Deregulation of the Notch pathway has been associated with T cell acute lymphoblastic leukaemia (T-ALL) and CLL (Rodriguez-Vicente et al., 2013). In CLL, constitutive activation of Notch1 prevents apoptosis and facilitates prolonged cell survival (Rosati et al., 2009). Mutations in *NOTCH1* mainly occur in the PEST domain, which is rich in proline (P), glutamic acid (E), serine (S), and threonine (T) amino acids. This leads to the accumulation of stable, active Notch1 isoforms in CLL cells and constitutive activation of the Notch1 signalling cascade (Puente et al., 2011) (Rosati et al., 2009). In a study of 391 B-CLL patients, 104 *NOTCH1* mutations were found in 86 (22%), and most occurred in UM-CLL (82%) (Nadeu et al., 2016). Similar to *BIRC3*, *NOTCH1* missense variants detected in tumour samples were present in the germline, raising the possibility that inheritance of germline *NOTCH1* variants may contribute to F-CLL.

1.3.5.5. Chromosome 6 and Other Rare Chromosomal Abnormalities

Other recurrent chromosomal abnormalities have been described in CLL including deletion 6q and trisomy 8 (Dohner et al., 2000) (Cuneo et al., 2004). Deletion 6q has been reported in ≈ 6% of CLL patients (Dohner et al., 2000), and has been categorized as an intermediate risk-group, with high WBC count, classical immunophenotype, positive CD38 and no association with *IGH* mutation status (Cuneo et al., 2004).

1.3.5.6. Chromosome 14

A small number of studies have implicated Chromosome 14 in B-CLL, and a linkage scan performed on a family with multiple cases of F-CLL by Fuller *et al*. found maximum linkage to a 200kb region of chromosome 14 between 14q24.1 and 14q31.2 which contains 175 genes with open reading frames encoding known and hypothetical proteins (Fuller et al., 2008). A candidate gene at this locus, *ZFP36L1*, has been linked to the apoptotic response to rituximab in CLL (Jackson et al., 2006, Baou et al., 2009), however no germ line variant was detected in the family. This family will be the subject of the gene and protein expression study reported in this thesis (Figure 1-1).

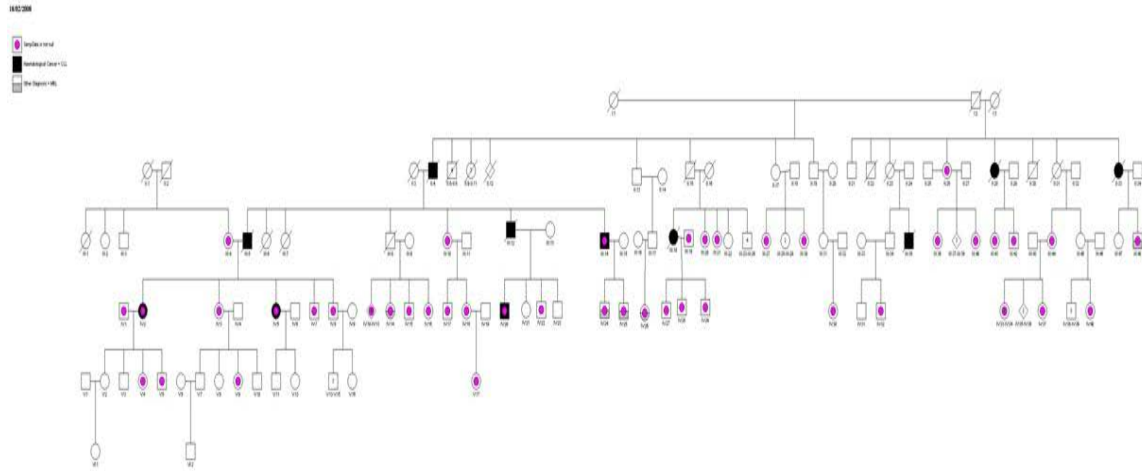

Figure 1-1. Pedigree of the family in abbreviated form

Pedigree of the family in abbreviated form showing segregation of chronic lymphocytic leukemia (CLL). Shaded symbols represent CLL cases, half-shaded MBL, and symbols with shaded pink circles, individuals from whom DNA was collected for the genetic linkage study reported by Fuller et al., (Fuller et al., 2008).

Chromosome 14 abnormalities are found in 6 - 14% of CLL cases. The most common region is located at band 14q32, which is the locus for the immunoglobulin heavy chain (*IGH*) gene, and rearrangements of this locus have been identified in mature B cell lymphomas. A translocation between the cyclin D1 (*CCND1*) gene at 11q13 and *IGH* at 14q32, t(11;14)(q13;q32), is the most common translocation in mantle cell lymphoma (Inamdar and Bueso-Ramos, 2007). A translocation involving *IGH* with the *BCL3* gene at 19q13, t(14;19)(q32;q13), has been associated with atypical CLL, which has 10% or less prolymphocytes, a younger age at onset and a progressive disease course (Inamdar and Bueso-Ramos, 2007) (Huh et al., 2011). Other translocations involving *IGH* have been reported with *BCL2* at 18q21, t(14;18)(q32;q21), in most cases of follicular lymphomas, and *MYC* at chromosome 8 in Burkitt's lymphoma patients (Martin-Subero et al., 2007).

The presence of chromosomal abnormalities involving 14q32 and *IGH* has been studied in 252 chronic lymphocytic leukaemia cases (Cavazzini et al., 2008). Translocations involving

14q32/*IGH* were identified as the sole aberration in 8/18 patients. Of these, 5 had a BCL2/*IGH* rearrangement while the remaining 3 cases had fusions of *IGH* with *BCL11A*, *CCND3* and *CDK6* (Cavazzini et al., 2008). This study also compared isolated 14q32/*IGH* translocation with the remaining cytogenetic risk groups: favourable risk [del (13q) and non-detectable chromosome abnormalities]; intermediate risk [trisomy 12, del (6q) or 1-2 detectable abnormalities]; and unfavourable risk [del (11q), del (17p) and complex abnormalities]. The results showed that CLL patients carrying the 14q32/*IGH* translocation have a shorter treatment-free survival (TFS = 2 months) compared to those with intermediate risk (TFS = 12months) or favourable risk (TFS = 20 months). A shorter overall survival was also observed for 14q32/*IGH* (OS = 18 months) compared to intermediate (OS = 50 months) and favourable risk groups (OS > 60 months). This study found that the 14q32/*IGH* predicted an unfavourable outcome, which was improved compared to del (11q), del (17p) and complex abnormalities, but worse compared to patients with del (13q), no detectable chromosome abnormalities, trisomy 12, del (6q) or 1-2 detectable abnormalities (Cavazzini et al., 2008).

1.3.6. Proteomic Profiling in B-CLL

The protein expression profiles of B-CLL compared to controls have been studied using a number of different proteomic methods (Alsagaby et al., 2014, Huang et al., 2016, Eagle et al., 2015, Perrot et al., 2011). However, most studies have aimed to identify protein markers which predict prognosis rather than identify proteins associated with neoplastic transformation, and only one proteomic study, which identified an association with downregulation of *DAPK1,* has been reported in F-CLL (Raval et al., 2007).

An indication that gene expression plays a role in the development of B-CLL has been the identification of differential expression of histones compared to normal B lymphocytes. A

proteolytic product of histone H2A (cH2A) has been found to be differentially abundant in B-CLL samples (Glibert et al., 2014), and histone profiles in B-CLL compared to control B lymphocytes has identified increased expression of a specific histone H2A isoform (H2A 1C) (Singh et al., 2015).

Protein profiles associated with prognosis have been studied in primary B-CLL samples and identified associations between increased expression of T-cell leukaemia/lymphoma protein 1A (TCL-1), thyroid hormone receptor-associated protein 3 (TR150), and S100A8 with high-risk CLL, and myosin-9 with low-risk disease (Alsagaby et al., 2014). A similar study identified 84 differentially abundant proteins, which have roles in cell proliferation, apoptosis, and DNA repair, between stable and progressive CLL (Huang et al., 2016). CLL proteomics is further reviewed in Chapter 6 of this thesis.

1.3.7. *IGH* Gene Usage and Mutation Status in F-CLL

Following T lymphocyte-dependent activation, B lymphocytes undergo a rapid proliferative phase in the germinal centre which is accompanied by immunoglobulin class switching, and affinity maturation is directed by the introduction of mutations into the immunoglobulin variable region genes. B lymphocytes are then selected according to the affinity of the encoded immunoglobulin for antigen, generating high-affinity memory B lymphocytes and plasma cells. CLL can be divided into a benign group, with a high load of mutations, and a progressive group, with a low load of mutations, by sequencing *IGH* genes and comparing with germline sequences. B-CLL cases with unmutated *IGH* genes (UM-CLL), or ≥ 98% sequence homology with germline, have a median survival of 8 years and those with mutated genes (M-CLL), or < 98% sequence homology with germline, have a median survival of 25 years (Damle et al., 1999, Hamblin et al., 1999).

In addition to providing prognostic information, distinguishing between these 2 groups is important to understanding the pathogenesis of CLL (Herve et al., 2005) (Klein and Dalla-Favera, 2010). Although it had been postulated that UM-CLL originates from a pre-germinal centre precursor with *IGH* genes lacking somatic mutations and M-CLL arises from a post-germinal centre B cell that expresses somatically hypermutated *IGH* (Stevenson and Caligaris-Cappio, 2004), UM-CLL and M-CLL both show similar gene expression profiles (Klein et al., 2001, Rosenwald et al., 2001). This similarity in gene expression profiles suggested a common cell of origin. Cloning and expressing *in vitro* recombinant antibodies from M-CLL and UM-CLL B lymphocytes and testing their specificity indicated that both M-CLLs and UM-CLL may originate from self-reactive B cell precursors (Herve et al., 2005). Somatic hypermutation is proposed to alter the original B cell receptor autoreactivity and disease progression (Herve et al., 2005).

Initial reports in small numbers of families had provided evidence of a restriction in *IGH* usage amongst CLL, compatible with selection by a common antigen (Shen et al., 1987). However, subsequent studies failed to provide confirmatory data (Sakai et al., 2000, Crowther-Swanepoel et al., 2008). The largest of these studies was of *IGH* usage in 327 F-CLL cases which were compared with 724 sporadic CLL cases (Crowther-Swanepoel et al., 2008). The frequency of M-CLL was higher in F-CLL and there was evidence of concordance in mutation status, however *IGH* usage was not different between F-CLL and S-CLL. Furthermore, *IGH* usage was not correlated between affected members within the same families (Crowther-Swanepoel et al., 2008). The failure of these studies to show a more restricted intra-familial phenotype with respect to *IGH* usage may be explained by an oncogenic event in the malignant cell of origin occurring before affinity maturation. The present study of mRNA and protein profiles in F-CLL patients, who share a similar genetic background, subclassified into M-CLL and UM-CLL may

provide further insight into the CLL cell of origin.

1.4. SUMMARY

In several reported families, CLL appears to be transmitted as an autosomal dominant disorder (Brown et al., 2012, Sellick et al., 2006a). However, genetic linkage studies have been unable to detect driver mutations due to the simultaneous presence of many low-penetrance associated genes. To address the rarity of large multigenerational B-CLL families, linkage studies have combined multiple families using data from affected subjects in each family (Goldin et al., 2003, Sellick et al., 2005b, Sellick et al., 2007, Ng et al., 2006, Ng et al., 2007b). However, since it is unlikely that each family will have identical susceptibility genes which map to the same disease loci, it is likely that linkage signals have conflicted. Consequently, most studies have been limited by weak evidence for linkage, and susceptibility loci have not shown concordance between studies (Sellick et al., 2005b).

To detect multiple low-risk loci, genome-wide association studies (GWAS) have used large numbers of cases and controls and dense-coverage SNP arrays (Di Bernardo et al., 2008, Speedy et al., 2014, Crowther-Swanepoel et al., 2010). To date, 30 risk variants have been identified, however these account for only ≈19% of CLL heritability (Speedy et al., 2014), suggesting that a large number of heritable factors remain undetectable by GWAS. A proportion of the remaining inherited CLL risk may be associated with non-DNA sequence-linked heritable information, including epigenetic modifications, which have been found to regulate oncogenes and tumour suppressor genes in a number of other hereditary cancers (Gazzoli et al., 2002, Esteller et al., 2001). Accordingly, identification of differentially expressed genes and proteins in CLL families may identify key targets of epigenetic modification, in addition to variations in transcription factor binding sites which affect gene regulation.

This present study will use high-resolution expression profiling microarrays and MS to identify differentially expressed genes and proteins in purified normal, premalignant and malignant B lymphocytes from a family with multiple cases of CLL and MBL (Nazarov et al., 2017, Meghann Palermo, 2014). For this family, my supervisor, A/Professor Fuller, has previously reported a genetic linkage scan which showed weak linkage to 14q24.1 and 14q31.2 (non-parametric linkage statistic = 2.24; p = 0.03) (Fuller et al., 2008).

AIMS

1. Determine if familial CLL/MBL (familial lymphoproliferative disease; F-LPD) B lymphocytes contain unique mRNA profiles compared to B lymphocytes from sporadic CLL (S-CLL) cases and control subjects.

2. Determine if F-LPD B lymphocytes contain unique protein profiles compared to B lymphocytes from S-CLL cases.

3. Use mRNA profiles to identify changes associated with progression of non-malignant normal B lymphocytes through premalignant MBL cells to malignant CLL cells.

4. Compare mRNA and protein profiles in *IGH* mutated and unmutated F-LPD and S-CLL.

An overview of the experimental design used in this thesis is shown in (Figure 1-2).

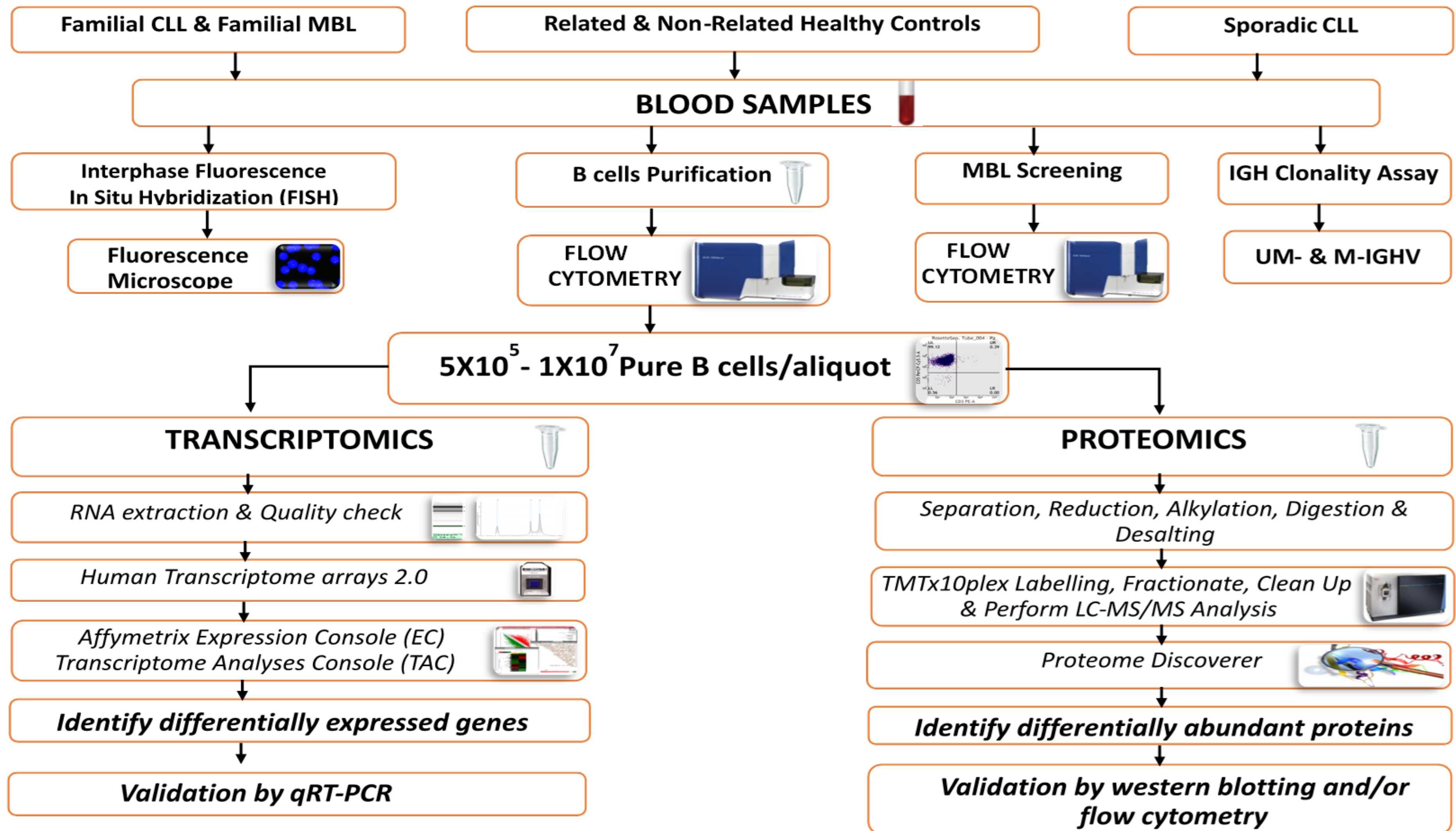

Figure 1-2. The overview of the experimental design used in this thesis

CHAPTER-2: MATERIALS AND METHODS

2.1. SAMPLE SETS

Subjects were recruited from the Nepean Hospital, Penrith, New South Wales, Australia after obtaining informed consent. The experimental protocol was approved by the Nepean and Blue Mountains Local Health District Human Research Ethics Committee (01/70) and the University of Sydney Human Research Ethics Committee.

2.1.1. Healthy Control Subjects

Peripheral blood samples were obtained from healthy volunteers after informed consent as approved by the Nepean and Blue Mountains Local Health District Human Research Ethics Committee (01/70) and the University of Sydney Human Ethics Committee.

2.1.2. Sporadic Chronic Lymphocytic Leukaemia (S-CLL) Subjects

Blood samples were obtained for analysis from >50 subjects with sporadic chronic lymphocytic leukaemia (S-CLL). The diagnosis of B cell CLL was made according to the current world health organisation (WHO) guidelines for the diagnosis and treatment of B-CLL (Hallek et al., 2008, Swerdlow et al., 2016). B-CLL subjects had a peripheral blood B lymphocyte count $\geq 5x10^9$/L for > 6 months with the presence of small to medium sized lymphocytes with condensed chromatin and smudge cells. The circulating B-CLL cells expressed immune-phenotypic markers: CD5 (T-cell antigen), CD10 (pre B-cell), CD19 (pan B-cell), CD20 (mature B-cell), CD22 (B-cell), CD23 (B-cell subset), CD79b (mature B-cell),CD38, FMC7, and surface immunoglobulin kappa or lambda light chains. Monoclonal B-cell lymphocytosis (MBL) was defined as the presence of monoclonal B-lymphocytes in the peripheral blood of $< 5x10^9$/L either with the phenotype of CLL, atypical CLL, or non-CLL (CD52) B cells (Swerdlow et al., 2016).

2.1.3.Familial Chronic Lymphocytic Leukaemia (F-CLL) subjects

Our group previously studied a large family from which 11 members over three generations had been diagnosed with B-CLL (Fuller et al., 2008). For the present study, samples were collected from 21 family members for analysis. Two subjects had been diagnosed with CLL (IV-02 & IV-05) and 4 with MBL (III-10, IV-13, IV-17 & IV-18) according to the criteria described in 2.1.2. The remaining 15 subjects were unaffected. The pedigree of the family members studied in this thesis is shown in (Figure 3-1).

2.2. CELL PURIFICATION

2.2.1.MATERIALS

Table 2-1 List of materials used for cell purification

	MATERIALS	COMPANY
1	Lithium heparin Vacutainer tubes	BD,Plymouth, UK Cat 367526
2	Ficoll-Paque PLUS	GE Healthcare Life sciences, Australia
3	15ml and 50ml conical Falcon centrifuge tubes	BD
4	0.2 µm syringe filter, PVDF	Merck Millipore Cat. SLGV033RS
5	Serological Pipettes	Greiner Bio One
6	Centrifuge	Heraeus Megafuge 1.0
7	Neubauer-improved bright line-counting chamber	Hirschmann, EM-techcolor
8	IX50 Inverted microscope	Olympus
9	RosetteSep™ Human B Cell Enrichment Cocktail	STEMCELL Technologies
10	Phosphate-Buffered Solution (PBS)	Amresco Inc.
11	Fetal calf serum (FCS)	
12	RNAlater®	Ambion®, USA, Cat. AM7020
13	Lysis buffer (RLY) from Isolate II RNA Mini Kit	Bioline, Taunton, MA, Cat. BIO-52072
14	2-Mercaptoethanol (β-ME)	Promega, Australia, Cat. Z523A-C
15	Urea	Sigma-Aldrich (St. Louis, MO, USA), Cat. No. U5378
16	Tris (hydroxymethyl) aminomethane	Amresco (Ohio, USA), Cat. No. 97061-794
17	Sodium Dodecyl Sulfate (SDS)	Amresco (Ohio, USA) Cat. No. 97064-470

2.2.2.METHOD

Peripheral blood samples from S-CLL patients and healthy volunteers were collected

in lithium heparin tubes at the Nepean Cancer Care Centre, Nepean Hospital, Penrith, New South Wales, Australia.

B lymphocytes were purified from peripheral blood using RosetteSep™ Human B Cell Enrichment Cocktail (StemCell Technologies, Australia). Briefly, 3 ml or 50 ml of a whole blood sample from CLL or healthy subjects, respectively, was incubated with RosetteSep antibody cocktail at a concentration of 70 µL cocktail/ml of blood in the dark at room temperature (RT) for 20 min. After incubation, the blood was diluted with one volume of 2% FCS in PBS. The diluted samples were then layered on top of one volume of Ficoll-Paque PLUS ™ density medium and centrifuged at 1200 g with brake off for 20 min. After centrifugation, three separating layers formed and the enriched B cells were harvested from the middle interface layer. The enriched cells were washed twice with 2% FCS in PBS and centrifuged for 10 min at 300 g. After washing, the enriched cells were counted using a Neubauer-improved bright line counting chamber and aliquoted for further analysis.

To check the efficiency of the separation and the purity of the enriched B cells by flow cytometry, a minimum concentration of 2 x 10^5 cells was set aside. (See sections 2.3.2.1.22.3.2.1.2 below).

For gene expression studies, a minimum of 1x10^6 of enriched B cells were pelleted and immediately lysed with 350 µL RLY + 3.5 µL 2-mercaptoethanol (β-ME) before being placed directly in a −80°C freezer until RNA isolation was performed (See section 2.4.2.2).

For protein analysis, a minimum of 1x10^6 enriched B cells were washed 4 times with 1ml PBS by centrifuging for 5 minutes at 2000 g, to remove FCS proteins. The washed pellet was homogenised by vigorous mixing in ≈200 µL of protein lysis buffer containing 8 M urea in 50 mM Tris-HCL with 0.1% (w/v) SDS, at pH 7.5, before storing at -80°C.

Unpurified PBMCs were isolated from whole blood at the same time by the same procedure described above except that the RossetteSep antibody cocktail was not used.

2.3. FLOW CYTOMETRY

2.3.1.MATERIALS

Table 2-2 List of materials used for Flow cytometry

	MATERIALS	COMPANY
1	BD FACS Round-Bottom Tubes	BD Biosciences, product. No. 352054
2	5% of fetal calf serum (FCS) in phosphate buffered saline (PBS)	
3	ethylene diamine tetra-acetic acid disodium salts dehydrate (EDTA)	Sigma-Aldrich (St. Louis, MO, USA), Product No: E5134-100G
4	5% FCS in PBS with 1mM EDTA) (FACS buffer)	
5	90-100% Paraformaldehyde	ProSciTech, Australia, Product No:C007
6	10X concentrate BD pharm lyse™ lysing buffer	BD Biosciences, Cat. No. 555899
7	The BD FACSuite™ CS&T research beads	BD Biosciences, Cat. No. 650621
ANTIBODIES FOR PURITY CHECK		
1	Anti-CD14-FITC	BD Biosciences, Cat. No. 347493
2	Anti-CD3-PE	BD Biosciences, Cat. No. 347347
3	Anti-CD5-PerCPCy5	BD Biosciences, Cat. No. 341089
4	Anti-CD20-APC	BD Biosciences, Cat. No. 340941
5	Anti-kappa-FITC & Anti-lambda-PE	BD Biosciences, Cat. No. 349516
ANTIBODIES FOR MBL SCREENING		
1	Anti-CD4-FITC	BD Pharmingen, Cat. No. 561005
2	Anti-CD19-FITC	BD Pharmingen, Cat. No. 560994
3	Anti-CD8-PE	BD Pharmingen, Cat. No. 561949
4	Anti-CD16-PE	BD Pharmingen, Cat. No. 560995
5	Anti-CD45-PerCPCy5.5	BD Pharmingen, Cat. No. 564106
6	Anti-CD3-APC	BD Pharmingen, Cat. No. 561810
7	Anti-IgM-FITC	BD Pharmingen, Cat. No. 562029
8	Anti-CD23-PE	BD Pharmingen, Cat. No. 561774
9	Anti-CD22-FITC	BD Pharmingen, Cat. No. 561771
10	Anti-CD79b-PE	BD Pharmingen, Cat. No. 561943
11	Anti-CD10-FITC	BD Pharmingen, Cat. No. 340925
12	Anti-CD38-PE	BD Pharmingen, Cat. No. 560981
13	Anti-CD20-APC	BD Pharmingen, Cat. No. 559776
14	Anti-CD5-PerCPCy5	BD Biosciences, Cat. No. 341089
15	Anti-kappa-FITC & Anti-lambda-PE	BD Biosciences, Cat. No. 349516
INSTRUMENTS		
1	BD FACSVerse flow cytometer	Becton Dickinson, Franklin Lakes, USA
2	BD FACSuite software	Becton Dickinson, Franklin Lakes, USA

2.3.2.METHODS FOR FLOW CYTOMETRIC ANALYSIS

2.3.2.1. Cell Staining

2.3.2.1.1. Cell Staining for Detecting MBL

Peripheral blood samples were collected from affected and unaffected family members into EDTA vacutainer tubes. Five antibodies cocktails were prepared as listed in (Table 2-3) by mixing equal volumes of each antibody.

Table 2-3 Antibody cocktails for MBL screening

	Cocktail 1	Cocktail 2	Cocktail 3	Cocktail 4	Cocktail 5	Unstained Tube
FITC	CD4/CD19	anti-Kappa	IgM	CD22	CD10	-
PE	CD8 /CD16	anti-Lambda	CD23	79b	CD38	-
PerCPCy5	CD45	CD5	CD5	CD5	CD5	-
APC	CD3	CD20	CD20	CD20	CD20	-

Antibody combinations (20 µL) were used to stain whole blood (100 µL) in 5 ml FACS tubes by incubation at 4°C for 30 minutes in the dark. Whole blood (100 µL) was used as an unstained control. After incubation, 2 mL of freshly prepared 1 X BD Pharm Lyse™ was added to each FACS tube, followed by gentle vortexing and further incubation for 15 min, at RT, in the dark. Cells were pelleted by centrifugation at 300 g for 5 min, at RT, and washed with 2 mL of FACS Buffer and centrifuged again at 300 g for 5 min. The washed cell pellet was re-suspended in 300 µL PFA before acquiring data by flow cytometry.

2.3.2.1.2. Cell Staining for Assessing B cell Enrichment

Fluorochrome-labelled antibodies were used to check the efficiency of the separation and purity of the enriched B cells isolated in section 2.2.2. Cells (2×10^5) were re-suspended

in 100 µl FACS buffer and antibodies were added according to (Table 2-4);

Table 2-4 Cell staining for assessing B cell enrichment

	Tube 1	Tube 2	Tube 3	Tube 4
Cell type	PBMCs	PBMCs	PBMCs	Enriched B cells
Total volume of cell suspension	100 µL	100 µL	100 µL	100 µL
Antibodies		- (10 µL) CD14-FITC - (7.5 µL) CD3-PE - (10 µL) CD5-PerCPCy5 - (5 µL) CD20-APC	- (10 µL) CD5-PerCPCy5 - (5 µL) CD20-APC - (5 µL) kappa-FITC & lambda-PE	- (10 µL) CD14-FITC - (7.5 µL) CD3-PE - (10 µL) CD5-PerCPCy5 - (5 µL) CD20-APC

All tubes were mixed with the appropriate amounts of monoclonal antibodies and incubated on ice for 30 min in the dark. After incubation, stained cells were washed twice with 2 mL PBS by centrifuging for 5 min at 300 g. The supernatant was carefully discarded and the pellet was resuspended in 300 µL of 4% paraformaldehyde (PFA) in PBS for fixation.

2.3.2.2. Flow Cytometry

2.3.2.2.1. Quality Control

Each day, the performance of the BD FACSVerse™ flow cytometer was checked using BD FACSuite™ CS&T research beads as described by the manufacturer.

2.3.2.2.2. Gating Strategy for Detecting MBL

Flow cytometry gating was conducted using a Becton Dickinson FACSVerse and for each test 50,000 cells were analysed using BD FACSuite software to identify B-CLL and MBL populations. A lymphocyte region for subset analysis was first established using forward light scatter (FSC) versus side scatter (SSC). Lymphocyte populations were first analysed using CD45 vs. SSC to distinguish lymphocytes from other cells. Lymphocyte were defined as CD45 bright with low SSC, and further gating on this population was performed to identify B- and T-

lymphocytes and NK cells. In the first cocktail, T-lymphocytes were evaluated by gating CD45 versus CD3 and further gating on CD3 populations to determine T cells subsets, T-helper (CD4) and T-cytotoxic (CD8), by analysing CD4 and CD8 against CD3. The inclusion of T cell markers in the first cocktail allowed evaluation and screening of lymphoid subsets. B-lymphocytes were identified in the 2nd cocktail tube by analysing CD19 against CD45, CD19 vs CD3 and CD19 vs CD16. NK cells were identified by CD16 vs CD45, CD16 vs CD19 and vs CD3.

B-lymphocyte populations were studied in other cocktail tubes by gating the lymphocyte region using forward light scatter FSC versus the side scatter SSC, then B-CLL and MBL populations were identified using CD20 vs.CD5 and in these populations the expression of CD10, CD22, CD23, CD38, CD79b, IgM and kappa and lambda light chains were evaluated. The use of kappa and lambda light chain antibodies is to determine monoclonality. The use of CD10, a marker of GC B-cells can be useful for identifying reactive B cell populations and differentiation from lymphoma. CD38, a marker for plasma cells, was included to assess B cell maturation and to identify abnormal populations within B cell populations. These combinations are required for diagnosis of monoclonal B-cell lymphocytosis which expressed CD19, CD5 and CD23 with weak or no expression of CD20, CD79b, CD22 and surface IgM.

2.3.2.2.3. Gating Strategy for Assessing B cell Enrichment

Each tube was analysed using a BD FACSVerse™ cytometer. Data from 10,000 events were collected using BD FACSuite™ software. A gate around all leukocytes was first established using FSC versus SSC. To determine the percentage of pure B-CLL cells, purified B-lymphocytes were analysed using CD5, CD3, CD14 and CD20. A high percentage of CD20 and CD5 with negative expression of CD3 and CD14 indicated high purity B cells. Un-purified PBMCs with the same panel were also included in a separate tube as a control sample.

2.4. RNA EXTRACTION

High quality RNA was required for gene expression studies. Many different kits are used to extract RNA from blood cells, but an effective and efficient one must be selected to prevent failure of downstream analysis, avoid contamination and reduce reagents costs. Since limited numbers of cells were available from some samples it was necessary to establish an optimal RNA extraction procedure. Two extraction methods; TRI reagent (Sigma-Aldrich, Australia, Cat No.T9424) and Isolate II RNA mini kit (Bioline, Taunton, MA, Cat No.BIO-52072) were compared in order to determine the best method to obtain high yield and quality RNA.

2.4.1.MATERIALS

Table 2-5 List of materials used for RNA Extraction

MATERIALS		COMPANY
1	1-Bromo-3-chloropropane	Sigma-Aldrich, Cat. No. B9673
2	2-Propanol	Sigma-Aldrich, Cat. No. I-9516
3	Ethyl alcohol, Pure	Sigma-Aldrich, Cat. No. E7023
4	Water, molecular biology grade	Sigma-Aldrich, Cat. No. W4502
5	2-Mercaptoethanol β-ME	Promega, Product code Z523A-C
KITS		
1	TRI reagent	Sigma-Aldrich, Cat No.T9424
2	Isolate II RNA mini kit	Bioline, Taunton, MA, Cat No.BIO-52072
INSTRUMENTS		
1	Varian Cary 1 Bio UV spectrophotometer	
2	Sigma Microcentrifuge	

2.4.2.METHODS

2.4.2.1. TRI Reagent

RNA was extracted from 5 x 10^6 PBMCs cells preserved in *RNAlater*®, as described in section 2.2.2, with TRI Reagent (Sigma-Aldrich, USA, Catalog No.T9424) according to the manufacturer's instructions. Briefly, 0.5 mL of TRI reagent was added to cells and mixed gently by repeated pipetting to form a homogenous lysate. Samples were left for 5 min at room

temperature and then 1-Bromo-3-chloropropane (0.05 mL) was added and shaken vigorously for 15 s. Samples were incubated for 10 min at room temperature and the resulting mixture was centrifuged at 12,000 g for 15 min at 4°C, separating the mixture into three phases. The colourless, upper-aqueous phase containing RNA was transferred into a new 1.5 ml micro-tube and 2-propanol (250 µL) was added, mixed and allowed to stand for 5–10 minutes at RT. After incubation, samples were centrifuged at 12,000 g for 10 min at 4°C so that the RNA precipitate formed a pellet on the side and bottom of the micro-tube. The supernatant was carefully discarded and the RNA pellet was washed by adding a minimum of 1 ml of 75% ethanol. This was followed by centrifuging at 10,000 g for 5 min at 4°C, the supernatant was carefully decanted and the RNA pellet was dried for 5-10 min using air-drying. RNA was dissolved in water (50 µL) and the quality and quantity of purified RNA was assessed by spectrophotometry at A260 and A280 nm.

2.4.2.2. Isolate II RNA Mini Kit (Bioline)

RNA was extracted from 5 x 10^6 PBMCs cells preserved in *RNAlater*®, as previously referred to in section 2.2.2, using the Isolate II mini RNA kit according to the manufacturer's protocol. Lysis buffer (RLY) (350 µL) and β-ME (3.5 µL) were added to the cell pellet and mixed several times vigorously by pipetting to form a homogenous lysate. This was passed through a filter to reduce viscosity and clear the lysate by centrifugation at 11,000 g for 1 min at RT. Ethanol (70%, 350 µL) was added to the filter and mixed by pipetting up and down 5 times and loaded onto the RNA isolation column, followed by centrifuging at 11,000 g for 30 s, trapping the RNA on the column. The column was then washed with 350 µL membrane desalting buffer and centrifuged at 11,000 g for 1 min. The column was then treated with DNase I to eliminate any genomic DNA contamination as described by the manufacturer. After further washing, RNA was eluted from the column with RNase-free water (60 µL) supplied with the kit,

centrifuged at 11,000 g for 2 min, and the quality and quantity of purified RNA was assessed by spectrophotometry at A260 and A280nm.

2.5. INTERPHASE FLUORESCENCE IN SITU HYBRIDIZATION (FISH)

Interphase Fluorescence in Situ Hybridization (FISH) was adapted from the Cytogenetic Laboratory, the Children's Hospital at Westmead, Sydney, NSW, Australia.

2.5.1.MATERIALS AND SOLUTIONS

Table 2-6 List of materials used for FISH Analysis

MATERIALS		COMPANY
1	Centrifuge	Heraeus Megafuge 1.0
2	Diamond-tipped pencil	
3	Water-bath	Thermoline Scientific
4	Dry block heater	Thermoline Scientific
5	Inverted microscope	Olympus IX50 Inverted Microscope
6	Microscope Slides	
7	15mm round coverglass	ProSciTech Pty Ltd
8	24x60mm coverglass	ProSciTech Pty Ltd
9	Coplin jars	
10	Potassium chloride (KCl)	Sigma-Aldrich, Product No. P 5405
11	Methanol	Sigma-Aldrich, Product No. P 34860
12	Acetic acid (glacial) 100%	Merck Millipore, Cat. No. 1000632500
13	Sodium chloride (NaCl)	Sigma-Aldrich, Cat No. 32038
14	Trisodium citrate dehydrate	Sigma-Aldrich, Cat No. S1804
15	Nonidet P-40 (NP40)	Roche, Cat No. 11754599001
16	4,6-Diamidino-2-phenylindole (DAPI)	Sigma-Aldrich, Cat No. D9542
17	Tween 20	Sigma-Aldrich, Cat No. P1379
18	XL CLL probe kit	MetaSystems, (D-5044-100-TC)
19	ECLIPSE 80i epi-fluorescence microscope	Nikon, Tokyo, Japan
20	Bottle Top Vacuum Filters, 0.22µm Pore 33.2cm² PES Membrane	500mL, Corning, Cat. No. 431118
21	1.15g DABCO (1,4-diazabicyclo-[2.2.2]octane (Triethylenediamine))	Sigma-Aldrich, Australia, Cat No. D27802
22	Glycerol	
23	Art cement	

Table 2-7 List of solutions used for FISH Analysis

SOLUTIONS		
1	0.075M KCl	5.592g KCL dissolved in 1L MilliQ H2O
2	Carnoy's fixative	3 parts of Methanol to 1 part of acetic acid
3	The XL CLL Probes	MetaSystems, D-5044-100-TC
4	20X saline sodium citrate (SSC)	87.66g NaCl was dissolved in 450mL Milli-Q water then 44.115g Trisodium citrate dehydrate was added to the NaCl solution. The volume was adjusted to 500mL and sterilised by filtration and stored protected from light.
5	Wash buffer 1: 0.4xSSC / 0.3% NP40	20mL of 20XSSC was transferred into a new sterile 500mL bottle and then up to 500mL Milli-Q water was added, mixed before adding 1.5mL NP40. Stored at RT
6	Wash buffer 2: 2xSSC / 0.1% NP40.	100mL of 20XSSC, 400mL Milli-Q water and 0.5mL NP-40 in 500mL sterile bottle. Stored at RT
7	DAPI Stock in 5mL volume	To 1mg DAPI, 0.5mL methanol was added and swirled to mix and then 4.5mL distilled water was added and mixed by inversion. Stored at -20°C.
8	4xSSC / 0.2% Tween 20	100mL of 20XSSC, 400mL Milli-Q water and 1mL Tween 20 in 500mL sterile bottle. Stored at RT.
9	Antifade Solution	1.15g DABCO was first dissolved in 1mL 1M TRIS-HCL pH 9.0, and 4mL Milli-Q water and then made up to 50mL with Glycerol. Stored at 4°C.

2.5.2.METHOD

2.5.2.1. Slide Preparation

Briefly, peripheral blood (6 mL) was collected in lithium heparin tubes from each CLL patient and 0.5 mL patient blood was diluted with 10 mL hypotonic solution (0.075 M KCL, 37°C) incubated in a 37°C water-bath for 10 min, followed by centrifugation at 300 g for 6 min, at RT. The supernatant was discarded and re-suspended, cells were washed with 10 mL of freshly prepared 5% acetic acid and centrifuged as before. The supernatant was removed and pellets were fixed by washing 3 times with freshly prepared Carnoy's fixative (10 mL), and centrifuged as before, to obtain a clear white pellet. After the last wash, 3 drops of Carnoy's fixative were added to re-suspend the pellet. For each probe, one slide was prepared, cleaned with fixative and labelled with patient details and probe name. One or two drops (≈10 µL) of cell suspension were dropped from a minimum height of 20 cm onto the centre of the slide

and the position of drops were marked by circling the underside of slide with a diamond pencil. A test slide was dried on the bench and the density of cellular dispersion was checked under phase contrast on a IX50 inverted microscope (Olympus). The quality assessment of the cell spreading on the slide is an essential step to avoid cells being too dense which could introduce background signals and lead to incorrect FISH results. A satisfactory preparation should have ≈20 interphase nuclei/field of view using a low power lens. The prepared slides were then used immediately or stored at -20°C. Unused cell suspension in Carnoy's fixative was stored at 4°C for future analyses.

2.5.2.2. Probe Preparation and Denaturation

The XL CLL probe kit (MetaSystems, D-5044-100-TC) was used to determine the most frequent recurrent genomic copy number aberrations in CLL. According to the manufacturer's instructions, this kit contains 2 vials: 1. vial contains the XL DLEU/*LAMP*/12cen that is a 3-colour probe-mix containing the DLEU gene region at 13q14 (including the D13S319 marker) in red, the LAMP gene region at 13q34, and a chromosome 12 centromeric probe labelled in green and; 2. vial contains the XL *ATM*/*TP53* locus-specific probe that detects a deletion in the long arm of chromosome 11 and the short arm of chromosome 17. The green labelled probe hybridises to a specific region at 11q22 covering *ATM* and the red labelled probe hybridises specifically to *TP53* and flanking regions at 17p13. The vials were thawed at room temperature for 15 min and gently vortexed before applying them on the prepared slides. From the appropriate vials, 5 µL were dispensed onto the corresponding slides as guided by the diamond pencil mark and covered with 15 mm round coverslips. Air bubbles were removed with forceps before sealing the coverslips with art cement. The slides were denatured by placing on a hotplate at 75°C for 4 min. The hotplate was turned off and the slides were left on the hotplate until the temperature had returned to 37°C.

2.5.2.3. Hybridisation

For hybridization, the slides were transferred into the humidified chamber at 37°C for overnight incubation. The art cement and coverslips were then carefully removed before applying a series of stringency washes.

2.5.2.4. Post-Hybridisation Washes

After overnight hybridization, slides were placed into a Coplin jar containing wash buffer 1 at 75°C for 2 min with gentle agitation every 10 s. Slides were transferred into another Coplin jar containing wash 2 buffer at room temperature for a minimum of 1 minute. The buffer was discarded and the Coplin jar containing the slides was left in a dark place to dry before applying DAPI counterstain solution.

2.5.2.5. Counter Staining

A freshly prepared working stock of DAPI counterstain solution was made by diluting DAPI Stock 1/1500 in 4XSSC / 0.2% Tween 20 (30 µL in 45 mL) and stored in the cupboard to protect from light until required. For the counterstaining process, DAPI working stock solution was decanted into the Coplin jar containing the slides and stained for 5 min in the dark place to protect from light. After staining, DAPI solution was discarded and the slides rinsed four times in Milli-Q water. The slides were then removed and left to air dry in the dark before mounting as described below.

2.5.2.6. Mounting

To prepare slides for mounting, three small drops of antifade were evenly spaced onto 24 x 60 mm coverslips. The slides were placed on top of the coverslips and left to spread. The excess antifade solution and bubbles were removed with tissues and forceps. The slides were

stored at 4°C in a cardboard rack labelled with the date and covered to protect from light for analysis.

2.5.2.7. Visualisation

Slides were analysed in a darkened room using an ECLIPSE 80i epi-fluorescence microscope equipped with a charge-coupled device camera and appropriate filters. A total of two-hundred images of interphase nuclei were captured for every probe set according to the manufacturer's instructions.

For the XL DLEU/LAMP/12cen probe set, an appropriate filter to detect *LAMP* was not available therefore a normal signal pattern was indicated by 2 Red/2 Green dots. A signal pattern of 2 Red/3 Green dots indicated trisomy 12 while a signal pattern of 1 Red/2 Green dots indicated deletion of DLEU (13q14).

The normal signal patterns for the XL *ATM*/*TP53* probe set is 2 x Green/2 x Red dots. A signal pattern of 1 x Green/2 x Red indicated *ATM* deletion (11q22) and a signal pattern of 2 x Green/1 x Red indicated a deletion of the *TP53* locus (17p13).

Results were considered to be abnormal when the percentage of cells with any given abnormality was > 5% in 200 interphase nuclei for Trisomy 12 and > 8% for deletions of 13q, 11q and 17p, as noted in Wawrzyniak *et al.* (Wawrzyniak et al., 2014) study. A case was considered to have a poor prognosis if the nuclei carried an 11q or 17p aberration, an intermediate prognosis if trisomy 12 was detected and a good prognosis if a 13q deletion was detected (Wawrzyniak et al., 2014) (Wolff et al., 2007).

2.6. *IGH* GENE CLONALITY ASSAY

This assay was carried out in collaboration with Maryam Hassanvand, an MD student

at the University of Sydney Nepean Clinical School.

2.6.1.MATERIALS

Table 2-8 List of materials used for IGH gene clonality assay

	MATERIALS	COMPANY
1	Wizard Genomic DNA purification kit	Promega, Cat No. A1120
2	Isopropanol	Sigma-Aldrich, Cat. No. I-9516
3	Pure, ethyl alcohol	Sigma-Aldrich, Cat. No. E7023
4	Water, molecular biology	Sigma-Aldrich, Cat. No. W4502
5	Thermo Scientific Nanodrop 2000 Spectrophotometer	Thermo Fisher Scientific, Wilmington, U.S.A.
6	*IGH* Gene Clonality Assay	Invivoscribe Technologies Inc., San Diego, USA
7	Thermo-cycler	PTC-225 Tetrad DNA Engine PCR System, MJ Research
8	The Mini-Sub cell GT agarose gel electrophoresis system	Bio-Rad Laboratory Pty., Ltd, NSW, Australia
9	Ultra-pure grade 25X Tris-acetate-EDTA (TAE) concentrate ready pack	Amresco, Cat No 0912-2PK, Australia
10	Grade Agarose 1	Amresco, Cat No 0710-500G, Australia
11	Safe DNA gel stain in 1 X TAE	Invitrogen, Cat No S33111
12	EZ Load 100 bp molecular ruler	Bio-Rad, Cat No170-8352
13	Safe imager blue light trans-illuminator	Invitrogen, S37102
14	Wizard SV Gel and PCR Clean-Up System	Promega, Cat No. A9282
15	25X Tris-acetate-EDTA (TAE) concentrate	One foil pouch containing 25X TAE powder was dissolved in 1L Milli-Q water.

2.6.2.METHOD

2.6.2.1. Isolation of DNA from peripheral blood

DNA was isolated from peripheral blood using the Wizard Genomic DNA purification kit (Promega), according to the manufacturer's protocol. Briefly, blood samples were collected in lithium heparin vacutainer tubes. Whole blood (300 µL) was added to a sterile 1.5 mL micro-centrifuge tube containing Cell Lysis Solution (900 µL) and mixed by inverting the tube 5-6 times. The tube was left at room temperature for 10 min to lyse the red blood cells. The sample was centrifuged at 13,000 g for 30 s at room temperature and the supernatant was carefully removed. The white cell pellet was re-suspended and Nuclei Lysis Solution (300 µL) was added to the tube and pipetted up and down gently 5 times to lyse the white blood cells

until the solution became very viscous. Protein Precipitation Solution (100 µL) was added to the nuclear lysate and vortexed vigorously for 20 s. The tube was centrifuged at 13,000g for 3 min at RT, and the supernatant containing the DNA was carefully transferred to a new sterile 1.5 mL micro-centrifuge tube containing isopropanol (300 µL). The solution was inverted gently until the white thread-like strands of DNA formed a visible mass, and then centrifuged at 13,000 g for 3 min at RT. After centrifugation, the DNA formed a small white pellet on the side and bottom of the microcentrifuge tube. The supernatant was carefully aspirated and then 70% ethanol (300 µL) at RT was added to the DNA pellet. The tube was inverted several times gently to wash the DNA pellet and the sides of the tube before centrifuging at 13,000 g for 3 min at RT. The supernatant was carefully aspirated leaving a small white pellet of DNA and the tube was inverted onto a clean absorbent paper to air-dry for 60 min. Afterward, DNA Rehydration Solution (100 µL) was added to the DNA pellet and incubated at 65°C for 1 hour or overnight at 4°C. Finally, the rehydrated DNA was quantified using a Thermo Scientific Nanodrop 2000 Spectrophotometer and its software (Thermo Fisher Scientific, Wilmington, U.S.A.).

2.6.2.2. IGH Gene Clonality Assay

Clonal immunoglobulin heavy chain gene (*IGH*) rearrangements were detected using the *IGH* Gene Clonality Assay (Invivoscribe Technologies Inc., San Diego, USA). The kit contains six primer master mixes; IGH tube A catalogue no. 2-101-0011, IGH tube B catalogue no. 2-101-0101, and IGH tube C catalogue no. 2-101-0031, for targeting the framework regions within the variable and the joining regions of *IGH*, while IGH tube D catalogue no. 2-101-0041 and IGH tube E catalogue no. 2-101-0051, targeted the diversity and joining regions. The last master mix tube is the specimen control size ladder (Catalogue no. 2-096-0021), which targeted multiple genes and generated a serious of amplicons to ensure that the quality and

quantity of input DNA was sufficient to yield valid results. Positive clonal controls DNA (IVS-0030, IVS-0019, IVS-0024, and IVS-0008) and negative polyclonal controls DNA (IVS-0000) were also included. Reactions were set up according to the manufacturer's instructions using AmpliTaq Gold DNA polymerase and amplification was performed on a thermo-cycler using the following PCR program,

1 x cycle	- Initial incubation at 95°C for 7 min.
35 x cycle	- Denaturation at 95°C for 45 s.
	- Annealing at 60°C for 45 s.
	- Extension at 72°C for 90 s.
1 x cycle	- Extension at 72°C for 10 min.

The PCR products were stored at 4°C before analysing them on a 2.5% agarose gel.

2.6.2.3. Agarose Gel Electrophoresis

Agarose gel electrophoresis was performed using the Mini-Sub cell GT agarose gel electrophoresis system, to separate and visualise DNA of various size. For a final concentration of 2.5% agarose gel, agarose (2.5 g) was dissolved in 1X TAE (100 mL) by heating in a microwave oven until the agarose was completely dissolved. SYBR safe DNA gel stain (5 µL) was added to the agarose solution and gently swirled before pouring the gel into a 7 x 10 cm Bio-Rad Mini-Sub cell gel-casting tray. An eight-tooth well-forming comb was inserted, and the gel was allowed to solidify for at least 45 min at RT. The casting tray containing the gel on it was inserted into the electrophoresis box with the sample wells near to the negative black cathode. The box was gradually filled with 1X TAE running buffer until the gel was submerged and the comb was then carefully removed from the solidified gel. EZ Load 100 bp molecular

ruler (5 µL) was loaded in the first well as a DNA size standard. 6X nucleic acid sample loading buffer (5 µL) was added to 25 µL of amplified DNA sample, and then 20 µL of sample mixture was carefully loaded into agarose well. Electrophoresis was carried out at room temperature at a constant voltage of 110V for 60 min. After electrophoresis, the gel was placed on the Safe Imager blue light trans-illuminator for nucleic acid visualisation and analysis. An image of the gel was collected, and stained agarose gel bands within the valid size range were excised using sterile surgical scalpels and transferred directly into a new sterile 1.5 micro-centrifuge tube for DNA purification.

2.6.2.4. Gel Slice DNA Purification

DNA was purified from gel slices using the Wizard SV Gel and PCR Clean-Up System, according to manufacturer's instructions. Briefly, membrane binding solution (500 µL) was added to a 1.5 mL micro-centrifuge tube containing the DNA gel slice, vortexed vigorously and incubated at 60°C for 5-10 min until the gel was completely dissolved. The dissolved DNA gel mixture was transferred to a SV mini-column in a collection tube and incubated for 1 min at room temperature. After that, the SV mini-column assembly was centrifuged at high speed (16,000 g) for 1 min, then the flow was discarded and the SV mini-column was washed with membrane wash solution (700 µL) and centrifuged again for 1 min at 16,000 g. Another washing step was performed by adding membrane wash solution (500 µL) followed by centrifugation for 5 min at 16,000 g. After centrifugation, the flow-through was carefully removed, and the SV mini-column in the collection tube was then re-centrifuged for 1 min with the micro-centrifuge lid open to completely evaporate any residual ethanol. The SV mini-column was transferred into a nuclease-free 1.5 mL micro-centrifuge tube and nuclease-free water (30 µL) supplied with the kit was added directly to the centre of the SV mini-column silica membrane without touching the membrane with the pipette tip. The tube was incubated

at RT for 1 min, followed by centrifuging at 16,000 g for 1 min. The quantity of purified DNA was measured using a Thermo Scientific Nanodrop 2000 Spectrophotometer and its software (Thermo Fisher Scientific, Wilmington, U.S.A.).

2.6.2.5. Sequencing Reaction

Sequencing of the purified fragments was performed at the Australian Genome Research Facility (AGRF) (Westmead Millennium Institute, Westmead, Australia) by high throughput Sanger sequencing using applied bio-systems 3730 and 3730xl capillary sequencer with big dye terminator (BDT) chemistry version 3.1 and standard sequencing protocols (https://www.agrf.org.au/docs/sanger-sequencing-sample-preparation-guide.pdf).

Nucleotide sequences were compared with the international ImMunoGeneTics information system (IMGT) (http://www.imgt.org/) and IgBLAST (https://www.ncbi.nlm.nih.gov/igblast/) databases and gene-usage and mutational status determined based on the recommended threshold of 2% to distinguish mutated from non-mutated B-CLL cases (Tobin et al., 2005) (Ghia et al., 2007).

2.7. Quantitative reverse-transcriptase PCR (qRT-PCR)

2.7.1.MATERIALS

Table 2-9 List of primers used for qRT-PCR validation of gene expression

	GENE	PRIMERS
1	GRASP[1]	F: GCTCAGGATTCCGCTGGAAGAA R: AGGTCACCATTTCCACACGCTG
2	TBC1D10C[1]	F: GCTCAGGATTCCGCTGGAAGAA R: AGGTCACCATTTCCACACGCTG
3	RASGEF1B[2]	F: CTGGATCCCTGGAAGCACTC R: TTGGCCATTAGCTCATACGGA
4	DFNB31[2]	F: TCTACGCTCTCCCAGCTCTC R: CCGCACTTTTCTTCACACGG
5	INPP5F[1]	F: TGGCACATCCTTAGAGGCTCTG

		R: CACTGGCACCTTCATCCAAAGG
6	SMAD3[1]	F: TGAGGCTGTCTACCAGTTGACC R: GTGAGGACCTTGTCAAGCCACT
7	GADPH[2]	F: CGAGATCCCTCCAAAATCAA R: TTCACACCCATGACGAACAT

[1] Sequences sourced from OriGene (https://www.origene.com)

[2]Sequences determined using Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast)

2.7.2.METHOD

To validate primers for use in qRT-PCR, primer pairs were checked for specificity, appropriate amplicon size, and for the absence of SNPs, using Primer-blast (www.ncbi.nlm.nih.gov/tools/primer-blast) and SNPCheck3 (https://secure.ngrl.org.uk/SNPCheck), respectively. Total RNA was isolated from purified B cells using Isolate II RNA mini kit (Bioline, Taunton, MA) as mentioned earlier. cDNA synthesis of up to 1 µg RNA was performed using a Tetro cDNA synthesis kit (Bioline, Taunton, MA.) according to the manufacturers' protocol using a 1:1 ratio of random hexamers and oligo dT for priming. qRT-PCR was performed using a Rotor-Gene 2000 (Corbett Research, Sydney, Australia) as described previously (Skarratt and Fuller, 2014) except that the annealing conditions were optimized to 58°C for 15 s , and the extension conditions to 72°C for 10 s over 35 cycles. Amplified products were verified by sequencing of gel purified amplicons as described in sections 2.6.2.4 2.6.2.5.

CHAPTER-3: A COMPARISON OF MESSENGER RNA PROFILES IN FAMILIAL MBL, FAMILIAL B-CLL AND SPORADIC B-CLL

3.1. INTRODUCTION

CLL is the commonest subtype of leukaemia in western countries, accounting for ≈25% of all cases (Howlader N, 2017). However, the incidence is very low in eastern countries and in immigrants from eastern to western countries, suggesting the presence of a genetic predisposition. Furthermore, there is no clear relationship between the development of CLL and environmental factors, including exposure to radiation (Hatch and Cardis, 2017).

Using GWAS data, the heritability of CLL has been estimated to be 46–59% (Di Bernardo et al., 2013, Berndt et al., 2013). Di Bernardo *et al*., used SNP genotyping data from a GWAS of 517 CLL cases and 2,930 CLL controls (Di Bernardo et al., 2013) and estimated heritability using the methods of Yang *et al*. and Lee *et al*. (Yang et al., 2010, Lee et al., 2011). In case-control studies the proportion of cases is usually larger than the prevalence in the population, therefore the data were transformed to account for disease prevalence, incomplete linkage-disequilibrium (LD) and ascertainment (Lee et al., 2011), and the heritability of CLL was estimated to be 0.59, where heritability ranges from 0 (no genetic contribution) to 1 (all trait differences reflect genetic variation) (Di Bernardo et al., 2013). Berndt *et al.* used genotyping data from 4 GWAS and 1 case control study (Crowther-Swanepoel et al., 2010b, Crowther-Swanepoel et al., 2011, Di Bernardo et al., 2008, Slager et al., 2011) to perform a meta-analysis of 3100 CLL cases and 7667 controls (Berndt et al., 2013). In this study, the contribution of all common variants to the genetic heritability of CLL was investigated using the same methods as Di Bernardo *et al*. (Yang et al., 2010, Lee et al., 2011, Di Bernardo et al., 2013), and common SNPs were estimated to explain up to ≈46% of the familial risk of CLL (Berndt et al., 2013).

The heritability of CLL is amongst the highest of any cancer (Sellick et al., 2006a), and

heritability studies of GWAS data are compatible with susceptibility to CLL mediated through many common SNPs. A family history of CLL or a related B-cell lymphoproliferative disorder is present in ≈ 10% of cases, and an inherited risk for developing CLL has been supported by twin studies, case control studies, and cohort studies (Bevan et al., 2000, Sellick et al., 2006a, Di Bernardo et al., 2013). Registry-based studies of related subjects have consistently demonstrated familial aggregation of CLL, with the two largest studies, based on the Swedish Family-Cancer Database and the Utah Population and Cancer Registry Database, showing an 8.5 and 6.1 fold first-degree familial relative risk of CLL, respectively (Czene et al., 2002, Goldgar et al., 1994, Kerber and O'Brien, 2005). Furthermore, identification of a higher frequency of MBL in families with multiple affected cases provides further evidence of inheritance in a subset of CLL patients (Rawstron et al., 2004, Rawstron, 2002).

To date, GWAS have identified 30 risk variants, however these account for only ≈19% of the familial risk of CLL and no definite predisposition genes have been identified (Speedy et al., 2014). It is likely that further advances will be made using GWAS- and WGS-based association studies, however genetic studies of large multigenerational families continue to provide opportunities to identify the remaining ≈80% of susceptibility genes.

Several families have been reported in which CLL appears to be transmitted as an autosomal dominant disorder (Sellick et al., 2006a). However, genetic linkage studies have been unable to detect any driver mutations, and the presence of numerous collaborating, genes has likely limited the usefulness of this approach. The largest family studied to date comprised 11 x F-CLL cases and 4 x F-MBL over 5 generations (Fuller et al., 2008). A genetic linkage scan of this family using a Genechip® Mapping 10K 2.0 Xba Array containing ≈10 200 SNPs in F-CLL cases and controls showed a maximal logarithm of odds (LOD) score of 1.01 ($p=0.06$) at 2q37.2, which contains a number of potential CLL candidate genes that could be

further studied for the presence of variants including *INPP5D*, which interacts with KRAS and NRAS in the B-cell receptor signalling pathway (Isnardi et al., 2006), and *SP140*, the lymphoid-restricted homologue of *SP100* which has a role in EBV-mediated B cell immortalization (Bloch et al., 1996), and was identified in the Di Bernardo *et al.* GWAS (Di Bernardo et al., 2008). The maximal NPL statistic of 2.24 (P=0.03) in the family was obtained between 14q24.1 and 14q31.2 (Fuller et al., 2008), containing *ZFP36L1* which has been associated with the apoptotic response to rituximab (Jackson et al., 2006, Baou et al., 2009). However, germline mutations that segregated with affected family members were not found in *ZFP36L1*.

It is likely that a proportion of genetic susceptibility to CLL results from mutations that affect gene regulation, including transcription factor binding sites (Law et al., 2017), and epigenetic modifications, rather than changes in DNA sequences that affect protein function. Mutations or epimutations, pathogenic alterations in DNA methylation or chromatin structure that do not alter DNA sequence, affect the function of the expressed imprinted gene allele and can result in imprinting disorders. Mechanisms that give rise to an imprinting disorder include uniparental disomy, intragenic mutations or copy number alterations that alter the function of a gene, and mutations or epimutations in imprinting control centres that alter imprinting or expression of genes (Abu-Amero et al., 2008, Bullman et al., 2008, Scott et al., 2008a, Scott et al., 2008b). A number of human imprinting syndromes are associated with cancer risk, including the Beckwith–Wiedemann syndrome imprinted domain at 11p15.5 which is associated with Wilms tumour, hepatoblastoma, and rhabdomyosarcoma (Kamikihara et al., 2005, Kaneda and Feinberg, 2005, Yuan et al., 2005); and the Prader–Willi syndrome imprinted domain at 15q11-q13 which is associated with myeloid leukemias (Davies et al., 2003).

HYPOTHESIS

It is proposed that inherited mutations or epimutations affecting the expression of imprinted genes can be inferred by differences in mRNA levels in controls, F-CLL cases, F-MBL, and S-CLL.

AIMS

1. Determine if family CLL/MBL (combined as familial lymphoproliferative disease; F-LPD) B lymphocytes contain unique mRNA profiles compared to B lymphocytes from unaffected family subjects and sporadic CLL (S-CLL) cases. Differences in gene expression between F-MBL, F-CLL, S-CLL and controls were assessed using analysis of variance (ANOVA), and false discovery rate (FDR) p-values for unsupervised hierarchical clustering (Pounds, 2006, Reiner et al., 2003).

2. Compare mRNA profiles between F-LPD and S-CLL groups, after removing genes that were differentially expressed as a result of genetic relatedness within the F-LPD group. To remove these genes, the F-LPD group was compared to the S-CLL group, and differentially abundant mRNAs were then compared to related family controls. Only mRNAs differentially abundant between F-LPD and S-CLL, and F-LPD and family controls, were used for semi-supervised hierarchical clustering (Bair, 2013, Bair and Tibshirani, 2004).

3.2. MATERIALS

Table 3-1 List of materials used for Affymetrix GeneChip Human Transcriptome Array 2.0

	MATERIALS	COMPANY
1	Affymetrix GeneChip Human Transcriptome Array 2.0	Affymetrix Inc, Santa Clara, California, USA
2	RNA Pico 6000 chip	Agilent Technologies
3	Agilent Bioanalyser	Agilent Technologies
4	GeneChip® WT Pico Kit	Affymetrix Inc, Santa Clara, California, USA, P/N 703262 Rev.5
5	GeneChip® Whole Transcript (WT) Expression Arrays	Affymetrix Inc, Santa Clara, California, USA
6	GeneChip® Expression wash, stain and scan for Cartridge arrays kit	Affymetrix Inc, Santa Clara, California, USA
7	Gene chip scanner 3000 7G	Affymetrix Inc, Santa Clara, California, USA
8	Affymetrix Expression Console software, version 1.4	Affymetrix Inc, Santa Clara, California, USA
9	Transcriptome analysis console (TAC) 3.0 software	Affymetrix Inc, Santa Clara, California, USA

3.3. METHODS

3.3.1.Patients and Samples

The experimental protocol was approved by the Nepean and Blue Mountains Local Health District Human Research Ethics Committee (01/70). Subjects were recruited from the Nepean Hospital, (Penrith, New South Wales, Australia) after obtaining informed consent. Peripheral blood samples were collected from 6 surviving patients (two with F-CLL, and 4 with F-MBL), and 3 controls from a family with multiple cases of F-LPD, (Figure 3-1) (Fuller et al., 2008). Blood samples were collected from a further 3 unrelated controls, and 6 S-CLL cases. All B-CLL subjects were treatment naïve.

The diagnosis of B-CLL was based on the presence of a monoclonal B lymphocyte count ≥5 X 10^9/L for ≥ 3 months, expression of CD19, CD5 and CD23, and weak or no expression of

CD20, CD79b, CD22 and surface IgM. The diagnosis of MBL was based on the same immunophenotypic profile, however clonal B cells were <5 × 10^9/L (Hallek et al., 2008).

B lymphocytes were enriched using a RosetteSep™ B-Cell isolation cocktail (StemCell Technologies Inc., Vancouver, BC, Canada) to provide >95% B lymphocyte purity, confirmed using flow cytometry (Essakali et al., 2008). PCR was used to confirm a clonal B-lymphocyte population by the presence of identical rearrangements of immunoglobulin heavy chain variable genes (*IGH*) in CLL/MBL cases and the presence of a polyclonal population in controls.

3.3.2. *IGH* Usage and Mutation Analysis

Analysis of *IGH* usage and mutation status was performed in collaboration with Ms. Maryam Hassanvand, Sydney Medical School Nepean, University of Sydney. Genomic DNA was extracted from peripheral blood using the Wizard genomic DNA purification kit according to the manufacturer's instructions and quantified using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Amplification by polymerase chain reaction (PCR) and sequence analysis of *IGH* rearrangements were conducted according to BIOMED-2 protocols using commercially available *IGH* gene clonality master mixes (InVivoScribe Technologies, San Diego, CA, USA) (van Dongen et al., 2003, van Krieken et al., 2007). Agarose gel electrophoresis was performed and the PCR products were visualized using SYBR Safe staining (Thermo Fisher Scientific, Waltham, MA, USA). The Wizard® SV Gel and PCR Clean-Up System was used to purify PCR products excised from the gel. The purified DNA was sequenced using a 3' JH consensus primer at the Australian Genome Research Facility, Brisbane, Australia.

The percent mutated *IGH* was determined by comparing the VH region sequence to the Ig blast database (US National Library of Medicine, National Center for Biotechnology Information). To ensure accuracy of the results, Ig blast GenBank and the IMGT/V-QUEST

portal for immunoglobulin and T cell receptor sequences (the International ImMunoGeneTics Information System) were used to analyse and align *IGH* sequences (Ghia et al., 2007). In accordance with published criteria, sequences with a germline homology ≥98% were considered as unmutated, and those displaying homology <98% were designated as mutated (Ghia et al., 2007).

3.3.3. Interphase Fluorescence in Situ Hybridization (FISH)

Interphase Fluorescence in Situ Hybridization (FISH) analyses for common abnormalities associated with B-CLL were performed in affected individuals [n=5, S-CLL 87, S-CLL 88, F-MBL (III-10), F-CLL (IV-5), and F-CLL (IV-2)] using the following probes: *DLEU/LAMP* at 13q14, chromosome 12 centromere, *ATM* at 11q22, and *TP53* at 17p13. Interphase FISH studies were performed based on techniques adapted from the Cytogenetics and the Molecular Genetics Laboratory, the Children's Hospital at Westmead, Sydney, NSW, Australia, as described in 2.5. A total of two-hundred images of interphase nuclei were captured for every probe set according to the manufacturer's instructions. Results were considered to be abnormal when the percentage of cells with any given abnormality was > 5% in 200 interphase nuclei for trisomy 12 and > 8% for deletions of 13q, 11q and 17p. A case was considered to have a poor prognosis if the nuclei carried an 11q or 17p aberration, an intermediate prognosis if trisomy 12 was detected and a good prognosis if a 13q deletion was detected (Wawrzyniak et al., 2014) (Wolff et al., 2007).

3.3.4. RNA Extraction

Purified cells from all selected samples (n=21) were extracted from heparinised blood by adding RossetteSep cocktail and centrifugation on Ficoll-Histopaque as described in 2.2.2. A minimum of 1 x 10^6 B-cells/aliquot were collected for gene expression profiles. Lysis buffer

(RLY) (350 µL) (Bioline, Taunton, MA, Cat No.BIO-52072) and 2-mercaptoethanol (β-ME) (3.5 µL) (Promega, Australia, Product code Z523A-C) were immediately added to the purified B cells and placed directly into the −80 °C freezer until RNA isolation was performed. Total RNA isolation from purified B cells was performed according to the manufacturer's protocol as described in 2.4.2.2.

Since limited numbers of cells were available from MBL and control samples it was necessary to optimise the RNA extraction procedure. Two extraction methods: TRI reagent (Sigma-Aldrich, Australia, Cat No. T9424) and Isolate II RNA mini kit (Bioline, Taunton, MA, Cat No.BIO-52072) were compared to determine the best method to obtain high yield and quality RNA.

RNA was extracted from PBMCs cells (5 x 10^6) preserved in RNA*later*®, as described in section 2.4, using TRI Reagent and an Isolate II mini RNA kit according to the manufacturer's protocol. The quality and quantity of purified RNA was assessed by spectrophotometry at A260 and A280 nm.

The yield and quality of RNA isolated using the isolate II RNA mini spin column kit (Bioline, Taunton, MA, Cat No.BIO-52072) was higher compared to the TRI kit. Ultraviolet (UV) spectrophotometer analysis showed that the purity of RNA was greater with the isolate II RNA mini kit (average A260/280 ratio =1.88 for isolate II RNA mini spin column isolation kit versus 1.22 for TRI isolation kit (Figure 9-4). The RNA yields/µg isolated using the spin column method were > 2.4/µg, compare to 2/µg using TRI reagent. Similarly, the concentration of RNA was higher in samples using the spin column compared to those obtained with the TRI protocol. Therefore, for gene expression profiles the Isolate II RNA mini spin column kit (Bioline, Taunton, MA, Cat No.BIO-52072) was used in this study to extract RNA from the purified B

lymphocytes.

Before gene expression microarray analysis, the quantity and purity of total RNA was checked. Samples were quantified and purity determined using a Thermo Scientific Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, U.S.A.). The concentration of total RNA (ng/µL) was determined at 260 nm and was used to calculate the total RNA yield (µg) (Figure 9-5). The RNA purity was assessed by measuring absorbance at 260 nm and at 280 nm. Samples > 1.8 at A260/A280 ratio were used for Affymetrix gene expression microarray analyses. All samples had A260/280 ratios above 2.0 as reported in (Figure 9-6).

Additional RNA quality assessments were performed using the Agilent 2200 TapeStation and 2100 Bioanalyser to determine ribosomal RNA ratio (28S/18S) and RNA integrity Number (RIN). High quality RNA samples were identified by measuring the fraction of the area under the 28S and 18S rRNA peaks of the electropherogram. This fraction should show well-defined 28S and 18S peaks with a 2:1 ratio, indicating that the RNA is completely intact. The RNA integrity number (RIN) provides a numerical quality value for RNA and a score from 1 to 10, with 1 representing the most degraded RNA profile and 10 being most intact. An RIN > 7 is recommended for microarray analysis or qRT-PCR (Affymetrix). All analysed RNA samples had a 28S/18S rRNA ratio > 2 and RIN > 7 (Figure 9-7; Figure 9-8).

3.3.5. Transcriptome Profiling

Total RNA (10 ng) from each sample was prepared as described for the GeneChip® WT Pico Reagent Kit (Affymetrix Inc, Santa Clara, CA, USA) and analysed using Affymetrix GeneChip® Human Transcriptome 2.0 Arrays. The arrays were washed, hybridized and scanned using Cartridge array kits (Affymetrix Inc, Santa Clara, CA, USA).

The quality of each Affymetrix HTA microarray was determined using Affymetrix spike-

in controls, perfect match expression and relative log expression (RLE) during data summarization and normalization in the Affymetrix expression console (EC) software, version 1.4.1. The Affymetrix transcriptome analysis console (TAC 3.0) software was used to perform statistical analysis and generate a list of differentially expressed mRNA.

3.3.6. Affymetrix Expression Console (EC) Software

Quality control checks for evaluating hybridisation on all files (.CEL) were performed using the expression console software, version 1.4.1, and the data were normalized using default settings to remove non-biological variants. Graphs of the QC metrics were generated to identify any potential outliers that might indicate problems with sample quality, hybridisation, labelling reactions or technical errors across all arrays (Figure 9-9; Figure 9-10; Figure 9-11; Figure 9-12).

3.3.7. Monitoring Sample Quality:

Sample quality metrics *pos_* vs_ *neg_*auc, were monitored as a first pass metric for overall data quality. This metric compares signal values for positive controls to the negative controls, a value < 0.8 indicates poor-quality data (Affymetrix). All arrays were > 0.9 (Figure 9-9).

Absolute Relative log expression *(RLE)* mean was also used to detect outlier arrays. This metric calculates the differences between the signals of one array against the median signal value of all arrays. *RLE mean* should be consistently low for biological replicates and arrays with a significantly higher signal value indicate outlier samples (Affymetrix).

3.3.8. Monitoring Hybridisation and Labelling Quality:

Hybridisation controls were used as an additional quality metric to ensure

hybridisation had correctly occurred on the arrays. Hybridization metrics showed the bacterial spike positive controls were displayed in the expected order: *BioB* < *BioC* < *BioD* < *Cre.* This indicated hybridisation had occurred correctly in all arrays (Figure 9-13).

To assess the efficiency of labelling reactions, labelling metrics were performed which showed the Poly-A spike controls were spiked in the correct rank order starting from *Lys* as the lowest*, Phe, Thr, to Dap* as the highest in most arrays. However, in one chip [F-CLL91 (IV-2)] the poly-A spikes (*Lys* and *Phe*) did not fall within the expected order (Figure 9-14). The bacterial spikes and positive controls for the relevant chips were repeated and the results were correct, suggesting the previous problem with the poly-A spikes were specific to those controls and may have indicated a pipetting error in their preparation (Jaksik et al., 2015).

3.3.9. Validation of Gene Microarrays Using qRT-PCR

To validate primers for use in qRT-PCR (Table 2-9), pairs were checked for specificity, appropriate amplicon size and for the absence of SNPs using Primer-blast (US National Library of Medicine, National Center for Biotechnology Information) and SNPCheck3 (National Genetics Reference Laboratory, Manchester, UK), respectively. cDNA synthesis of RNA (1 µg) was performed using a Tetro cDNA synthesis kit (Bioline, Taunton, MA, USA) according to the manufacturer's protocol using a 1:1 ratio of random hexamers and oligo dT for priming. RTPCR was performed using a Rotorgene 2000 (Corbett Research, Sydney, NSW, Australia) as described previously (Skarratt and Fuller, 2014), except annealing conditions were optimised to 58°C for 15 s, and the extension conditions to 72 C for 10 s over 35 cycles. Amplified products were verified by sequencing of gel purified amplicons.

3.3.10. Ingenuity Pathway Analysis

To identify relevant biological functions and pathways, differentially expressed genes (n = 1678) were exported to Microsoft Excel (Microsoft Corporation, Redmond, WA) and uploaded into IPA software using standard settings and criteria restricted to human, immune cells, mononuclear cells, lymphocyte, B-lymphocytes and peripheral blood lymphocytes. Core analysis was run to find the most significant interactions and associations in datasets compared to the IPA database. To determine the association of identified genes with the canonical pathways, significance values were calculated using both a right tailed Fisher's exact test p-value < 0.05, and the ratio of the number of genes involved in the canonical pathway, divided by the total number of genes in this pathway. Significance p-values were set by default at the -log10 (0.05) cut-off of 1.3, which indicated > 95% confidence that genes were not generated by chance. The score was used to rank networks and therefore the higher focus molecules in a network generated higher scores.

3.3.11. Statistical and Bioinformatics Analyses

Assessments for differences in gene expression between F-MBL, F-CLL, S-CLL and controls were performed using analysis of variance (ANOVA), and false discovery rate (FDR) p-values. Unsupervised hierarchical clustering was performed using Affymetrix software (Euclidean distance) (Pounds, 2006, Reiner et al., 2003). In the F-MBL, F-CLL, S-CLL analysis, F-MBL subjects and F-CLL clustered together as an F-LPD group, and within this group, F-MBL subjects were not differentiated from F-CLL. Therefore, for subsequent analyses, F-MBL and F-CLL were considered a single F-LPD group.

In order to identify genes that separated F-LPD from S-CLL which were not due to genetic relatedness between family members, mRNAs differentially abundant in F-LPD

compared to both S-CLL and family controls were identified, and semi-supervised clustering was performed using these mRNAs (Bair, 2013, Bair and Tibshirani, 2004). First, a 2-sample t-test was performed to compare F-LPD and S-CLL groups. All genes that had a log fold change > 2, and FDR p-value < 0.01 were checked to ensure that they were also different on a 2-sample t-test comparison between (A) controls and S-CLL and (B) controls and F-LPD. mRNAs that had FDR p-values < 0.01 were considered for semi-supervised hierarchical clustering (Bair, 2013, Bair and Tibshirani, 2004). Fourteen mRNAs and one miRNA met these criteria and were clustered using Affymetrix software.

3.4. RESULTS

3.4.1. Patients and Samples

Peripheral blood samples were collected from 6 surviving patients (two with F-CLL, and 4 with F-MBL), and 3 unaffected family members from a family with multiple cases of F-LPD (Figure 3-1) (Fuller et al., 2008). Blood samples were collected from a further 3 normal unrelated controls, and 6 S-CLL cases. All CLL subjects were treatment naïve (Table 3-2). B lymphocytes were enriched to provide >95% B lymphocyte purity, confirmed using flow cytometry (Figure 9-1) (Essakali et al., 2008). PCR was used to confirm a clonal B lymphocyte population by the presence of identical rearrangements of immunoglobulin heavy chain variable genes (*IGH*) in CLL/MBL cases and the presence of a polyclonal population in controls.

3.4.2. Interphase Fluorescence in Situ Hybridization (FISH) Results

FISH analyses were performed in 4 B-CLL cases and 1 MBL (S-CLL 87, S-CLL 88, F-MBL III-10, F-CLL IV-5, and F-CLL IV-2). Among these samples, two cases showed del*TP53* with ≥ 30% abnormal cells, one (S-CLL 88) was 50% del*TP53* combined with del (13q14) (13%) and

another (S-CLL 87) was 30% del*TP53*. Del13q14 (30% of cells) alone was detected in 1 case (F-CLL IV-5) and no chromosomal abnormalities were detected in the remaining 2 cases (F-MBL III-10 and F-CLL IV-2) (Figure 9-3).

3.4.3. *IGH* Gene Clonality Assay Results

IGH gene clonality analyses were performed in 6 S-CLL cases, 2 F-CLL and 4 F-MBL (Table 3-2). Clonal populations of B lymphocytes were identified in all cases. Three of 6 S-CLL cases (S-CLL 53, S-CLL87 and S-CLL88) and 1 of 6 F-LPD cases (F-CLL IV-2) had UM-CLL. Profiles for mRNA and protein expression in S-CLL and F-CLL based on *IGH* and familial M- and UM-CLL will be analysed in Chapter 6.

Table 3-2 Summary of sample characteristics

Characteristic	Unrelated control (n=3)	Family control (n=3)	S-CLL (n=6)	F-LPD (n=6)
F-CLL	-	-	-	2 (33%)
F-MBL	-	-	-	4 (66%)
Male gender	2 (66%)	1 (33%)	2 (33%)	3 (50%)
Age*	Mean 49, SD 5	Mean 48 SD 8	Mean 74. SD 12	Mean 60. SD 10
Absolute lymphocyte count (x 10^9/L)	Mean 2. SD 0.4	Mean 2. SD 0.2	Mean 41. SD 35	Mean 6. SD 7
FISH analysis				
• Del*TP53*	-	-	2	
• Del13q14	-	-		1
• Normal	-	-		2
IGHV mutational status				
• Mutated	-	-	3	5
• Unmutated	-	-	3	1
IGHV region containing the mutation				
• Unknown	-	-	2	1
• IGH1	-	-	1	-
• IGH2	-	-	1	
• IGH3	-	-	2	2
• IGH4	-	-	-	2
• IGH5	-	-	-	1

* The S-CLL cases were older than the F-LPD (mean 74 versus 60; Student's t-test p=0.05), both groups were matched for *IGHV* mutation status (F-CLL: 1 mutated, I unmutated; S-CLL: 3 mutated, 3 unmutated). Ages were similar between the unrelated and family normal control groups (mean 48 for both groups). S-CLL; sporadic chronic lymphocytic leukaemia, F-CLL; familial chronic lymphocytic leukaemia, F-MBL; monoclonal B-cell lymphocytosis (F-MBL) family members, SD: standard deviation

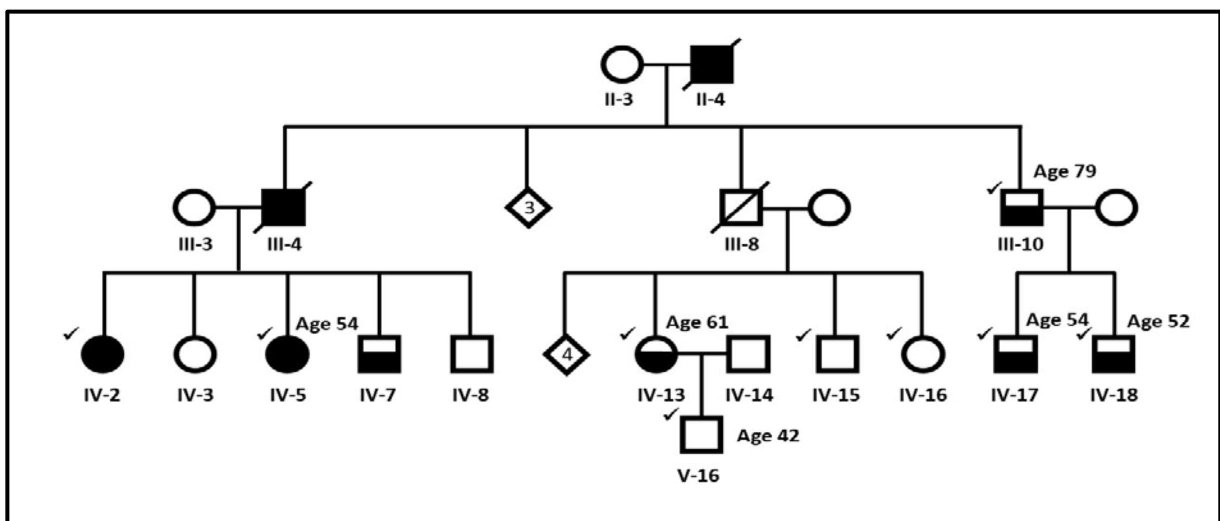

Figure 3-1 Pedigree of the family.

The pedigree abbreviated from (Figure 1-1) shows segregation of CLL. Blackened symbols are individuals affected with CLL; ticked symbols denote individuals studied from whom mRNA and protein were collected; half-shaded symbols denote family members with MBL; diamonds represent grouped siblings.

3.4.4. Unsupervised Hierarchical Clustering of Gene Expression Data for Family Members with MBL or CLL and Sporadic CLL Cases

Purified B lymphocytes from F-MBL and F-CLL cases, S-CLL cases and related and unrelated control subjects, were compared. The control samples were from 3 family members and 3 unrelated controls, there were 2 F-CLL cases, 4 F-MBL, and 6 S-CLL. Comparison of all 3 groups together identified 1893 cDNA elements, representing 1678 genes, which were differentially abundant (FDR < 0.05). Hierarchical clustering segregated cases of F-MBL and F-CLL cases from S-CLL and controls (Figure 3-2). In this analysis, F-CLL and F-MBL cases did not segregate and were combined as “F-LPD” for subsequent comparisons. Unsupervised hierarchical clustering segregated related normal controls (V-16, IV-15 and IV-16) from unrelated normal controls (1 - 3 NC). Based on FDR, the highest ranked genes upregulated in S-CLL versus normal controls and further upregulated in F-LPD were *LEF1* (p = 9.69E-09), *ROR1* (p = 2.54E-09). *ABCA6* (p = 2.54E-08) and *MIR4524A* (p = 4.80E-07). The most highly ranked genes downregulated in S-CLL versus normal controls and further downregulated in F-LPD were *SH3RF1* (p<0.0001), *PLD4* (p<0.0001), *FAM135A* (p<0.0001), and *SNX22* (p<0.0001). Consistent with previous studies, upregulation of *BCL2* was found in B-lymphocytes from S-CLL and F-LPD (p = 0.002).

These findings confirmed that gene expression profiles in B lymphocytes from F-CLL and F-MBL, combined as F-LPD cases, were unique compared to B lymphocytes from unaffected subjects and sporadic CLL cases. Furthermore, gene expression profiles segregated family normal controls and unrelated controls.

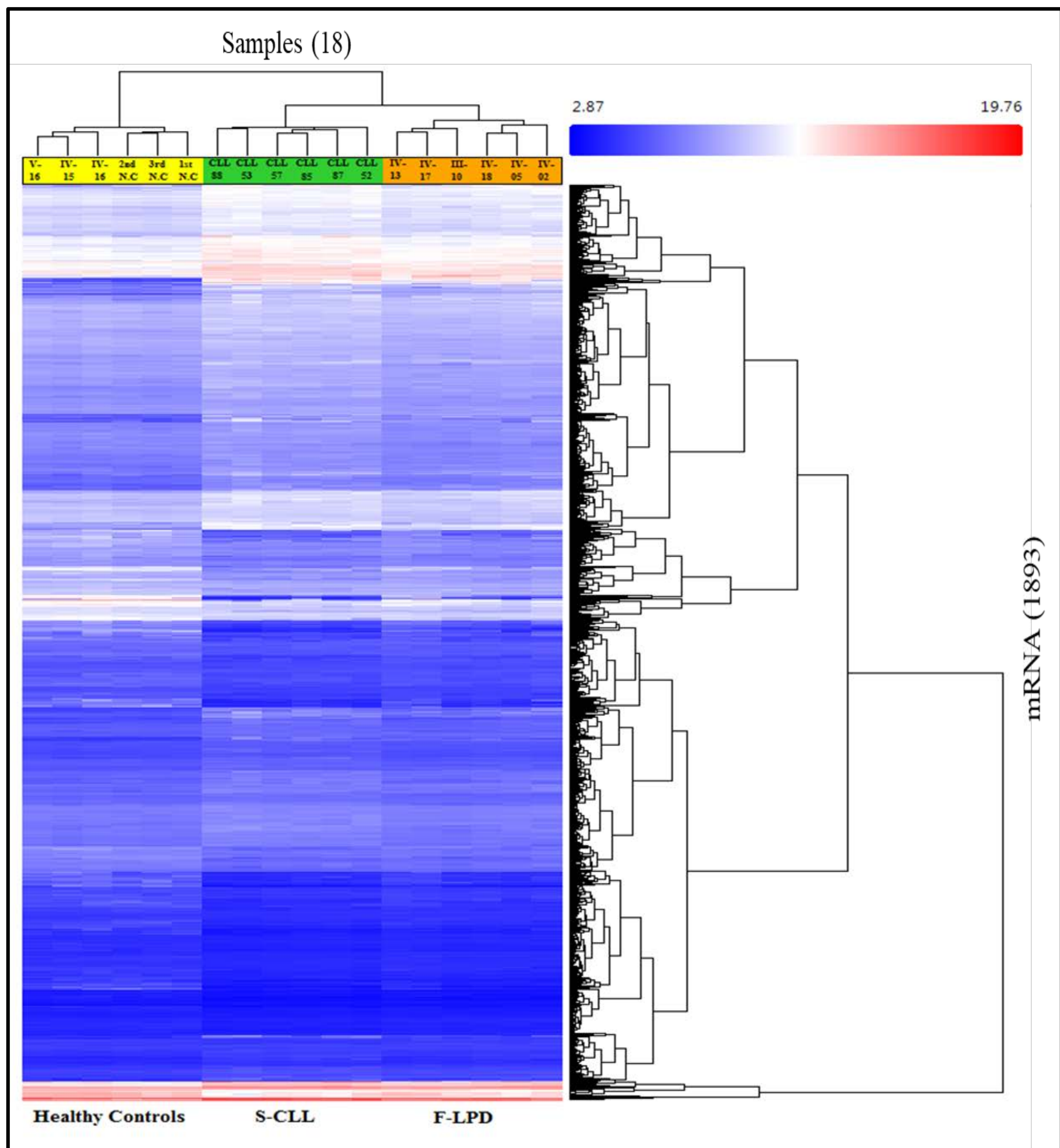

Figure 3-2 Unsupervised hierarchical clustering of B lymphocyte gene expression

Unsupervised hierarchical clustering of 18 samples from familial-lymphoproliferative disease (F-LPD), sporadic chronic lymphocytic leukemia (S-CLL) and related and unrelated control subjects. Array elements that significantly varied (FDR < 0.05) were included (1893 mRNAs). Data are displayed as a heat map where rows represent unique cDNA elements and columns represent experimental samples. Coloured pixels capture the magnitude of the response for any gene, where shades of red and blue represent induction and repression, respectively, relative to the median for each gene. *IGH* genes did not cluster and were removed from the analysis. The cluster dendrograms at the right segregated healthy controls, F-LPD, and S-CLL.

3.4.5.Ingenuity Pathway Analysis

Comparing F-LPD, S-CLL and normal controls using the IPA application, the top five canonical pathways were identified for the datasets. These included xenobiotic metabolism signalling (p-value = 5.34E-04), pancreatic adenocarcinoma signalling (p-value = 6.99E-04), chronic myeloid leukemia signalling (p-value = 1.81E-03), 3-phosphoinositide biosynthesis (3.58E-03) and super-pathway of inositol phosphate compounds (3.85E-03), with the following top upstream regulators ordered by overlap p-value: CD44 (3.49E-02), IGHM (3.49E-02), RELA (3.49E-02), and IL15 (4.27E-02). Molecular and cellular functions differing between these groups included cell signalling, molecular transport, vitamin and mineral metabolism, cell cycle, and gene expression. The top scoring network showed a score of 10 with 15 network eligible molecules involved in cancer, cell death and survival, organismal injury and abnormalities. A summary of the IPA results for 1893 cDNA elements differentially abundant (FDR < 0.05) between F-LPD, S-CLL and normal controls is shown in (Table 3-3).

Table 3-3 A summary of the IPA results for 1893 cDNA elements differentially abundant (FDR < 0.05) between F-LPD, S-CLL and normal controls

Top Canonical Pathways	**p-value**	**Overlap ***	**Ratio ** **
Xenobiotic Metabolism Signalling	5.34E-04	33/163	0.20
Pancreatic Adenocarcinoma Signalling	6.99E-04	21/89	0.23
Chronic Myeloid Leukaemia Signalling	1.81E-03	19/83	0.22
3-phosphoinositide Biosynthesis	3.58E-03	25/127	0.19
Superpathway of Inositol Phosphate Compounds	3.85E-03	28/148	0.18
Upstream Regulator	**p-value of overlap**	**Predicted activation**	
CD44	3.49E-02		
IGHM	3.49E-02		
RELA	3.49E-02		
IL15	4.27E-02		
Molecular and Cellular Functions	**p-value**	**#Molecules**	
Cell Signalling	7.69E-03	4	
Molecular transport	7.69E-03	4	
Vitamin and mineral metabolism	7.69E-03	4	
Cell cycle	3.51E-02	2	
Gene expression	3.51E-02	2	
Physiological System Development and Function	**p-value**	**#Molecules**	
Embryonic development	4.97E-03	4	
Haematological System Development and Function	4.97E-03	4	
Haematopoiesis	4.97E-03	4	
Humoral Immune Response	4.97E-03	4	
Lymphoid Tissue Structure and Development	4.97E-03	4	
Top Networks		**Score**	**Focus Molecules**
1. Cancer, Cell Death and Survival, Organismal Injury and Abnormalities		10	15
2. Cellular Development, Cellular Growth and Proliferation, Haematological System Development and Function		9	19
3. Cardiovascular Disease, Gastrointestinal Disease, Hepatic System Disease		1	2
4. Cellular Movement, Cellular Assembly and Organization, Cellular Function and Maintenance		1	1
5. Cell-To-Cell Signalling and Interaction, Cancer, Cellular Movement1		1	1

* Overlap column indicates the number of observed molecules from our dataset that met the filter criteria and participate in a canonical pathway to the total number of molecules that participate in the same canonical pathway from the Ingenuity knowledge base.

** The Ratio column indicate by taking the number of observed molecules from our dataset that participate in a canonical pathway divided by the total number of molecules that participate in the same canonical pathway from the Ingenuity knowledge base.

3.4.6. Semi-Supervised Hierarchical Clustering of Gene Expression Data for S-CLL Cases and F-LPD

From 1893 differentially expressed mRNAs with FDR $p < 0.05$, differential expression analysis was used to generate a list of F-LPD mRNAs which had been corrected for genetic relatedness. These mRNAs were used for semi-supervised hierarchical clustering (Bair, 2013, Bair and Tibshirani, 2004). To exclude mRNAs that were similarly abundant as a result of being part of the same family, mRNAs that were differentially abundant in F-LPD versus S-CLL, and also differentially abundant in F-LPD versus controls were identified (Figure 3-3).

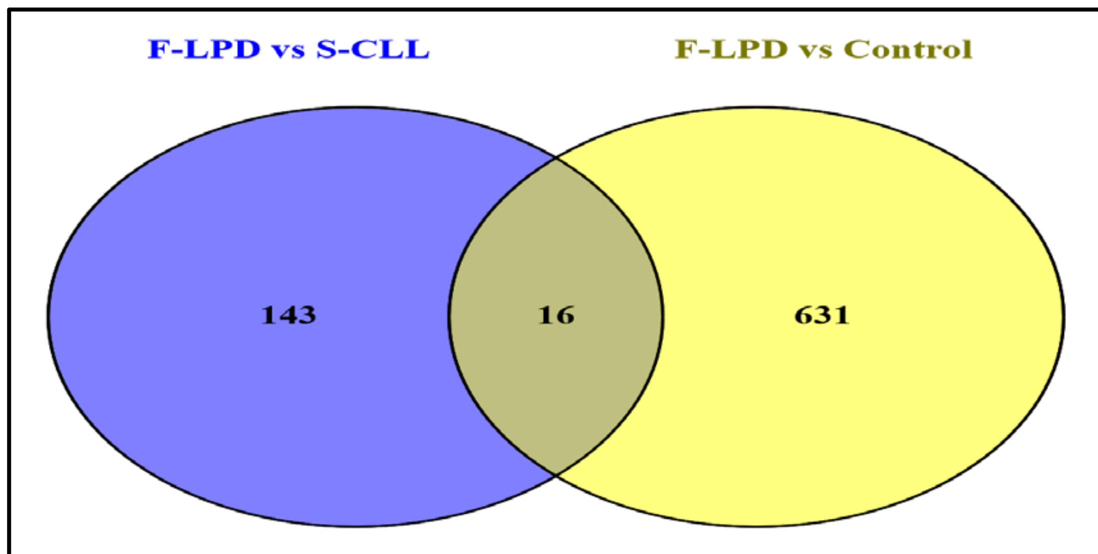

Figure 3-3. Venn diagram illustrating overlap between F-LPD vs S-CLL, and F-LPD vs Controls

Sixteen mRNAs were differentially abundant with t-test $p < 0.01$; >2-fold change in F-LPD versus S-CLL, and also differentially abundant in F-LPD versus controls. F-LPD; familial-lymphoproliferative disease, S-CLL; sporadic chronic lymphocytic leukaemia. Retrieved from http://bioinfogp.cnb.csic.es/tools/venny/index.html on March 7th 2018.

For the F-LPD versus S-CLL comparison, 16 mRNAs representing 14 genes and 1 microRNA (miRNA) were also differentially expressed in the F-LPD versus control comparison. Of these, 14 genes and 1 miRNA showed a 2-fold change in regulation (Figure 3-4; Table 3-4).

Inclusion of *ITGA4* and *PAG1* produced hierarchical clustering of one F-LPD patient (IV-05) with S-CLL patients (Figure 3-4).

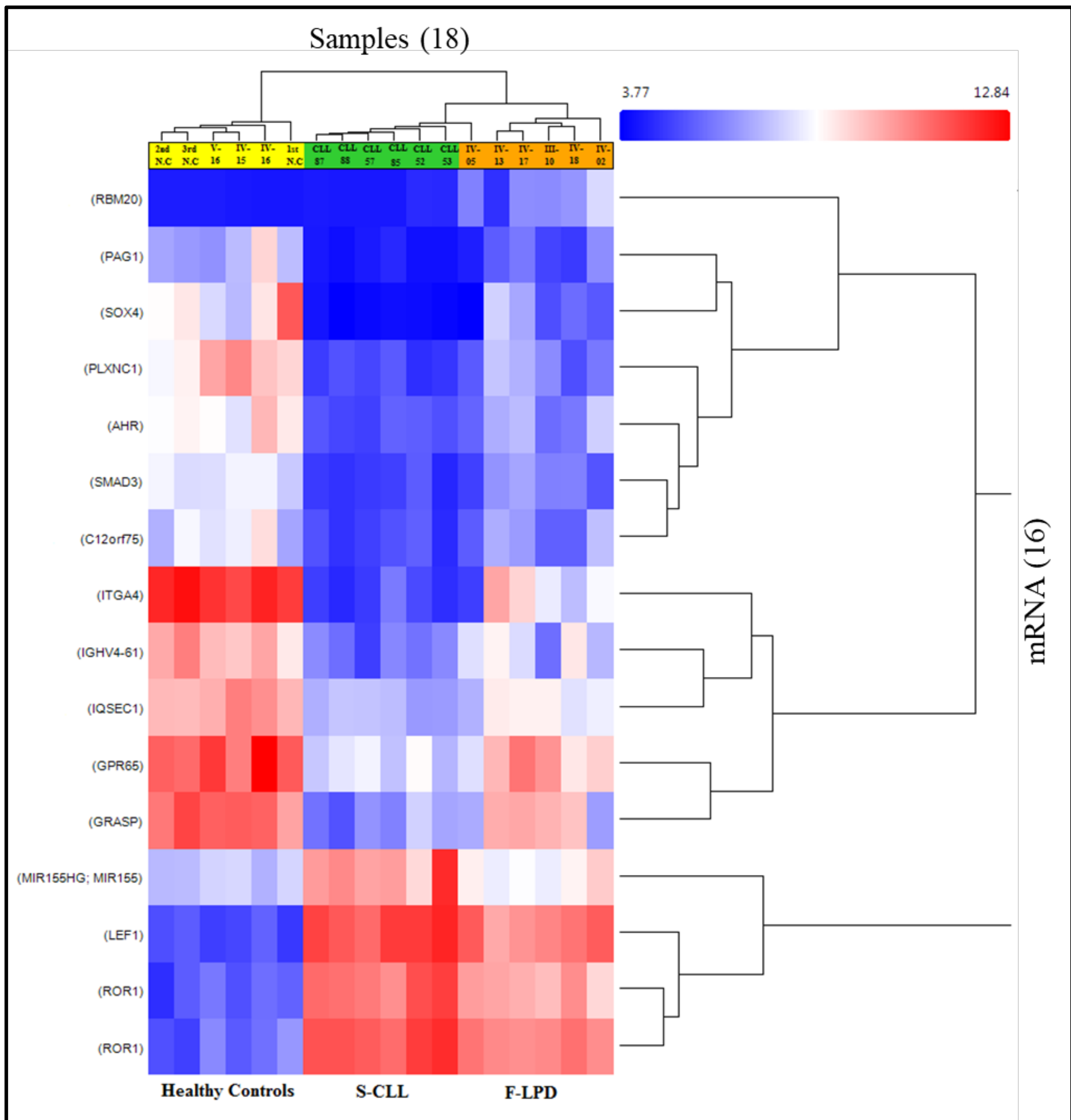

Figure 3-4. Semi-supervised hierarchical clustering of B lymphocyte mRNA levels following correction for relatedness

Supervised hierarchical clustering of 14 genes and 1 microRNA (miRNA) which showed a >2-fold change (t-test $p < 0.01$) in regulation between normal familial, F-LPD and S-CLL. Inclusion of *ITGA4* and *PAG1* produced hierarchical clustering of one F-LPD patient (IV-05) with S-CLL.

Increased *ITGA4* (CD49) mRNA has been correlated with unmutated *IGVH* genes and poor prognosis (Baumann et al., 2016). The mean mRNA level of *ITGA4* in the F-LPD group was higher compared with S-CLL, however F-CLL (IV-05), who had mutated *IGVH* genes, had low *ITGA4* mRNA levels. In addition, F-CLL (IV-05) had low *PAG1* expression, which has been found to be downregulated in MBL cells compared to memory B cells (Lanasa et al., 2011). The combination of low *PAG1* and *ITGA4* expression in F-CLL (IV-05) resulted in this patient clustering with S-CLL patients. On the assumption that downregulation of *PAG1* was related to progression from normal B lymphocytes to MBL, rather than associated with F-LPD *per se*, and low *ITGA4* expression in F-CLL (IV-05) was associated with mutated *IGH*, both *PAG1* and *ITGA4* were removed from the final semi-supervised hierarchical clustering analysis.

Following removal of *ITGA4* and *PAG1*, semi-supervised hierarchical clustering using gene expression data for 12 genes and 1 miRNA from the control familial samples (n=3), unrelated control (n=3), F-CLL (n=6) and S-CLL (n=6) samples segregated healthy controls, F-LPD, and S-CLL (Table 3-4; Figure 3-5).

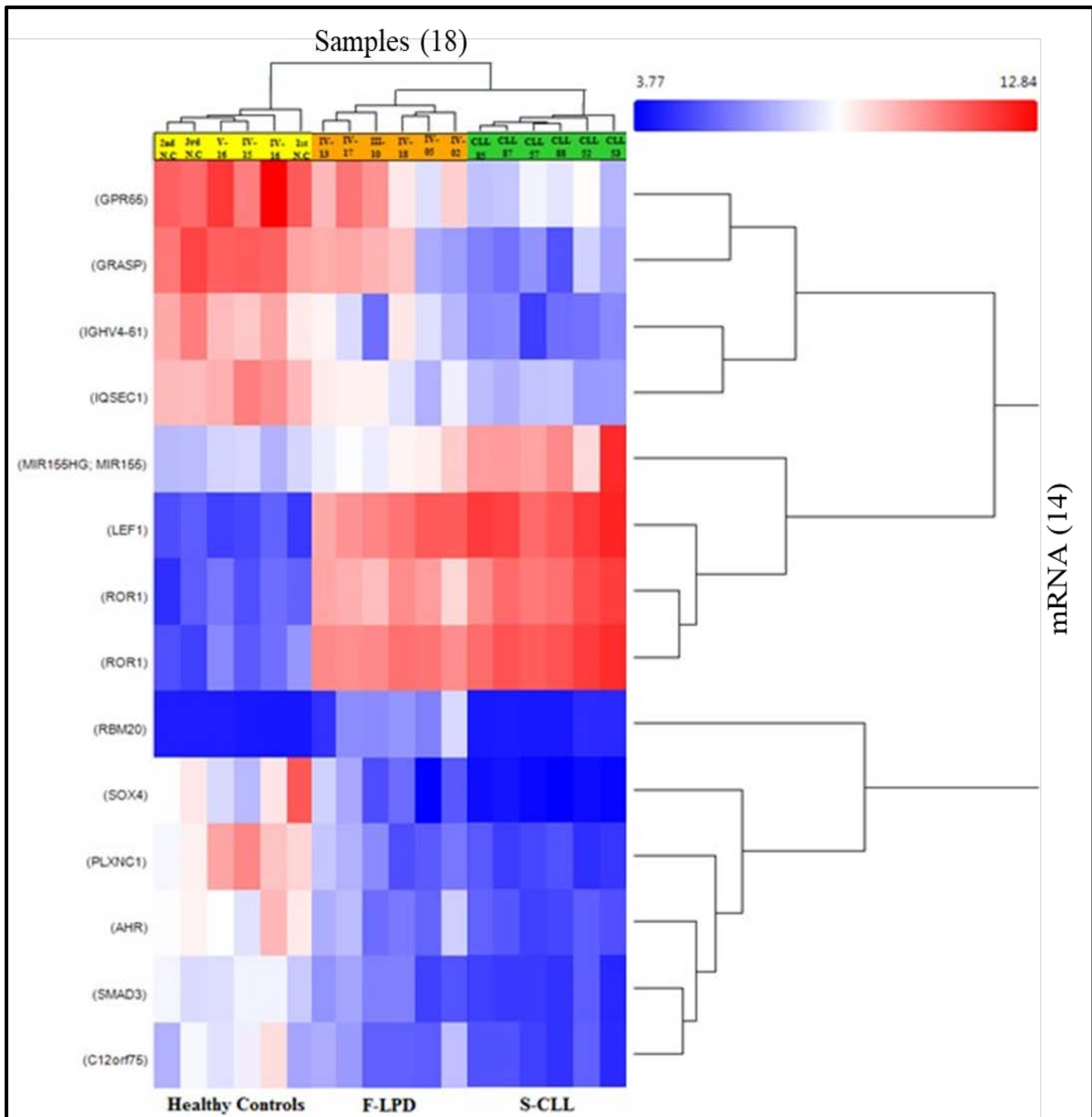

Figure 3-5. Semi-supervised hierarchical clustering of B lymphocyte mRNA levels following removal of PAG1 and ITGA4

Supervised hierarchical clustering was performed using expression data for 12 genes and 1 microRNA (miRNA) from control family samples (n=3), unrelated controls (n=3), F-CLL (n=6) and S-CLL (n=6). Twelve genes and 1 microRNA (miRNA) showed a >2-fold change in level. The cluster dendrogram segregated healthy controls, F-LPD, and S-CLL.

Table 3-4 Differential levels of mRNAs from grouped data for F-LPD and S-CLL

mRNA	Gene name	Fold change	p-value	Cellular location	Function (Molecule Types)
GRASP	*general receptor for phosphoinositides 1 associated scaffold protein*	10.6	0.007	Plasma membrane	other
ITGA4	*integrin subunit alpha 4*	10.4	0.003	Plasma membrane	transmembrane receptor
RBM20	*RNA binding motif protein 20*	4.1	0.0007	Nucleus	Other
GPR65	*G protein-coupled receptor 65*	3.4	0.006	Plasma membrane	G-protein coupled receptor
SOX4	*SRY-box 4*	3.2	0.009	Nucleus	transcription regulator
IGH4-61	*immunoglobulin heavy variable 4-61*	3.2	0.006	Other	other
IQSEC1	*IQ motif and Sec7 domain 1*	2.5	0.004	Cytoplasm	Other
SMAD3	*SMAD family member 3*	2.3	0.006	Nucleus	transcription regulator
PLXNC1	*plexin C1*	2.3	0.008	Plasma membrane	transmembrane receptor
AHR	*aryl hydrocarbon receptor*	2.3	0.006	Nucleus	ligand-dependent nuclear receptor
PAG1	*phosphoprotein membrane anchor with glycosphingolipid microdomains 1*	2.2	0.005	Plasma membrane	Other
C12orf75	*chromosome 12 open reading frame 75*	2.1	0.0001	Other	Other
ROR1	*receptor tyrosine kinase like orphan receptor 1*	-2.0	0.0006	Plasma membrane	kinase
LEF1	*lymphoid enhancer binding factor 1*	-2.1	0.007	Nucleus	transcription regulator
ROR1	*receptor tyrosine kinase like orphan receptor 1*	-2.2	0.001	Plasma membrane	kinase
MIR155HG	*MIR155 host gene*	-3.3	0.002	Other	Other

© 2000-2017 QIAGEN. All rights reserved

Compared to normal controls, 2 mRNAs (*LEF1* and *ROR1*) and *MIR155HG* were upregulated in F-LPD and further upregulated in S-CLL cases. Ten genes (*IGH4-61*, *SOX4*, *RBM20*, *SMAD3*, *PLXNC1*, *C12orf75*, *AHR*, *GRASP*, *GPR65/TDAG8* and *IQSEC1*) were downregulated in F-LPD, and further downregulated in S-CLL cases. One gene, *RBM20*, was differentially upregulated in F-LPD compared to both controls and S-CLL. Two genes *(ROR1*, and *LEF1*), have previously been found to be upregulated in CLL, one (*GRASP*) downregulated, and two (*MIR155HG* and *GPR65/TDAG8*) linked to CLL biology (Cui et al., 2014, Rosko et al., 2014, Li et al., 2013, Justus et al., 2017, Cui et al., 2016, Liao et al., 2015, McCarthy et al., 2015).

3.4.7. Validation of Gene Microarrays Using qRT-PCR

To validate genes differentially expressed between the S-CLL and the F-LPD group, qRT-PCR was performed using a Rotor-Gene 2000 (Corbett Research, Sydney, Australia) as described previously in section 2.7.2.

Twelve genes and 1 miRNA were identified as differentially expressed between the S-CLL group and the F-LPD group. Two of these genes, *GRASP* and *SMAD3*, were also identified in this thesis as differentially associated with progression from normal B-lymphocytes to pre-malignant MBL to malignant CLL cells. (See section 5.4). In microarray analysis, these two genes were expressed more highly in F-LPD than in S-CLL. Therefore, *GRASP* and *SMAD3* were selected to validate microarray results using qRT-PCR. To determine the relative expression of these two genes, qRT-PCR was performed on 3 S-CLL (CLL57, CLL87 and CLL88) samples, and 6 F-LPD (2 F-CLL & 4 F-MBL) cases and changes in expression were determined relative to *GAPDH* (delta Ct).

The result showed that expression of *GRASP* and *SMAD3* relative to *GAPDH* was higher in F-LPD compared to S-CLL, which correlated with the microarray results. The qRT-PCR results for these genes are shown in (Figure 3-6).

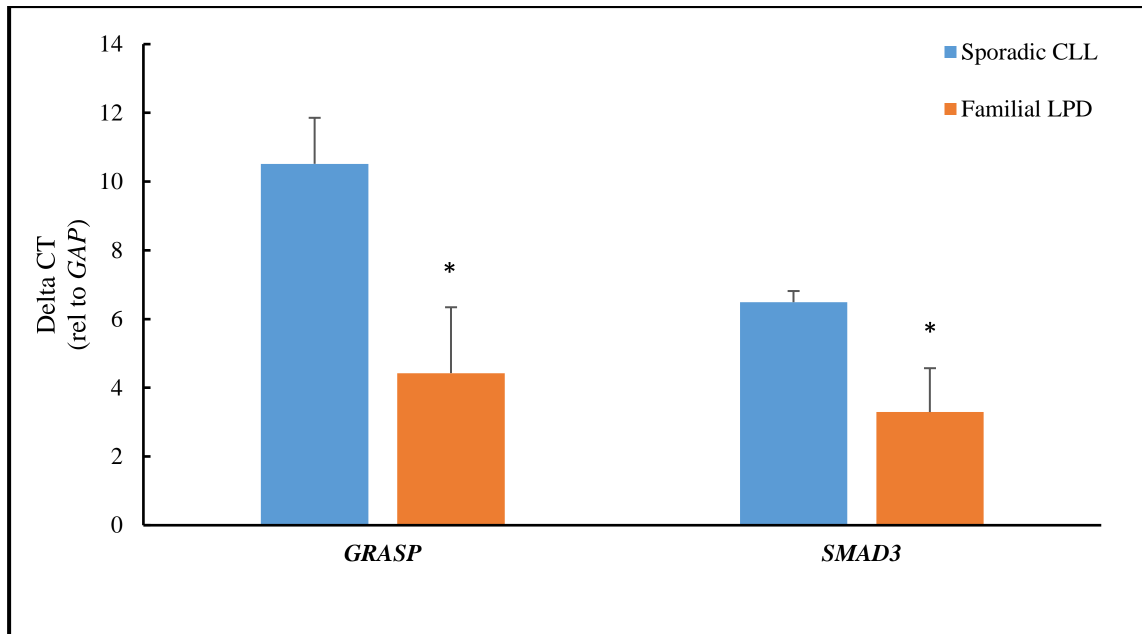

Figure 3-6. Validation of the expression of GRASP and SMAD3 relative to GAPDH (delta Ct) in B-lymphocytes from S-CLL and F-LPD subjects

Delta Ct (cycle threshold) is defined by the number of cycles required for the fluorescent signal to cross threshold background fluorescence. Ct levels are inversely proportional to the amount of target mRNA, that is, the lower the Ct the greater the amount of target mRNA in the sample. Both GRASP and SMAD3 mRNAs were more abundant in F-LPD compared to S-CLL.

3.5. DISCUSSION

Genome-wide association studies have identified a large number of mutations associated with B-CLL. However, these mutations account for only ≈19% of familial risk for developing CLL (Law et al., 2017). Of note, GWAS haplotypes are enriched in regulatory elements including key B-cell transcription binding factor sites (Law et al., 2017). Therefore, it is likely that a proportion of genetic susceptibility to B-CLL results from mutations that affect gene regulation, including transcription factor binding and epigenetic modification, rather than changes in DNA sequences that affect protein function.

Using unsupervised hierarchical clustering, we have shown that mRNA profiles segregate S-CLL from F-LPD. This profile was distinct from those found in normal B-

lymphocytes in family and unrelated controls.

To identify genes differentially expressed between S-CLL and F-LPD, we removed genes that were similarly expressed between normal family controls and F-LPD. An advantage of this family study compared to association studies of unrelated subjects was that background genetic variation, was partially controlled as a function of the degree of relationship, increasing the contribution of epigenetic and/or environmental modifiers to variation in phenotype.

The resultant panel of differentially expressed genes were then studied at an individual level using semi-supervised hierarchical clustering to determine if they segregated normal B-lymphocytes from F-LPD and S-CLL. Twelve genes and 1 miRNA were identified that segregated F-LPD from S-CLL. *LEF1*, *ROR1* and *MIR155HG* were expressed at higher levels in F-LPD and S-CLL compared to normal B-lymphocytes. The following were downregulated in F-LPD and S-CLL: *IGH4-61*, *SOX4*, *RBM20*, *SMAD3*, *PLXNC1*, *C12orf75*, *AHR*, *GRASP*, *GPR65/TDAG8* and *IQSEC1*. Three genes (*ROR1*, *LEF1*, and *GRASP*), have previously been found to be differentially expressed in CLL and are further discussed in Chapter 5 (Cui et al., 2016, Liao et al., 2015, McCarthy et al., 2015), and two additional genes (*MIR155HG* and *GPR65/TDAG8*) have been linked to CLL biology (Cui et al., 2014, Rosko et al., 2014, Li et al., 2013, Justus et al., 2017). Of note, *GPR65/TDAG8*, is located on chromosome 14q31.3, close to the chromosomal region 14q24.1-14q31.2 identified in a previous linkage study of this CLL family (Fuller et al., 2008).

MiR-155 is a noncoding RNA that plays an essential role in the regulation of gene expression. It has been found to have multifunctional roles in several biological processes, and is implicated in several diseases (Faraoni et al., 2009). *MiR-155* has been found to be

upregulated in leukaemia and lymphoma, suggesting a direct or indirect role of *miR-155* in synthesising a protein with tumour-suppressor or pro-apoptotic function (Eis et al., 2005). An early study by (Costinean et al., 2006) reported that *miR-155* was enhanced in B-cell precursors, which induced a pre-lymphoproliferative disease and later B-cell malignancy in *Eμ-mmu-miR155* transgenic mice. In CLL patients, *miR-155* has been described as a signature marker of CLL in combination with the most common genomic aberrations (Rossi et al., 2010, Visone et al., 2009). Another study by (Cui et al., 2014), found that high expression of *miR-155* can be used as an additional independent prognostic value to segregate patients at relative risk for disease progression, but levels of *miR-155* can vary between CLL cells in the same patients. In the same study, the authors found that B-CLL cells with high expression of *miR-155* were more responsive to BCR ligation compared to B-CLL cells with low expression levels of *miR-155*, suggesting a role in the regulation of the BCR signalling pathway in CLL cells and disease progression.

T-cell death-associated gene 8 (*TDAG8*, also known as *GPR65*) is a member of the proton-sensing G-protein-coupled receptor family which can be activated by extracellular acidosis. It has been mapped to chromosome 14q31-32.1, where cytogenetic abnormalities in T cell lymphoma and leukaemia are located (Justus et al., 2017). It is also located close to the chromosomal region 14q24.1-14q31.2 identified in the previous linkage study of this CLL family (Fuller et al., 2008). The oncogenic activity of *GPR65/TDAG8* was reported by Ihara *et al.*, who found that overexpression of *GPR65/TDAG8* on the surface of tumour cells facilitates tumour development by sensing the acidic environment (Ihara et al., 2010). Another report has focused on the role of *GPR65/TDAG8* expression in CLL cells and the potential relationship of its expression to various anti- and pro-apoptotic Bcl-2 family members. This study found a correlation between *GPR65/TDAG8* expression and the anti-apoptotic proteins Bcl-2, Mcl-1

and Bcl-xl, whereas no correlation was detected between *GPR65/TDAG8* and the pro-apoptotic proteins BIM, PUMA or NOXA. This study suggested that the expression of *GPR65/TDAG8* can have a potential function in the survival of CLL cell in the microenvironment (Rosko et al., 2014). The expression of *GPR65/TDAG8* has recently been identified in multiple forms of blood cancers including B-CLL, and *GPR65/TDAG8* gene expression was reduced 2.9-fold in B-CLL compared to normal blood cells. This work suggested that *GPR65/TDAG8* acts as a tumour suppressor by mediating Gα13 G-protein/Rho GTPase signalling to reduce *c-myc* oncogene expression (Justus et al., 2017).

SOX4 (SRY-related HMG-box) is another candidate gene differentially expressed in F-LPD and S-CLL. It has been found that the *SOX4* transcription factor mediates early B-cell differentiation and knock-out of *SOX4* leads to arrested B-cell development at the pro-B cell stage (Wetering et al., 1993, Schilham et al., 1996). Another study suggested that *SOX4* is required for survival of pro-B and pre-B cells but not required for the survival of later stage B cells. This work also suggested a functional role of *SOX4* in protecting pro-B cells from apoptosis by interacting with c-kit and Bcl-2 (Sun et al., 2013). In addition to maintaining survival of B cells, *SOX4* has been reported as a central mediator of oncogenic PI3K/AKT and MAPK signalling in acute lymphoblastic leukaemia (ALL) (Ramezani-Rad et al., 2013). Also, it has been shown to enhance β-catenin/T-cell factor (TCF) complex activity and modulate the transcription of Wnt-target genes in prostate and colon cancers (Liu et al., 2006, Sinner et al., 2007). However, the functional role of *SOX4* was reported to control the stability of β-catenin protein in these carcinoma cells and may act as a transforming oncogene. In a large-scale analysis of DNA methylation in CLL, *SOX4* was discovered, along with other SOX family members, to function as a negative regulator of the WNT signalling pathway (Rahmatpanah et al., 2009). In addition, an integrated genetics approach showed *SOX4* is required at multiple

stages of B cell development by suppressing Wnt/β-catenin signalling and activating immunoglobulin gene recombination (Mallampati et al., 2014). *SOX4* has also been found to have tumour suppressor activity by inducing cell cycle arrest and apoptosis, and inhibiting tumorigenesis in a p53-dependent manner (Pan et al., 2009).

In conclusion, the screening of this B-CLL family using mRNA profiling has identified a number of promising candidate genes associated with the pathogenesis of S-CLL and F-CLL. These differential profiles of mRNAs between categories of F-LPD and S-CLL should be useful for rapid diagnosis and provide a basis for understanding the mechanisms that drive neoplastic transformation. Furthermore, the methods described here could be used for other cancers and complex diseases that show heritability.

CHAPTER- 4: A COMPARISON OF PROTEIN PROFILES IN FAMILIAL MBL, FAMILIAL B-CLL AND SPORADIC B-CLL

4.1. Proteomic Studies in CLL

Proteomics aims to characterize protein structure, expression, and interactions, which determine the temporal and spatial functions of molecules in individual cells under specific conditions (Di Palma et al., 2012). Over the last decade, proteomic techniques have developed rapidly and are now able to characterise large protein datasets in complex mixtures. The protein expression profiles of B-CLL compared to controls have been studied using a number of relative and quantitative proteomic assays (Alsagaby et al., 2014, Huang et al., 2016, Eagle et al., 2015, Perrot et al., 2011). Isobaric tags for relative and absolute quantitation (iTRAQ), a quantitative proteomic method that uses tandem mass spectrometry (MS/MS), identified differential expression of a proteolytic product of histone H2A (cH2A) in B-CLL samples (Glibert et al., 2014). Another study used a liquid chromatography/tandem mass spectrometry (LC-MS/MS) based approach to determine histone profiles in normal B cells and B-CLL cells (Singh et al., 2015). In this study, several histone isoforms were identified that have specific roles in tumour biology, and the presence of specific H2A isoforms were associated with poor prognosis and shorter time to treatment. A specific histone H2A isoform (H2A type 1C) was also elevated in CLL and distinguished CLL from healthy controls (Singh et al., 2015).

In addition to comparing protein profiles between control B lymphocytes and B-CLL, studies have identified protein markers which predict prognosis. Protein profiles associated with prognosis have been identified in primary B-CLL samples using two-dimensional nano-LC coupled with matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry (MALDI-TOF/TOF-MS) and iTRAQ reagents (Alsagaby et al., 2014) . In this study, T-cell leukaemia/lymphoma protein 1A (TCL-1), thyroid hormone receptor-associated protein 3 (TR150), and S100 Calcium Binding Protein A8 (S100A8) were associated with high-risk B-

CLL, while myosin-9 was associated with lower risk disease

Another study identified 84 differentially abundant proteins between stable and progressive CLL using iTRAQ coupled to 2D-LC-MS/MS, and 32 of these proteins were quantified by selected reaction monitoring (SRM) analysis (Huang et al., 2016). Hierarchical clustering analysis showed that patients with progressive disease could be distinguished from those with stable disease. These proteins have roles in cell proliferation, cell death and survival, granzyme A signalling, and DNA repair (Huang et al., 2016).

The proteomes of M-CLL and UM-CLL have been compared using iTRAQ-based MS (Eagle et al., 2015). Differentially expressed proteins between 9 M-CLL and 9 UM-CLL subjects were identified, and functions of the protein subsets were analysed using a system biology approach. This study identifed 3521 proteins, and among these proteins, 274 showed significant differences in abundance between M-CLL and UM-CLL. Of these 274 proteins, 127 were expressed at higher levels and 147 at lower levels in the UM-CLL compared to M-CLL cases (Eagle et al., 2015). The functions of most of these proteins were associated with cell migration/adhesion pathways, of which 35 were expressed at significantly lower levels in the UM-CLL samples. These findings suggested that UM-CLL cells were less migratory and more adhesive than M-CLL cells, resulting in retention in lymph nodes, where the malignant cells are exposed to proliferative and pro-survival signals (Eagle et al., 2015).

Using DotScan CD antibody microarrays followed by validation with LC-MS/MS quantification, a correlation has been identified between cytogenetic alterations and protein expression patterns in B-CLL (Huang et al., 2014). This approach identified downregulation in protein kinase C (PKC) family members, which are involved in cell signalling pathways including apoptosis, cell proliferation and activation (Díez et al., 2016). In enriched B-CLL samples the

antibody microarray identified 27 antigens that were differentially abundant in progressive CLL with an accuracy of 79%, a sensitivity of 84% and specificity of 73% (Huang et al., 2014).

Electrophoresis combined with MS has been used to investigate differences in protein expression between UM-CLL and M-CLL, and $CD38^-$ and $CD38^+$ status (Cochran et al., 2003). Principal components analysis (PCA) identified several proteins that were expressed at higher levels in M-CLL compared to the UM-CLL, including F-actin-capping protein β subunit and laminin-binding protein precursor. However, PCA of $CD38^-$ versus $CD38^+$ did not show any significant change between these two groups (Cochran et al., 2003). A similar study used quantitative 2D-fluorescence difference gel electrophoresis (2D-DIGE) to study tumour cells of six CLL patients (3 M-CLL and 3 UM-CLL) selected according to their *IGH* mutation status and ZAP70 expression (Perrot et al., 2011). In this study, 2D-DIGE was used to compare 48 proteomic profiles of 2 CLL subsets before and after sIgM stimulation, followed by mass spectrometry using MALDI-TOF to identify differentially expressed proteins. This showed that UM-CLL cells display distinct proteomic profiles after BCR stimulation compared with M-CLL cells (Perrot et al., 2011).

In summary, there have been detailed proteomic analyses of cohorts of S-CLL patients to identify proteins which can be used to predict prognosis and guide earlier treatment, but only a limited number of studies which have compared protein expression profiles in control B lymphocytes with S-CLL, and no studies of F-CLL. A proportion of the inherited risk of CLL is likely to be associated with non-DNA sequence, including epigenetic modifications that regulate oncogenes and tumour suppressor genes, described in a number of hereditary cancers (Gazzoli et al., 2002, Esteller et al., 2001). Identification of differentially abundant proteins in B-CLL families offers an opportunity to identify candidate genes which are affected by epigenetic modification or variations in transcription factor binding sites. In addition,

recent advances in MS techniques provide further opportunities to identify differences in protein expression profiles between control B lymphocytes and B-CLL, and B-CLL subgroups.

4.2. Advances in Mass Spectrometry

Combining methods to separate proteins with mass spectrometers has minimised collision of ions during analysis and allowed detection of a greater number of peptides (Grebe and Singh, 2011). High-pressure liquid chromatography (HPLC) is used to resolve complex peptide mixtures and maximise peptide separation for mass spectrometry (Issaq et al., 2005). There are a number of different types of HPLC used, including reversed-phase liquid chromatography (RP-LC), normal phase liquid chromatography (NP-LC) and hydrophilic interaction liquid chromatography (HILIC) (Zhang et al., 2010) (Buszewski and Noga, 2012) (Yoshida, 2004). In comparison with other methods, HILIC retains very polar compounds and overcomes the poor solubility of hydrophilic compounds often observed in NP-LC. HILIC has also been shown to be compatible with MS coupled to LC (Buszewski and Noga, 2012). In summary, HILIC overcomes the limitations of both NP-LC and RP-LC techniques and was used as the separation protocol for the study reported in this Chapter.

The trapping capacity and efficiency of new generation ion traps has improved significantly over the last decade. The Orbitrap is a new type of mass spectrometer analyser developed in a hybrid device which provides high mass resolution and mass accuracy (Hu et al., 2005) (Zubarev and Makarov, 2013). For the present study, a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer was used which combines three mass analysers; a quadrupole mass filter, an ultra-high field Orbitrap mass analyser and dual-pressure linear ion trap analyser. These configurations offer high scan rate speed, multiple fragmentation techniques, and high mass accuracy and resolution to identify more low abundance proteins

(Senko et al., 2013).

In addition to peptide identification, quantitative differential proteomics can be used to calculate the relative amount of related protein from one sample to another or to measure the absolute amount of each protein. This quantification is performed by labelling samples with a stable isotope, and the quantity ratio of peptides is calculated and compared to other samples (Chen and Pramanik, 2009). Chemical or metabolic labelling has been most commonly used as labelling techniques for quantitative proteomics (Chen et al., 2015). Two isobaric tagging methods, iTRAQ and tandem mass tag (TMT), have been introduced to perform quantitative protein analysis and provide measurements of the relative abundances of proteins (Chahrour et al., 2015). TMT labelling is compatible with tandem MS and was used in the present study.

4.3. Bioinformatics Tools for Database Searching and Analysis

Two main approaches are used to perform peptide identification: 1. *de novo* sequencing and; 2. database search methods. In the first method, peptide sequencing is performed directly from the original spectra without using a sequence database, while the second approach uses a database dependent search. The database search algorithm is most commonly used and considered to be a more accurate method for peptide identification (Matthiesen, 2007). Several different algorithms are used to search sequence databases such as SEQUEST (Eng et al., 1994), Mascot (Perkins et al., 1999), XTandem (Craig and Beavis, 2004) and MS Amanda (Dorfer et al., 2014). All operate similarly by comparing experimental tandem mass spectra with theoretical spectra from the database, but they differ in their scoring systems (Tu et al., 2015) (Nesvizhskii, 2010). Other parameters have been recommended for improving the rate of confident peptide identifications. For example, the target-decoy false

positive rate search strategy is a simple method and one of the most common approaches to estimate the FDR. In this strategy, the experimental MS/MS spectra is searched against a concatenated database containing a target forward-sequences database and a match set of decoy sequences (reversed, or randomised, or shuffled) the same size as the target database. The top scoring peptide match for each spectrum is selected for further analysis and the application automatically counts the number of peptide-spectrum matches (PSM), and filter based on a given threshold to estimate the FDR (Elias and Gygi, 2007).

Further statistical improvements have achieved optimal separation between correct and incorrect PSMs. PeptideProphet (Keller et al., 2002) and the percolator (Kall et al., 2007) are the most common post-processing tools used to distinguish PSMs. The Percolator algorithm uses a semi-supervised machine learning method to iteratively train a support vector machine (SVM) classifier, which improves the discrimination between correct "target" and "decoy" spectrum identifications. The algorithm automatically calculates the q-value, SVM score and posterior error probabilities for each spectrum and assigns more reliable statistical confidence in peptide measurements (Kall et al., 2007).

In the study reported in this Chapter, bottom-up workflow and TMT labelling methods were used for identifying proteins differentially abundant between F-CLL and S-CLL and stable/progressive CLL samples (Chapter 6). In this workflow, B lymphocytes were enriched from all CLL-subgroups and controls, followed by protein denaturation. The experimental procedure for using the TMT10plex labelling approach is illustrated in (Figure 4-1).

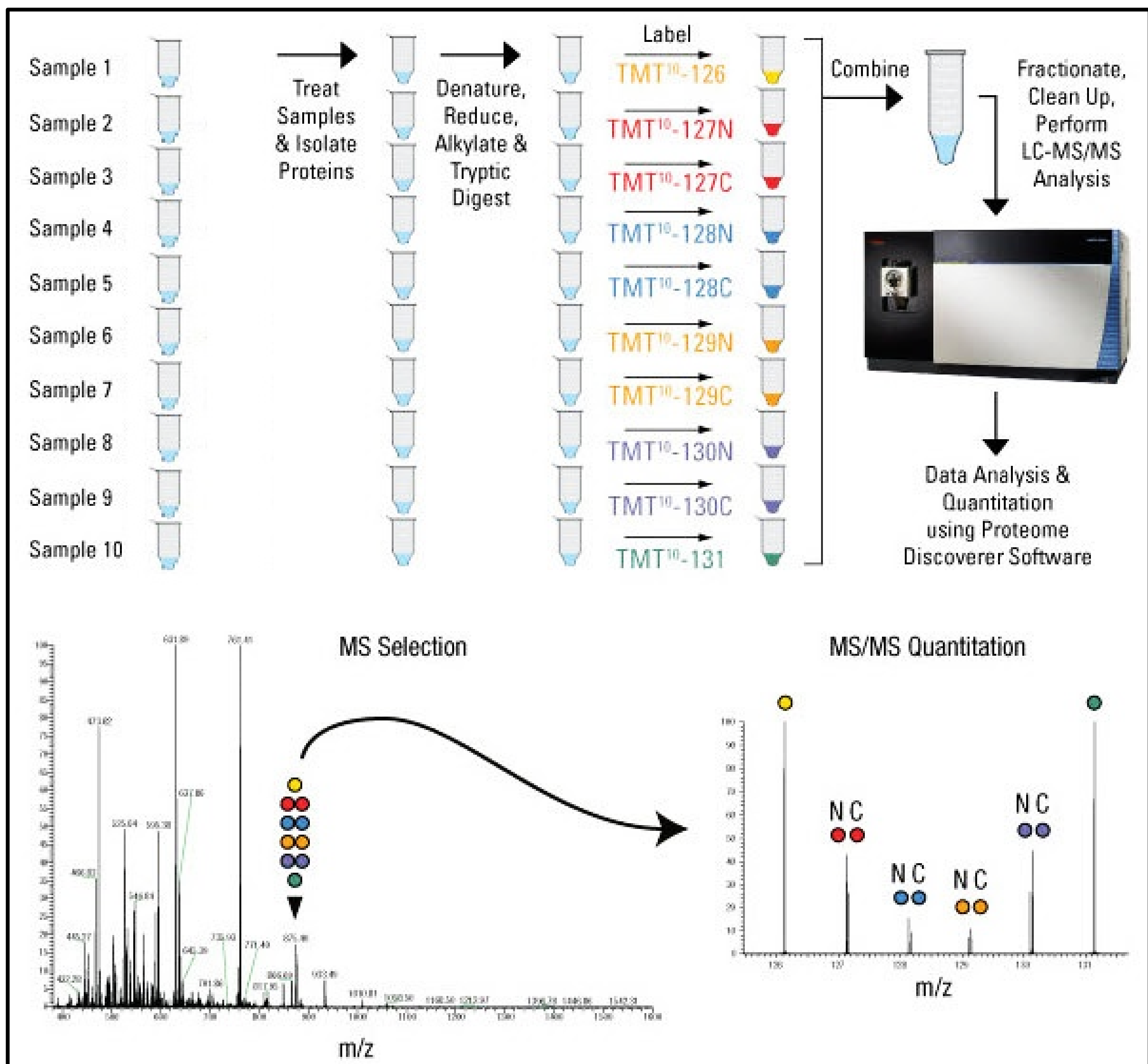

Figure 4-1. Experimental procedure for quantitative differential proteomics using TMT10plex labelling

The general workflow of the Thermo Scientific TMT10plex Isobaric Mass Tagging experiment. The protocol involves extracting proteins from cells, followed by reduction, alkylation, and digestion. Samples are labelled with TMT10plex reagents and the resulting TMT-labelled peptides are pooled at equal concentrations before fractionation and clean-up. The TMT-labelled samples are analysed by high-resolution Orbitrap LC-MS/MS. In the first MS scan, the labelled peptides are indistinguishable and appear as a single precursor. Following fragmentation of the precursor ion during MS/MS, the tag generates a unique reporter ion. The reporter ion intensity indicates the relative amount of the peptide in each sample.

HYPOTHESES

Identification of differentially abundant proteins in F-LPD and S-CLL will provide protein profiles that can be used as disease signatures and will identify proteins that may act as 'cancer drivers' for different subtypes of CLL.

AIMS

1. Determine if F-LPD B lymphocytes contain unique protein signatures compared to B lymphocytes from controls and S-CLL cases using a combination of quadrupole, ion trap and Orbitrap mass spectrometer analysis, and unsupervised hierarchical clustering.

2. Determine if F-LPD B lymphocytes contain unique protein signatures compared to B lymphocytes from controls and S-CLL cases using semi-supervised hierarchical clustering.

4.4. MATERIALS

Table 4-1 List of materials used for proteomics

Materials	Supplier
Tapered microtip sonicator (5 mm)	(Branson B-250 Sonicator, Danbury, CT)
4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES)	Sigma-Aldrich (St. Louis, MO, USA), Cat. No. H3375
Acetone	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. 270725
Dithiothreitol (DTT)	DTT (Dithiothreitol, Thermo Scientific Inc.)
Eppendorf concentrator 5301	(Eppendorf AG, Hamburg, Germany)
Iodoacetamide (IAA)	IAA (Iodoacetamide, Thermo Scientific Inc.)
Oasis hydrophilic-hydrophobic-balanced (HLB) plus short cartridges	(Waters, Milford, Massachusetts (MA), USA)
Qubit® 2.0 Fluorometer	(Life Technologies, Carlsbad, CA, USA)
Qubit® Protein Assay Kits	(Life Technologies, Carlsbad, California (CA), USA)
Sodium Dodecyl Sulfate (SDS)	Amresco (Ohio, USA) Cat. No. 97064-470
Thiourea	Sigma-Aldrich (St. Louis, MO, USA)
TMT 10plex Mass Tag labelling kit	(Thermo Scientific Inc.)
Triethylammonium bicarbonate	Sigma-Aldrich (St. Louis, MO, USA)
Trifluoroacetic Acid (TFA)	(Pierce TM Trifluoroacetic Acid, Thermo Scientific Inc.)
Tris (hydroxymethyl) aminomethane	Amresco (Ohio, USA), Cat. No. 97061-794
Trypsin	(Pierce Trypsin Protease, MS Grade, Thermo Scientific Inc.), Cat. No. 90057
TSK-Amide 80 3um HILIC column	(Tosoh Bioscience, Tokyo, Japan)
Urea	Sigma-Aldrich (St. Louis, MO, USA), Cat. No. U5378
V bottom 96 well plate	(Greiner, polypropylene)

Table 4-2. List of materials used for western blot

Materials	Supplier
Urea	Sigma-Aldrich (St. Louis, MO, USA), Cat. No. U5378
Tris (hydroxymethyl) aminomethane	Amresco (Ohio, USA), Cat. No. 97061-794
Acetone	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. 270725
Sodium Dodecyl Sulfate (SDS)	Sigma-Aldrich (St. Louis, MO, USA) Cat. No.L4390
Protease Inhibitor Cocktail (PI)	Roche Cat. No.0589279001
1,4-dithiothreitol (DTT	Thermo Scientific Cat. No. 20290
Acrylamide	Amresco Cat. No. 97064-542
NuPageLDS sample buffer (4x) buffer	Life Technologies Cat No. NP0007
NAP blocker	G-biosciences Cat No. 786-190
Novex sharp prestained protein standard	Invitrogen Cat No. LC5800
Polyvinylidene fluoride (PVDF) Membrane	Amersham Cat No 1060030
Methanol	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. 34860
Temed	Sigma-Aldrich (St. Louis, MO, USA) Cat. No.T9281
Ammonium persulfate (APS)	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. A3678
Gentle Review Stripping Buffer	VWR N552
p-Coumaric acid	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. C9008

Luminol	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. A8511
30% H_2O_2	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. 216763
Tricine	Sigma-Aldrich (St. Louis, MO, USA) Cat. No. T0337
Glycine	Amresco Cat. No. 97063-736

Table 4-3. List of reagents required for western blot

Buffers	Reagents
Urea lysis Buffer	8M Urea in 50 mM Tris-HCl; pH 7.5, 0.1% SDS
Tris-buffered saline with Tween 20 (TBST)	Tris-buffered saline (TBS); pH 7.5 plus 0.05 % Tween-20
Enhanced chemiluminiescence (ECL) detection buffer	1.25mM Luminol, 0.2 mM p-coumaric acid, 0.1 H2O2 in 0.1 M Tris pH 6.8
Running Buffer	50mM Tricine, 50mM Tris, 0.1% SDS,pH 8.24
Towbin Transfer Buffer	25mM Tris, 192mM Glycine, pH 8.3, 20 % Methanol
10% APS	10g APS dissolved in 100 ml MilliQ H2O
Blocking buffer	1:1 NAP Blocker : TBST
Primary antibody diluting buffer	1NAP Blocker : TBST

Table 4-4. Gel composition to make a 3–15% gradient gel

Reagents	3% stacking gel	8% tris-acetate gel	15% tris-acetate gel
15 x Tris acetate buffer	0.27 ml	0.67ml	0.4 ml
40% acrylamide	0.30 ml	2.0 ml	2.25 ml
Pure water	3.43	9.4	7.6
Total 40ml	4 ml	10 ml	6 ml
Temed	5 µL	12.5 µL	7.5 µL
10% APS	19 µL	37.5 µL	28.5 µL

Table 4-5. List of primary antibodies used for western blot

Antibodies	Company	Dilution
Mouse anti human CYBB	Santa Cruz Biotechnology, Cat. No. sc-130543	1:200
Mouse anti human GAPDH	Santa Cruz Biotechnology, Cat. No. sc-32233	1:500
Goat anti-mouse IgG-HRP	Santa Cruz Biotechnology, Cat. No. sc-2005	1:5000
Goat anti-mouse IgM-HRP	Thermo Fisher Scientific Cat. No. 62-6820	1:5000

4.5. METHOD

4.5.1. Negative Selection of B-CLL Cells from Blood Samples

Purified B-CLL cells were isolated and a minimum of $1x10^6$ enriched B cells were

washed 4 times with PBS (1 ml) by centrifuging for 5 min at 2000 g, to remove FCS proteins, as described in section 2.2.2. The washed pellet was homogenised by vigorous mixing in protein lysis buffer (200 µL) containing 8 M urea in 50mM Tris-HCL with 0.1% (w/v) SDS, at pH 7.5, before storing at -80°C.

4.5.2.Preparing Samples for Protein Profiles at the MSCF

All protein studies were performed by the candidate at the Mass Spectrometry Core Facility (MSCF), Charles Perkins Centre (CPC), the University of Sydney, with supervision by Ms. Angela Connolly. Enriched B cells, in dissolution buffer (8 M urea, 50 mM Tris-HCL with 0.1% (w/v) SDS, at pH 7.5), were thawed and resuspended in 200 µL lysis buffer (6 M urea, 2 M thiourea, 100 mM HEPES buffer, pH 7.5). The mixture was then tip-probe sonicated for 2 × 20 s using a 5 mm tapered microtip sonicator (Branson B-250 Sonicator, Danbury, CT) with 1 min on ice between each round. Samples were immersed in ice immediately before processing for acetone precipitation and clean-up.

4.5.3.Protein Precipitation and Clean-Up

Acetone precipitation was used to precipitate and concentrate proteins and remove excess salts, ionic detergents, lipids, and nucleic acids that may interfere with downstream studies. Samples were centrifuged at 16000 g for 5 min and the supernatant was collected in new Eppendorf tubes and precipitated with 1 mL ice-cold acetone (100%) before placing at -30°C overnight. Acetone was removed from each sample by centrifugation at 1000 g for 10 min at RT and the pellet was washed by adding ice-cold 80% acetone (1 mL). This was followed by centrifuging at 1000 g for 10 min at RT, the supernatant was carefully decanted and the protein pellet was dried for 5-10 min by air-drying. The protein pellet was resuspended in 100 µL lysis buffer (6 M urea, 2 M thiourea, 100 mM HEPES buffer, pH 7.5) and then 2 µL from

each sample was taken for protein quantitation using a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

4.5.4. Protein Assay

Protein concentrations were assessed using Qubit® Protein Assay Kits (Invitrogen, Life Technologies, Carlsbad, California (CA), USA) and Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, California (CA), USA). According to the manufacturer's instructions, Qubit working solution was prepared for the samples and 3 standards, using 199 µL Qubit buffer (Solution B) per 1 µL Qubit Reagent (Solution A) at a ratio of 1:200 for each sample and standard. For each standards tube, a total of 190 µL working solution was transferred into thin, clear 0.5 mL optical grade qRT-PCR tubes and 10 µL of each Qubit standard added to the appropriate tube for a final volume of 200 µL. The assay tubes were prepared in a final volume of 200 µL by loading 2 µL of extracted proteins into individual assay tubes containing 198 µL. Tubes were gently vortexed for 3 s and incubated at RT for 15 min. Concentration measurements were performed in duplicate using the Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, California (CA), USA) and protein concentrations for each sample calculated against standards and displayed in µg/mL, as per the manufacturer's instructions.

4.5.5. Protein Reduction and Alkylation

After determining the protein concentration, protein extracts in the homogenised buffer (6 M urea, 2 M thiourea, 100 mM HEPES buffer, pH 7.5) were reduced with 10 mM DTT (Dithiothreitol, Thermo Scientific Inc.) for 30 min at RT. After reduction, samples were alkylated with 25 mM IAA (Iodoacetamide, Thermo Scientific Inc.) in the dark for 30 min at RT. This step irreversibly prevents the free sulfhydryl groups on the cysteine residues from reforming disulphide bonds. The reaction was quenched with DDT to make a final

concentration of 20 mM. Samples were diluted (1 in 5) with 100 mM HEPES and proteins were digested with trypsin.

4.5.6. Protein Digestion

Proteolytic digestions were performed using trypsin (Pierce Trypsin Protease, MS Grade, Thermo Scientific Inc.) at a ratio of 1:20, trypsin to protein in 0.1 M TEAB, and incubated overnight at 30°C. Trypsin (0.1 µg/µL) was prepared by dissolving 20 µg in 200 µL 0.1 M tetraethylammonium bicarbonate (TEAB).

The digested samples were acidified with 1% TFA (v/v) (Pierce ™ Trifluoroacetic Acid, Thermo Scientific Inc.) per mL of sample to make a final concentration of 0.1% and then centrifuged at 16,000 g for 5 min to remove insoluble materials. After this step, the peptides were ready for desalting and concentrating through Oasis HLB plus short cartridges (Waters Corporation).

4.5.7. Peptides Desalting and Concentration

After proteolytic digestion, Oasis hydrophilic-hydrophobic-balanced (HLB) plus short cartridges (Waters, Milford, Massachusetts (MA), USA) were used to remove excess salts, detergents and buffers that significantly influence the ionisation efficiency process and the quality of mass spectrum analysis. Briefly, the Oasis hydrophilic-hydrophobic-balanced plus short cartridge was equilibrated with 100% methanol (1 mL), followed by 1 mL 100% acetonitrile (ACN) before washing the cartridge with 0.1% TFA (1 mL). Samples were loaded and the flow-through collected into a clear new Eppendorf tube. This step was repeated once to ensure maximum binding and the flow-through kept at -20C for recovering. The cartridge was washed with 0.1% TFA (5 mL). To elute peptides, 50% ACN (1 mL) in 0.1% TFA was slowly

loaded, and the peptides were eluted from the cartridge after the first drop and subsequently collected into a new 1.5 mL Eppendorf tube. The volume of each eluted peptide sample was reduced overnight to ≈20 µL by speed vacuum centrifugation using an Eppendorf concentrator 5301 (Eppendorf AG, Hamburg, Germany). The eluted peptides were resuspended in up to 100 µL with 100mM HEPES buffer pH 8.0, sonicated for 5 min, and pH 8.0 was checked for all samples. The eluted peptides were measured using the Qubit® Protein Assay Kits as described previously in 4.5.4. After quantification, the peptide samples with ≥ 100 µg were selected and reconstituted in 100 mM HEPES buffer (pH 8.0) for TMT tagging.

4.5.8.TMT10plex Mass Tag Labelling

In preparation for TMT labelling, 2 separate experiments were designed as follows: 1. to compare between F-CLL, S-CLL and controls; 2. to compare between M-CLL and UM-CLL. Three controls from a family with multiple cases of F-LPD were pooled in one sample, whereas a further 3 unrelated controls were pooled in another tube. For experiments 1 and 2, samples were labelled with TMT as shown in (Figure 4-2).

The labelling process using the TMT 10plex Mass Tag labelling (Thermo Scientific) was performed according to the manufacturer’s instructions. The TMT labelling reagents were prepared by adding anhydrous acetonitrile (41 µL) to each tube containing labelling reagent (0.8 mg). The reagent was allowed to dissolve for 5 min with occasional vortexing and brief spinning. TMT tag (20 µL) was added to each peptide sample (20 µg). The reaction was incubated at RT for one hour, followed by adding 5% hydroxylamine (8 µL) to each sample and then incubating for 15 min to quench the reaction. TMT-labelled samples for the 2 experiments (10plex TMT, experiment 1; 10plex TMT, experiment 2) were combined in 2 separate tubes (Figure 4-2). Both tubes were diluted with 0.1% TFA up to a final volume of 200

µL and purified using Oasis hydrophilic-hydrophobic-balanced (HLB) plus short cartridges (Waters, Milford, MA, USA) as described in 4.5.7, for HILIC separation.

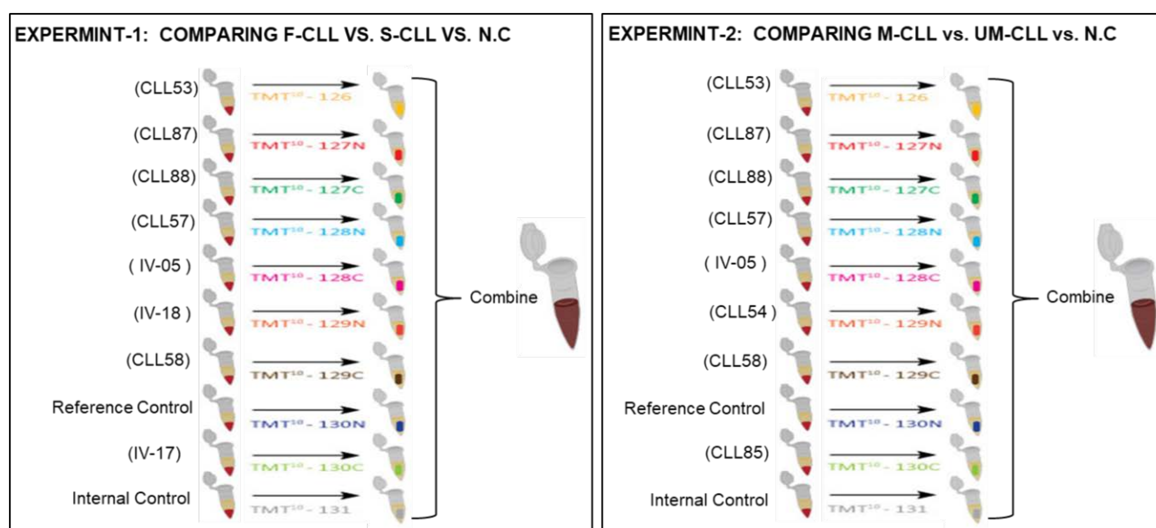

Figure 4-2 TMT10plex Mass Tag Labelling design for both experiments

Two separate experiments of TMT-labelled peptides were prepared and performed according to the manufacturer's recommendations. Each experimental set consisted of the reference control, internal control and 8 different samples. The reference control was a pool of unrelated controls, while the internal control was a pool of related family controls. These controls were assigned to the same reporter ion channels (130N and 131) in both experimental sets. (Left) Experiment-1 contained peptides from 2 controls, 3 F-CLL and 5 S-CLL samples labelled with a unique TMT reagent and then combined in equal amounts for analysis. (Right) Experiment-2 contained peptides from 2 controls, 3 UM-CLL and 5 M-CLL samples labelled with a unique TMT reagent and then combined in equal amounts for analysis.

4.5.9.Offline HILIC Separation:

Samples were fractionated by HILIC on offline mode using an Agilent 1200 chromatography system. Briefly, labelled peptides were resuspended in 100% HILIC buffer B (90% Acetonitrile, 10% Milli-Q water, 0.1% TFA) followed by HILIC fractionation using in-house packed TSK-Amide 80 HILIC columns (3 µm particle size, 320 µm inner diameter, 450 µm outer diameter, 17 cm length) (Tosoh Bioscience, Tokyo, Japan) with an attached PEEK filter (UpChurch Scientific, Thermo Scientific). Peptides were loaded onto the column in 100% HILIC buffer B for 15 min at 6 µL/min flow rate and eluted with a gradient of 100-60% HILIC buffer

B over 20 min at 6 µL/min, followed by a column re-equilibration step for 15 min. Fractions (8-10) were manually collected in a V bottom 96 well plate (Greiner Bio-One Gloucestershire, UK) at 2 min intervals after UV detection at 210 nm and the plate dried by vacuum centrifugation before LC-MS/MS analysis.

4.5.10. Liquid Chromatography Tandem Mass Spectrometry Analysis Using an Orbitrap Fusion Tribrid™ Mass Spectrometer

The TMT labelled HILIC fractions were resuspended in 6 µL of MS Loading buffer (3% acetonitrile with 0.1% formic acid) and analysed online by nano-capillary liquid chromatography tandem mass spectrometry (LC-MS/MS) using a Dionex Ultimate 3000 HPLC system and a Thermo Scientific Orbitrap Fusion Tribrid™ mass spectrometer. Peptides were loaded onto a Dionex Ultimate 3000 HPLC system (Thermo Scientific) and separated using an in-house packed 75 µm internal diameter capillary x 40 cm pulled column with 1.9 µm particle size, C18-AQ (Dr Maisch, Ammerbuch-Entringen, Germany). Two buffers were used: HPLC-grade 80% v/v Acetonitrile with 0.1% v/v formic acid (buffer B) and 0.1% v/v formic acid (buffer A). Peptides were eluted over a 150 min gradient at a flow rate of 250 nL/min. The gradient used for the analysis is presented below:

Time/ (min.)	**0**	**0**	**30**	**30**	**130**	**133**	**137**	**137**	**150**
% B	5	5	5	10	35	95	95	5	5

To identify and quantify TMT-labelled peptides, the Orbitrap Fusion Tribrid MS was programmed in a data-dependent mode for MS2 and multi-notch synchronous precursor selection MS3 scans and data acquired using Thermo Scientific Xcalibur software. To increase the number of identified peptides and protein groups in more complex mixtures, a full MS

(MS1) survey scan (at 375-1575 m/z) was performed in the Orbitrap Fusion Tribrid MS with a high resolving power of 120,000 FWHM, full-width-at-half-maximum. The significance of using high resolution was to separate two mass spectral peaks that have certain small mass differences and reduce the number of misidentifications (Strupat et al., 2016). The automatic gain control (AGC) target of 4 x 10^5 was enabled to regulate the number of ions in the mass analyser; and the maximum injection time for accumulation of the desired number of ions in the trapping device was set at 100 ms. To identify precursor ion masses for isolation and fragmentation at MS2, several parameters such as monoisotopic precursor selection, peptide were enabled, to include only peptide precursors with a charge state of 2-7, and intensity threshold above 5000 counts. The dynamic exclusion duration was set to 90 s, to minimize repeated sequencing of peptides and allow for new precursor ions to undergo fragmentation. These parameters improve protein and peptide identification rates (Kalli et al., 2013). The 12 most intense precursor ions in the MS1 survey scan were subjected to collision induced dissociation (CID) fragmentation. The MS2 scan was performed in the linear ion trap using the following settings: quadrupole isolation mode, CID activation type, AGC target 2 x 10^4, maximum injection time 70 ms, and rapid scan rate. The normalised collision energy was set to 35% and the activation (q) to 0.25. Following fragmentation, synchronous precursor selection (SPS) was used to select the 10 most abundant precursor ions in MS2 which were fragmented by HCD at a normalised collision energy of 55%. The MS3 scans were acquired in the Orbitrap Fusion Tribrid MS at a resolution of 60 000 FWHM with a 1 x 10^5 AGC target, and maximum injection time 120 ms.

4.5.11.Bioinformatic Tools for Database Searching and Analyses

The MS data.RAW files acquired by the Orbitrap Fusion Tribrid MS and Xcalibur data

system (Thermo Scientific) were directly imported into proteome discoverer version 2.1.0.81 (Thermo Scientific) and searched using SEQUEST HT with percolator validation. The database searches were performed using the following criteria: (i) UniProt Homo Sapiens with isoforms protein database (updated November – 2015.fasta) with common contaminants; (ii) enzyme name; trypsin; (iii) two maximum missed cleavage sites; (iv) dynamic modifications included: methionine (Met) Oxidation (+15.9949Da), and protein N-terminal acetylation (+42.01057); (v) carbamidomethylation of cysteine residues (57.02146Da) and TMT tags on peptide N-terminus and on lysine residues (229.16293Da) set as static modifications. The precursor mass tolerance was set to 20 ppm and fragment mass tolerance at 0.6 Da. FDR p-value were determined using the Percolator algorithm (version 2.05), and p-values were set to < 0.01 (1% FDR) and 0.05 (5% FDR) at both peptide and protein levels. Peptides < 7 amino acids in length were excluded and TMT reporter ions were quantified from the MS3 scan using an integration tolerance of 20 ppm with the most confident centroid. The parameter settings for peptide and protein quantifier node were set by applying a quantification value correction to true, co-isolation threshold to 50, and average reporter signal to noise (S/N) threshold value to 10. Proteome Discoverer calculates abundance ratios for each sample against the normal control (subject 130N).

Ingenuity pathway analysis (IPA) was used to determine cellular location, molecular functions, network signalling and associated pathways for identified proteins. The "Significant Proteins" datasets were uploaded separately into the IPA software using standard settings and criteria restricted to human, immune cells, mononuclear cells, lymphocytes, B-lymphocytes and peripheral blood lymphocyte. Core analysis was run to find the most significant interactions and associations in the datasets compared to the IPA database. Significance values were calculated using either a right-tailed Fisher's exact test p-value of <

0.05, or the ratio of the number of genes from the list of the dataset that are involved in the canonical pathway divided by the total numbers of genes in this pathway. To display only the most significant Canonical Pathways, the cut-off threshold was set by default at a negative log (p-value) greater than 1.3, which means that pathways with a p-value ≥ 0.05 are hidden. (IPA; ingenuity systems, http://www.ingenuity.com).

4.5.12. Statistical Analyses

Proteome Discoverer (Thermo Fisher Scientific, Waltham, MA, USA) was used to calculate abundance ratios for each sample against a normal control and protein quantification values were exported for further analysis to Microsoft Excel (2016). Assessment for the difference between normalised protein abundance between F-LPD and B-CLL patients were performed using independent two-sample t-tests. Proteins with log fold changes > 2 and t-test p-values < 0.05 were considered in further analyses. A total of 30 proteins were identified and a heat map using complete linkage and a Euclidean distance metric was constructed using the function "heatmap.2" from the gplots package in R (Warnes et al., 2016).

4.5.13. Western Blot validation of Mass Spectrometry

Because CYBB protein and mRNA were differentially expressed, CYBB was chosen to validate the MS data using western blotting. In brief, pellets of purified B cells were lysed in 1mL Urea lysis buffer + PI. Samples were vortexed for 5 s and left in ice for 60 s. This step was repeated 5 times before passing the lysate through a 23-gauge needle 3 times. The solution was left on ice for 20 min and centrifuged for 10 min at 4°C. After centrifugation, the supernatant was collected and precipitated with 4 volumes of ice-cold Acetone (100%), vortexed and stored at -20 overnight. Precipitated protein was pelleted by centrifugation at

16 000 g for 10 min at 4°C, and the pellet was washed 3 times with ice-cold 80% acetone, centrifuging as above. After the final wash, residual supernatant was completely removed and the pellet was air dried for 10 min at RT. The protein pellet was resuspended in 10 µL 1xLDS sample buffer + 100mM DTT and then heated at 70°C for 10 min with intermittent mixing before loading samples into the wells of an 8-15% Tris-acetate gel with a 3 % stacking gel. Nupage prestained protein standards were run alongside the samples. The proteins were separated by running the gel at 150 V for ~ 1hr in running buffer. The gel was soaked for 5 min in Towbin transfer buffer before assembling the blotting sandwich to transfer the separated proteins from the gel to PVDF membrane at 350mA for 1 hr at 4°C in Towbin transfer buffer. Subsequently, the membrane was rinsed with TBST and incubated for 1 hr in blocking buffer at room temperature with rocking. The primary antibody against CYBB was diluted as indicated in (Table 4-5) into primary antibody dilution buffer and the membrane was incubated with this mixture overnight at 4°C with rocking. The membrane was washed three times with TBST for 10 min with rocking at RT. After washing, the membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies: goat anti-mouse IgG-HRP and anti-IgM-HRP at 1:5000 dilution in TBST for 1 hr at RT with rocking. The membrane was again washed 3 times with TBST as previously. Enhanced chemiluminescence was used to develop the blot according to the method described by (Mruk and Cheng, 2011). Non saturating images were collected using a Gel Doc System (Bio-Rad). After collecting the images, antibodies were stripped from the blot by incubating in Gentle ReView stripping buffer for 30 min at RT with rocking. After washing as with TBST, the membrane was reblocked as before and probed with anti GAPDH and HRP secondary antibody as previously described. Assessment of the band intensity was performed using imageJ (https://imagej.nih.gov/ij/), according to the described by (http://lukemiller.org/index.php/2010/11/analyzing-gels-and-

western-blots-with-image-j/).

4.6. RESULTS

To determine if F-LPD has a unique protein profile in addition to a unique mRNA profile, purified B lymphocytes from 3 F-LPD cases were compared to S-CLL cases and normal subjects. One F-CLL patient (IV-2) had required treatment with fludarabine, cyclophosphamide and rituximab (FCR) between the times of collection of mRNA and protein samples, and this subject was not included in the proteomic analysis. In addition, adequate amounts of protein could not be extracted from small B lymphocyte clones in 2 F-MBL subjects (III-10 and IV-13). Consequently, the proteomic analysis was limited to 3 familial LPD cases (one F-CLL and 2 F-MBL), 3 family controls, 3 unrelated normal controls and 5 S-CLL cases.

Analysis of protein levels using a combination of quadrupole, ion trap and Orbitrap MS analysis identified 4672 proteins that after normalisation to the 6 control samples, segregated F-LPD, S-CLL and healthy control groups using unsupervised hierarchical clustering based on protein level patterns (Figure 4-3). Concordant with mRNA hierarchical clustering reported in 3.4.4, F-CLL and F-MBL cases did not segregate: one F-MBL case segregated with the F-CLL case before the second F-MBL case (Figure 4-3). Consequently, F-CLL and F-MBL cases were combined as “F-LPD” for subsequent comparisons of protein levels.

For the F-LPD group, the highest levels for upregulated proteins were for ADP ribosylation factor interacting protein 2 (ARFIP2) (log 2-fold change = 5.8; p-value = 0.02), C14orf2 (log 2-fold change = 5.2; p-value = 0.008), and macrophage migration inhibitory factor (MIF) (log 2-fold change = 3.9; p-value = 0.003) (Table 4-6). Whereas for downregulated

proteins, the lowest expression levels were for HLA-B (log 2-fold change = -3.5; p-value = 7.2E-06), arylsulfatase family member I (ARSI) (log 2-fold change = -3.0; p-value = 0.01) and zinc finger protein 648 (ZNF648) (log 2-fold change = -2.9; p-value = 0.02).

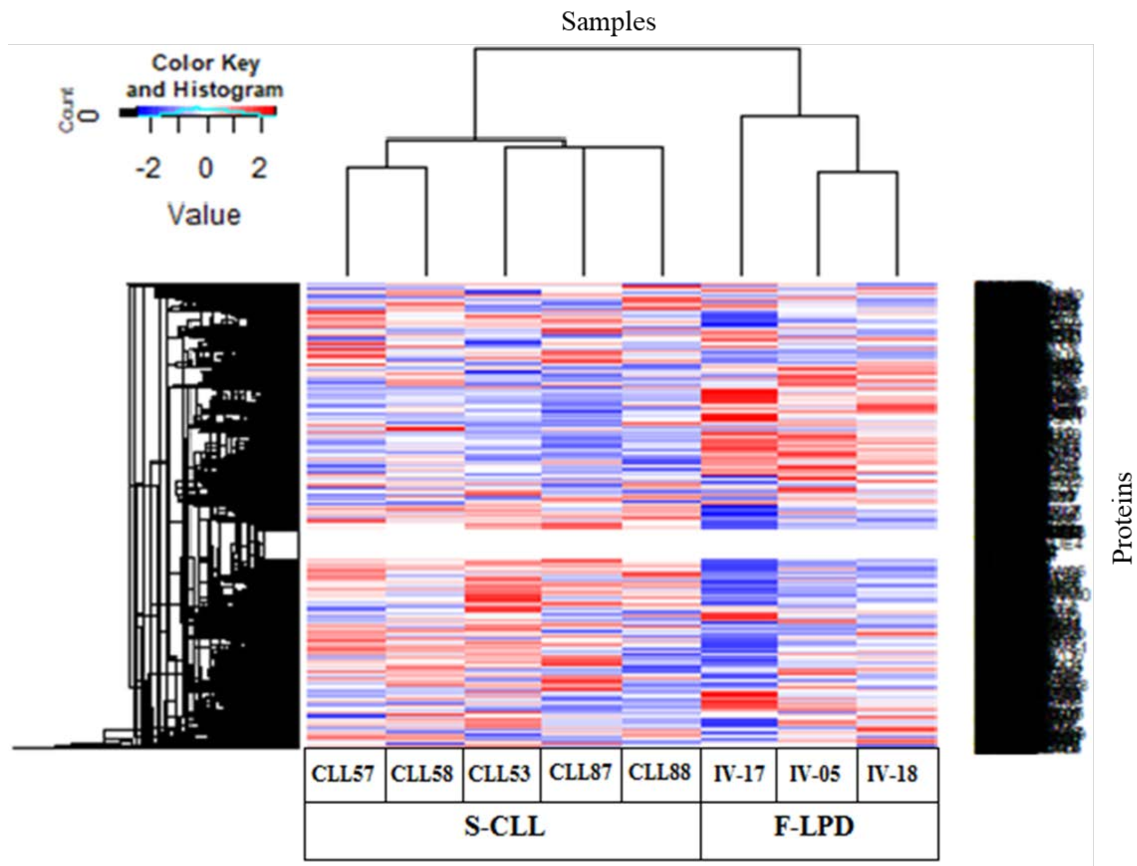

Figure 4-3 Unsupervised hierarchical cluster image of protein expression in S-CLL and F-LPD (combined F-MBL and F-CLL)

Differential abundance analysis of grouped data for F-LPD and S-CLL proteins was used to select proteins for further analysis by semi-supervised hierarchical clustering. A 2-fold difference in upregulation or downregulation of protein levels was chosen as biologically relevant and significance was set at $p < 0.05$ (Eagle et al., 2015). Following the first data analysis, HLA class II histocompatibility antigen, DRB1-13 beta chain and GRB2-associated-binding protein 2 showed no difference between groups and were removed from further

analyses. The results of log 2-fold change plotted against -log10 p value are shown in a volcano plot (Figure 4-4), which identified 30 proteins to be used for individual expression analysis (Table 4-6). The highest significance for upregulation was for HACD (log 2-fold change = 2.6; p < 0.0001) and MIF (log 2-fold change = 3.9; p=0.003), and for downregulation, SERPINH1 (log 2-fold change = -2.603; p=0.0003).

Figure 4-4 Volcano plot of protein expression for F-LPD versus S-CLL

The volcano plot shows the results of log 2-fold change plotted against -log10 p value. The green diamond symbols represent proteins that were differentially expressed between F-LPD and S-CLL.

Table 4-6. Proteins differentially expressed between F-LPD vs S-CLL with > 2-fold change and p-value < 0.05

Symbol	Entrez Gene Name	Log 2-fold change	p-value *	Location	Type(s)
ARFIP2	ADP ribosylation factor interacting protein 2	5.8	0.02	Cytoplasm	other
C14orf2	chromosome 14 open reading frame 2	5.2	0.008	Cytoplasm	other
MIF	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	3.9	0.003	Extracellular Space	cytokine
CYBB	cytochrome b-245 beta chain	3.2	0.04	Cytoplasm	enzyme
ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	3.0	0.02	Plasma Membrane	enzyme
MAPK3	mitogen-activated protein kinase 3	2.9	0.004	Cytoplasm	kinase
ACTBL2	actin, beta like 2	2.9	0.006	Nucleus	other
HIST1H1C	histone cluster 1 H1 family member c	2.7	0.03	Nucleus	other
TGFBRAP1	transforming growth factor beta receptor associated protein 1	2.7	0.009	Cytoplasm	other
HLA-DRB4	major histocompatibility complex, class II, DR beta 4	2.7	0.04	Plasma Membrane	transmembrane receptor
HACD3	3-hydroxyacyl-CoA dehydratase 3	2.6	<0.0001	Cytoplasm	enzyme
HIST1H1D	histone cluster 1 H1 family member d	2.6	0.04	Nucleus	other
CD74	CD74 molecule	2.6	0.007	Plasma Membrane	transmembrane receptor
RPS19	ribosomal protein S19	2.4	0.005	Cytoplasm	other
RPL14	ribosomal protein L14	2.4	0.03	Cytoplasm	other
FIP1L1	factor interacting with PAPOLA and CPSF1	2.3	0.007	Nucleus	other
SRSF2	serine and arginine rich splicing factor 2	2.3	0.04	Nucleus	transcription regulator
ADAMTS16	ADAM metallopeptidase with thrombospondin type 1 motif 16	2.1	0.02	Extracellular Space	other
NIFK	nucleolar protein interacting with the FHA domain of MKI67	2.1	0.04	Nucleus	other
EIF4EBP2	eukaryotic translation initiation factor 4E binding protein 2	2.0	0.03	Cytoplasm	translation regulator
TDRD1	tudor domain containing 1	2.0	0.03	Cytoplasm	other
LYAR	Ly1 antibody reactive	2.0	0.01	Plasma Membrane	other
PEA15	phosphoprotein enriched in astrocytes 15	-2.2	0.05	Cytoplasm	transporter
MTOR	mechanistic target of rapamycin	-2.3	0.009	Nucleus	kinase
HNRNPD	heterogeneous nuclear ribonucleoprotein D	-2.5	0.005	Nucleus	transcription regulator
SERPINH1	serpin family H member 1	-2.6	0.0003	Extracellular Space	other
ZNF292	zinc finger protein 292	-2.7	0.03	Nucleus	transcription regulator
ZNF648	zinc finger protein 648	-2.9	0.02	Other	other
ARSI	arylsulfatase family member I	-3.0	0.01	Extracellular Space	enzyme
HLA-B	major histocompatibility complex, class I, B	-3.5	7.2E-06	Plasma Membrane	transmembrane receptor

* t-test p-values

Semi-supervised hierarchical clustering was performed in F-LPD (n=3) and S-CLL (n=5) cases using expression data for the 30 proteins differentially expressed between F-LPD and S-CLL (Figure 4-5). The cluster dendrogram segregated F-LPD and S-CLL cases (Figure 4-5). Hierarchical clustering showed earlier segregation of one F-CLL case (IV-05) with an F-MBL case (IV-18), rather than segregation of the two F-MBL cases (Figure 4-5).

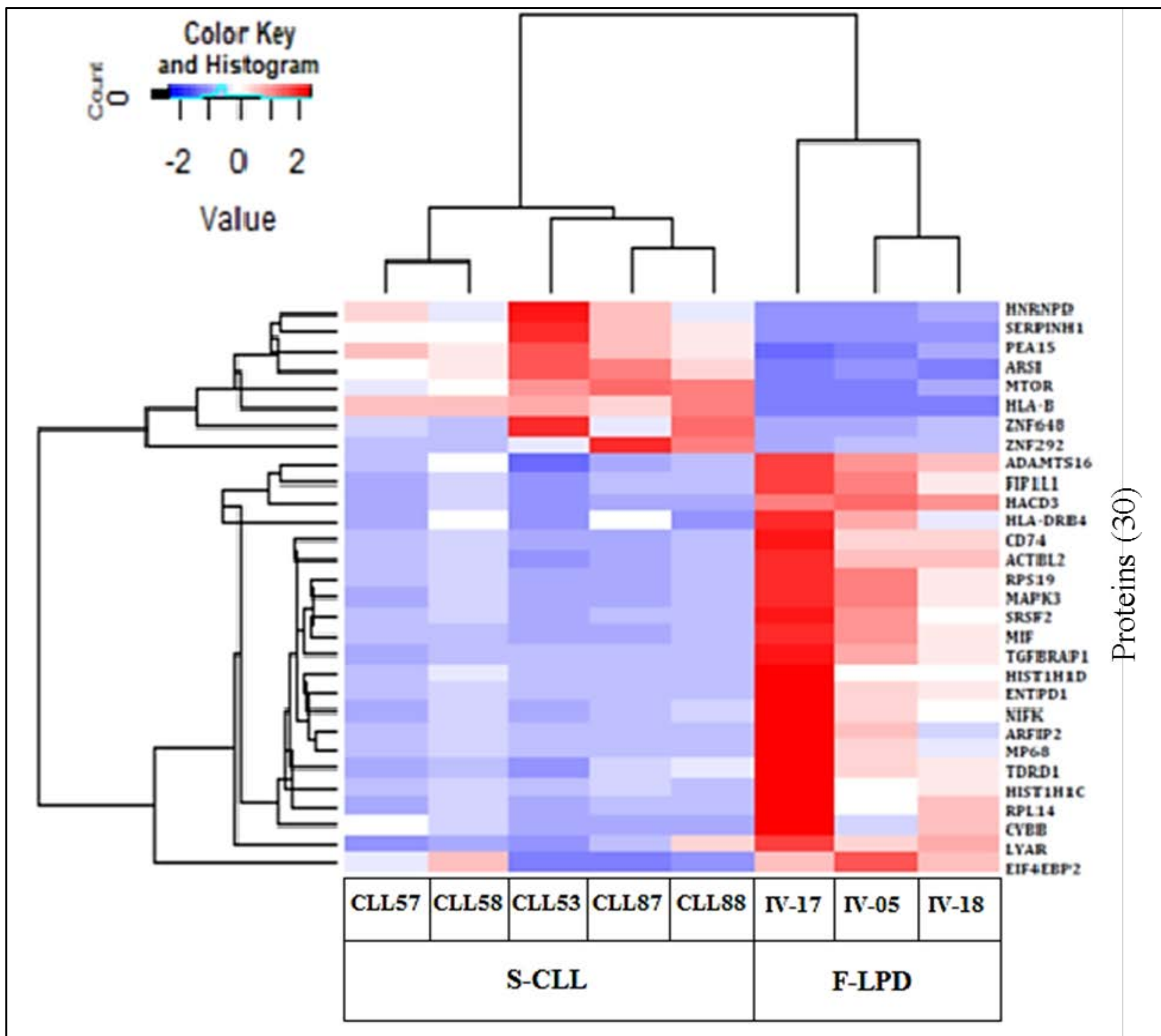

Figure 4-5 Clustering analysis of proteins differentially expressed between F-LPD and S-CLL

For the 30 differentially expressed proteins, correlations with mRNA levels were studied using individual gene expression data. Three proteins (MIF, SERPINH1 and CYBB) showed differential gene expression between familial controls, F-LPD and S-CLL. However,

only *CYBB* mRNA levels showed a significant difference, being higher in familial controls versus F-LPD and higher compared to S-CLL (FDR < 0.05). For F-LPD compared to S-CLL, CYBB protein levels were also higher (log 2-fold change = 3.222; p=0.04). Because CYBB protein and mRNA were differentially expressed, CYBB was chosen to validate the MS data using western blotting. Western blotting of CYBB showed higher expression in individual F-MBL (IV-17) compared to S-CLL57 and normalized to GAPDH. (Figure 4-6). These data validated the quantitative proteomics data.

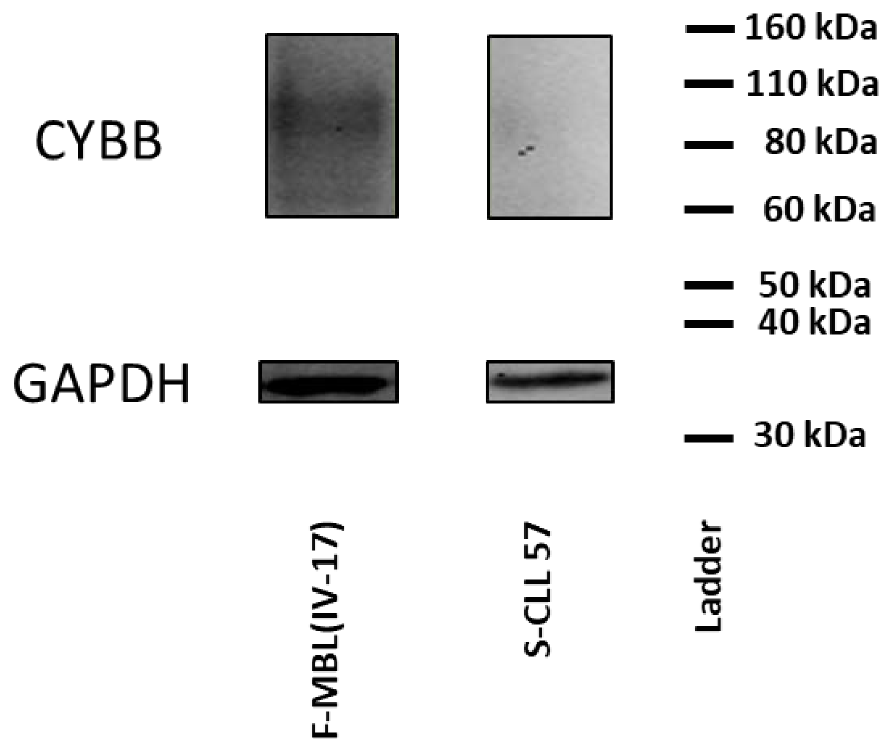

Figure 4-6. Expression of CYBB protein in F-MBL (IV-17) compared to S-CLL57

Proteins were separated by 8-15% Tris-acetate gel followed by western blotting and antibody detection using specific antibody to CYBB and an antibody to GAPDH. The western blot shows an increase in the expression of CYBB protein in F-MBL (IV-17) compared to S-CLL57.

Ingenuity pathway analysis (IPA) was used to determine cellular location, molecular functions, network signalling and associated pathways for identified proteins. A summary of the IPA results are presented in (Table 4-7 Integrated pathway analysis of proteins

differentially abundant between F-LPD & S-CLL. The top 5 canonical pathways include MIF-mediated glucocorticoid regulation (p-value=2.68E-05; 3/23 molecules), MIF regulation of innate immunity (p-value=5.48E-05; 3/29 molecules), regulation of eIF4 and p70S6K signalling (p-value=2.72E-04; 4/124 molecules), Rac signaling (p-value=1.44E-03; 3/87 molecules), antigen presentation pathway (p-value=2.91E-03; 2/31 molecules). Molecular and cellular functions differing between these two groups include cell-to-cell signalling and interaction, cellular development, and cellular growth and proliferation. This comparison identified only one significant network had a score of 2 with 1 focus molecule (CD74) linked to antigen presentation, inflammatory response, cellular assembly and organization.

IPA analysis indicated that 11 proteins were localized to the cytoplasm, 10 to the nucleus, 5 to the plasma membrane, and 4 to the extracellular space. All identified proteins had a range of different functions including cytokine activity (MIF), kinase activity (MAPK3 and mTOR), transmembrane receptor (HLA-DRB4, CD74, and HLA-B) and regulation of transcription (SRSF2, HNRNPD and ZNF292). Three molecules (HLA-DRB4, CD74 and MIF) were predicted to be involved in B-cell proliferation and peripheral T lymphocyte responses (Figure 4-7).

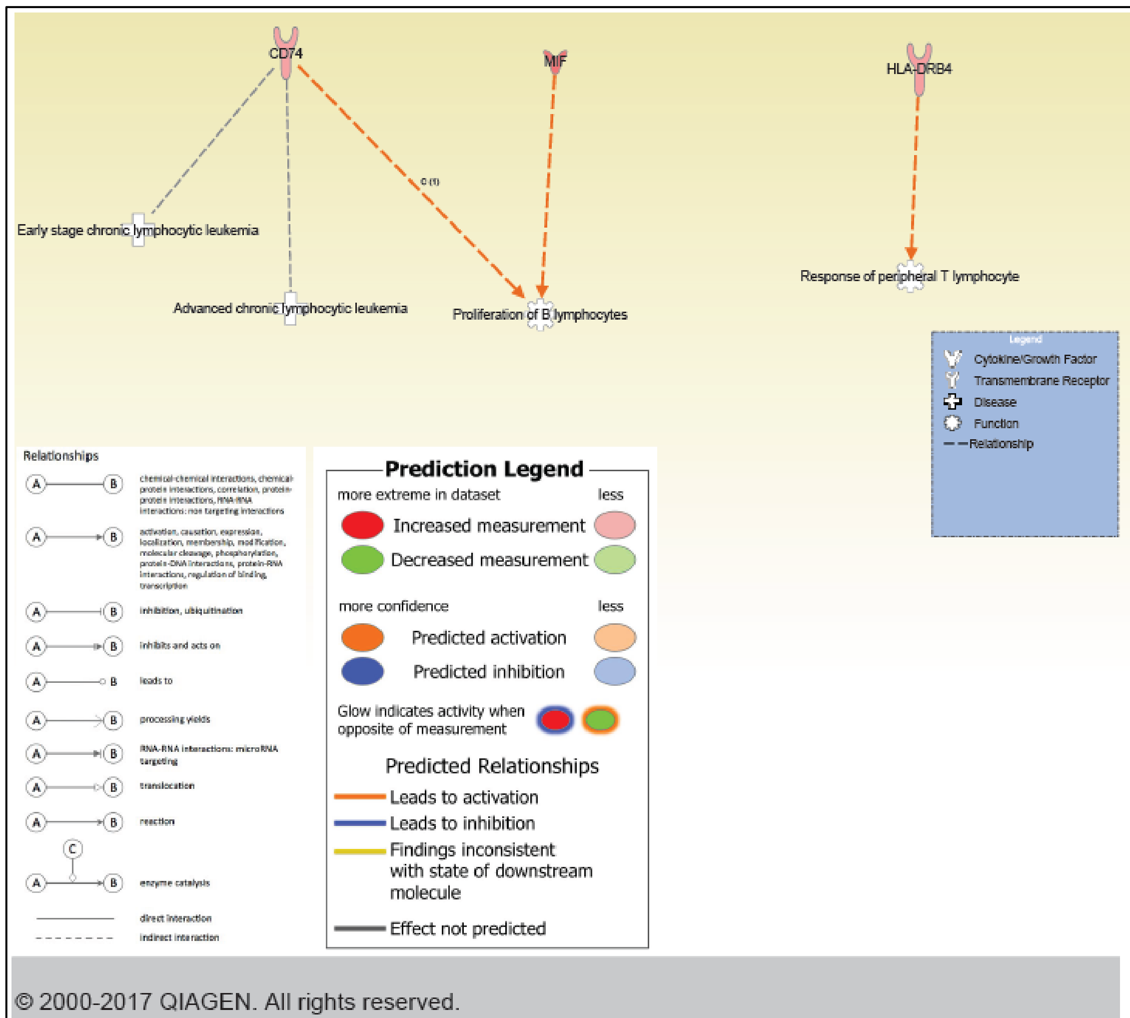

Figure 4-7 Proteins implicated in early and advanced CLL and B-cell proliferation.

Ingenuity pathway analysis identified HLA-DRB4, CD74 and MIF as being implicated in early and advanced CLL and B-cell proliferation. Overexpression of CD74 and MIF are predicted to be indirectly involved in activation and proliferation of B-lymphocytes and HLA-DRB4 in peripheral T lymphocyte response (orange dashed lines). Grey lines indicate over-expression of CD74 is indirectly involved in early and advanced B-CLL, however the biological effect of CD74 on early and advanced CLL cannot be predicted.

Table 4-7 Integrated pathway analysis of proteins differentially abundant between F-LPD & S-CLL

Top Canonical Pathways	**p-value**	**Overlap ***	**Ratio ****
MIF-mediated Glucocorticoid Regulation	2.68E-05	3/23	0.13
MIF Regulation of Innate Immunity	5.48E-05	3/29	0.103
Regulation of eIF4 and p70S6K Signalling	2.72E-04	4/124	0.0323
Rac Signalling	1.44E-03	3/87	0.0345
Antigen Presentation Pathway	2.91E-03	2/31	0.0645
Diseases and Disorders	**p-value**	**#Molecules**	
Cancer	2.62E-03	1	
Hematological Disease	2.62E-03	1	
Immunological Disease	2.62E-03	1	
Organismal Injury and Abnormalities	2.62E-03	1	
Molecular and Cellular Functions	**p-value**	**#Molecules**	
Cell-To-Cell Signalling and Interaction	7.84E-03	1	
Cellular Development	1.09E-02	2	
Cellular Growth and Proliferation	1.09E-02	2	
Physiological System Development and Function	**p-value**	**#Molecules**	
Hematological System Development and Function	7.84E-03	3	
Humoral Immune Response	1.09E-02	2	
Lymphoid Tissue Structure and Development	1.09E-02	2	
Top Networks		**Score**	
1. Antigen Presentation, Inflammatory Response, Cellular Assembly and Organization		2	

©2000-2017 QIAGEN. All rights reserved

* Overlap column indicates the number of observed molecules from our dataset that met the filter criteria and participate in a canonical pathway to the total number of molecules that participate in the same canonical pathway from the Ingenuity knowledge base.

** The Ratio column is calculated by taking the number of observed molecules from our dataset that participate in a canonical pathway divided by the total number of molecules that participate in the same canonical pathway from the Ingenuity knowledge base.

4.7. DISCUSSION

The protein expression profiles of B-CLL have been studied using a number of different methods (Alsagaby et al., 2014, Huang et al., 2016, Eagle et al., 2015, Perrot et al., 2011). The majority of these studies have identified protein markers which predict prognosis (Alsagaby et al., 2014, Eagle et al., 2015, Huang et al., 2016, Perrot et al., 2011), with only a limited number studying mechanisms of B lymphocyte neoplastic transformation (Alsagaby et al., 2014, Diez et al., 2017, Singh et al., 2015), and no studies comparing F-CLL/F-LPD with S-CLL. The aim of this study was to compare protein expression profiles between F-LPD and S-CLL. It was proposed that a study of a family with multiple CLL and MBL cases would provide some degree of control for genetic background and enhance the power to detect novel changes in protein expression associated with CLL development.

Of the 4672 identified proteins, 30 were differentially abundant between F-LPD and S-CLL. The highest significance for upregulation was for HACD and MIF, and for downregulation, SERPINH1. MIF protein binds to the B lymphocyte surface receptor CD74 (Binsky et al., 2010), which regulates VLA-4 expression, involved with homing and survival of CLL cells (Binsky et al., 2010).Three other proteins, ENTPD1/CD39, MTOR and SRSF2, have been associated with CLL biology (Abousamra et al., 2015, Decker et al., 2003, Garza et al., 2016, Pulte et al., 2007a, Pulte et al., 2007b, Tamburini et al., 2008). Of significance, increased levels of CYBB/gp91-phox protein correlated with levels of mRNA. Expression of *CYBB*/gp91phox gene has been described in B lymphocytes and in association with monocytic acute myeloid leukemia (Suzuki et al., 1998, Aurelius et al., 2012), although the present study is the first to identify increased expression in B-CLL cells.

Of the 30 proteins, 22 were expressed at higher levels and 8 expressed at lower levels

in F-LPD. These proteins were localized to the cytoplasm (11), nucleus (10), plasma membrane (5), and extracellular space (4). Cytoplasmic proteins are essential to maintaining cell structure and are involved in cellular activities and intracellular signalling, while nuclear proteins are responsible for cellular regulation, including regulation of gene expression (Thurgood et al., 2017).

Extracellular space proteins mediate extracellular signals and initiate intracellular signals. The 4 extracellular proteins identified in this study included MIF, ADAMTS16, SERPINH1 and ARSI. The pro-inflammatory cytokine, MIF, plays a critical role in immune regulation and inflammation and has been implicated in the pathogenesis of inflammatory and autoimmune diseases including rheumatoid arthritis, systemic lupus erythematosus (SLE) and atherosclerosis (Santos and Morand, 2009). In addition, MIF is overexpressed in skin, brain, breast, colon, prostate, and lung cancers and correlates with tumour aggressiveness and metastatic potential (Ouertatani-Sakouhi et al., 2010, Bernhagen et al., 2007, Lue et al., 2007, Meyer-Siegler et al., 2007, Rendon et al., 2009). MIF binding to CD74 induces sustained activation of ERK1/2 MAPK, in addition to transient activation via a Src-type kinase, which is co-regulated by the cellular MIF binding protein JAB1/CSN5 (Lue et al., 2006). Both pathways, which involve extra- and possibly intracellular MIF, regulate a number of cell functions including gene expression, proliferation and apoptosis (Calandra and Roger, 2003).

In CLL, MIF has been reported to promote the survival of B-CLL cells through CD74 signalling pathways (Binsky et al., 2010). MIF binds to the cell surface receptor CD74, the extracellularly expressed form of the MHC class II invariant chain (Leng et al., 2003). This interaction induces a pathway leading to the activation of NF-kB resulting in increased production of survival factor IL-8 (Binsky et al., 2010). Secreted IL-8 then further activates Bcl-2, inducing resistance to apoptosis in B-CLL (Binsky et al., 2007). In addition, binding of CD74

by MIF leads to upregulation of TAp63, a tumour suppressor protein and member of the p53 family (Su et al., 2010). Expression of Tap63 upregulates CLL expression of the integrin VLA-4 (ITGA4), promoting migration of circulating CLL cells to the bone marrow (Binsky et al., 2010). Importantly, blocking of MIF, CD74 or IL-8 decreases Bcl-2 expression and induces apoptosis (Binsky et al., 2010), suggesting that these molecules may be candidates for targeted therapies. In a murine *Eμ-TCL1* transgenic mouse model of CLL (Bichi et al., 2002), MIF protein is expressed at higher levels in B cells in leukaemic Eμ-TCL1 mice compared to TCL1 wild-type (Reinart et al., 2013). Knockout of MIF in these mice delayed the development of CLL, reduced splenomegaly and hepatomegaly, and prolonged survival (Reinart et al., 2013).

In the present study, several cytoplasmic proteins were identified to be differentially abundant in F-LPD including ADP ribosylation factor interacting protein 2 (ARFIP2), chromosome 14 open reading frame 2 (C14orf2), cytochrome b-245 beta chain (CYBB), mitogen-activated protein kinase 3 (MAPK3), transforming growth factor beta receptor associated protein 1 (TGFBRAP1), 3-hydroxyacyl-CoA dehydratase 3 (HACD3), ribosomal protein S19 (RPS19), ribosomal protein L14 (RPL14), eukaryotic translation initiation factor 4E binding protein 2 (EIF4EBP2), tudor domain containing 1 (TDRD1), and phosphoprotein enriched in astrocytes 15 (PEA15).

MAP kinases, or extracellular signal-regulated kinases (ERKs), act in a signalling cascade that regulates a number of cellular processes including proliferation, differentiation, and cell cycle progression in response to extracellular signals (Chang and Karin, 2001). The 2 MAPKs that play an important role in the MAPK/ERK signalling cascade are MAPK1/ERK2 and MAPK3/ERK1. As a result of activation, cytoplasmic MAPK3 regulates the activity of several transcription factors including FOS, Myc and signal transducer activation of transcription 3 (STAT3) (Cargnello and Roux, 2011). High levels of MAPK-Erk1/2 pathway activation have been

found to be associated with leukaemic cell survival, including B-CLL cells (Crassini et al., 2013), and the ERK pathway is activated after BCR ligation in B-CLL cells (Kawauchi et al., 2002). The activation of ERK in response to BCR stimulation has been reported to be associated with the expression of the proto-oncoprotein Myc, which is an essential positive regulator of cell cycle progression and cell growth (Krysov et al., 2012). Overexpression of MAPK3 in F-LPD B cells in the present study is consistent with previous studies which show deregulation of MAPK signalling plays a role in the pathogenesis of CLL (Shukla et al., 2017).

Members of the transforming growth factor-beta (TGF-ß) family, act through type II membrane receptor serine-threonine kinases, including TGFBR2, which leads to transphosphorylation of type I receptor serine-threonine kinases, including TGFBR1 (Wurthner et al., 2001). Downstream signalling events include embryogenesis, wound healing, tissue homeostasis, fibrosis, and immunomodulation (Derynck et al., 1998, Kulkarni et al., 1993). Activated TGFBR1 phosphorylates SMAD2/SMAD3, which binds to a common mediator, SMAD4, to form a SMAD complex that translocates to the nucleus to regulate transcription of target genes (Wurthner et al., 2001). Furthermore, TGF-ß can activate other signalling cascades including Erk, JNK, and p38 MAPK kinase pathways to initiate transcriptional responses independently of SMAD activation (Derynck and Zhang, 2003).

Dysregulation of TGF-ß, either decreased or increased but altered signaling, has been reported in association with progression and metastasis of malignancies (Pasche, 2001, Tang et al., 1998, Kyrtsonis et al., 1998). Expression levels of the TGF-ß receptor are decreased on B-CLL cells, or are less sensitive to the growth-inhibitory effects of TGF-ß, compared to normal B-lymphocytes (Douglas et al., 1997) (Lagneaux et al., 1997). In addition, TGF-ß is released at higher levels in B-CLL patients than controls (Lagneaux et al., 1995).

The plasma membrane proteins identified in the present study were ENTPD1, HLA-DRB4, CD74, LYAR, and HLA-B. These proteins are involved in a number of cellular processes including cell-cell interactions, transport, and signalling functions. The role of HLA class I proteins is to present peptides from endogenous proteins to CD8+ cytotoxic T-lymphocytes, while HLA class II proteins present peptides derived from exogenous proteins to CD4+ helper T-cells (Gragert et al., 2014). CLL risk has been associated with several HLA alleles including HLA-DQA1 and HLA-DRB5 (Slager et al., 2011), and HLA-DRB1 has been associated with F-CLL (Theodorou et al., 2002). Our results identified novel upregulation of HLA-DRB4 in F-LPD compared to S-CLL. Upregulation of HLA-DRB4 has been demonstrated to be associated with CLL susceptibility (Machulla et al., 2001) (Gragert et al., 2014), but not previous associated with F-CLL. HLA-B class I has also been found to be associated with increased risk of CLL in a US white population, but more work is needed to confirm this association (Gragert et al., 2014).

Our study found an increase in six, and decrease in four, nuclear proteins in F-LPD compared to S-CLL. These included ACTBL2, HIST1H1C, HIST1H1D, FIP1L1, SRSF2, NIFK, mTOR, HNRNPD, ZNF292, and ZNF648. Mechanistic (or mammalian) target of rapamycin (mTOR) is a protein kinase that belongs to the phosphatidylinositol 3 kinase (PI3K) family and has been implicated in the regulation of various cellular processes including cell cycle progression and cellular proliferation (Thoreen et al., 2012, Laplante and Sabatini, 2012). There are at least 2 multi-protein complexes in which mTOR is found, mTORC1 and mTORC2, which are defined by partner proteins, substrate specificities and differential sensitivity to rapamycin (Chapuis et al., 2010). The mTORC1 pathway is activated either by extracellular growth factors or changes in cell metabolism (Sarbassov et al., 2005),whereas mTORC2 is usually rapamycin-insensitive and has a role in cell survival and proliferation through activation of the

serine/threonine kinase Akt (Jacinto et al., 2004).

The activity of mTOR is deregulated in a number of haematological malignancies (Chapuis et al., 2010). In B-CLL, mTOR is regulated by PI3K/Akt signalling cascades which are involved in cell survival and proliferation (Blunt et al., 2015), and targeting mTOR in B-CLL induces cell cycle arrest and apoptosis (Decker et al., 2003) (Blunt et al., 2015). However, more studies are required to assess the safety and efficacy of therapies that target mTOR (Roohi and Hojjat-Farsangi, 2017).

Increased abundance levels of histone cluster 1, H1c (HIST1H1C) and histone cluster 1, H1d (HIST1H1D) were found in F-LPD compared to S-CLL cases. Histones are an integral part of chromatin structure and are responsible for the stability of DNA. The four core histones (H2A, H2B, H3 and H4) are essential for the formation of the histone octamer, while the linker histone, H1 and its subtypes, connect the nucleosome core particles at DNA entry and exit sites (Happel and Doenecke, 2009). The abundance of numerous histone post-translational modifications and histone variants has been widely studied in CLL (Glibert et al., 2014, Singh et al., 2015, Diez et al., 2017). HIST1H1C (or H1.2) and HIST1H1D (H1.3) are linker histones and it has been reported that the HIST1H1C is translocated from the nuclease to the cytoplasm in response to drug treatments in primary B-CLL cells (Gine et al., 2008) (Harshman et al., 2013). In addition, HIST1H1C initiates apoptotic cascades after exposure to X-ray irradiation and following DNA double-strand breaks in a p53-dependent manner (Gine et al., 2004) and (Gine et al., 2008).

SUMMARY

In this study, MS was used to identify differentially abundant proteins in B lymphocytes from a family with multiple cases of CLL and MBL. For this family, CLL appeared to be transmitted as an autosomal dominant disorder. However, a previous genetic linkage scan had provided no significant evidence for a single gene model of disease susceptibility (Fuller et al., 2008). As with the mRNA study reported in Chapter 3, the advantage of this proteomic association study was that background genetic variation was partially controlled as a function of the degree of relationship between family subjects, potentially increasing the contribution of epigenetic and/or environmental modifiers to variation in protein expression profiles (Borecki and Province, 2008).

Analysis of protein levels using a combination of quadrupole, ion trap and Orbitrap MS analysis identified 4672 proteins that after normalisation to control samples, segregated F-LPD, S-CLL and healthy control groups using unsupervised hierarchical clustering based on protein level patterns. Assessment for the difference between normalised protein abundance between F-LPD and B-CLL patients was performed using independent two-sample t-tests. Proteins with log fold changes > 2 and t-test p-values < 0.05 were considered in further analyses. A total of 30 proteins were identified and a heat map using complete linkage and a Euclidean distance metric was constructed (Warnes et al., 2016). Of these proteins, 22 were overexpressed, and 8 proteins were underexpressed in F-LPD compared to S-CLL. A number of proteins were identified that are involved in the regulation of a number of cellular processes including cell cycle progression and cellular proliferation. These proteins were localized to the cytoplasm, nucleus, plasma membrane, and extracellular space. All of the identified proteins have different molecular functions including cytokine activity (MIF), kinase activity (MAPK3

and mTOR), transmembrane receptors (HLA-DRB4, CD74, and HLA-B) and transcription regulator (SRSF2, HNRNPD and ZNF292). Using Ingenuity Pathway Analysis (IPA) analysis, three molecules (HLA-DRB4, CD74 and MIF) were implicated in early and advanced CLL and B-cell proliferation (Figure 4-7). Over expression of CD74 and MIF are predicted to be indirectly involved in activation and proliferation of B-lymphocytes and HLA-DRB4 in peripheral T lymphocyte response. However, the mechanism by which CD74 is related to B-lymphocyte activation and proliferation cannot be predicted using the IPA knowledgebase.

CHAPTER-5: CHANGES IN MESSENGER RNA PROFILES ASSOCIATED WITH PROGRESSION FROM NORMAL B LYMPHOCYTES TO F-MBL AND F-CLL

5.1. INTRODUCTION

Chapter 3 reported studies which found mRNA profiles differentiated B cells from control subjects, F-LPD and S-CLL. This Chapter will investigate whether changes in mRNA profiles are associated with progression from normal B-lymphocytes through pre-malignant MBL cells to malignant CLL.

The presence of CLL and MBL in several family members allowed comparison of mRNA profiles associated with progression to CLL that were to some extent controlled for a common genetic background. In this family, CLL appeared to be transmitted as an autosomal dominant disorder. However, a genetic linkage scan had provided no evidence for a single gene model of disease susceptibility (Fuller et al., 2008). In the study reported in this Chapter, background genetic variation was partially controlled as a function of the degree of relationship between family subjects, potentially increasing the contribution of epigenetic and/or environmental modifiers to variations in gene expression profiles (Borecki and Province, 2008).

Increases in the frequency of diagnostic blood testing have led to the recognition that MBL is a clinical precursor to CLL (Landgren et al., 2009, Parikh et al., 2013, Rawstron, 2002). MBL cases have the same immunophenotypic markers as CLL, expressing CD5, CD19 and CD23, with low levels of CD20 and surface immunoglobulin, and either kappa or lambda light chains (Boehler et al., 2011). Patients are considered to have MBL if they have $< 5 \times 10^9$/L monoclonal B cells in the peripheral blood, with no symptoms of fever, weight loss or night sweats, and no sign of lymphadenopathy, hepatomegaly, splenomegaly or cytopenias (Parikh et al., 2013).

MBL precedes almost all cases of CLL/ SLL (Landgren et al., 2009), and is subclassified into “low-count” MBL, defined as a peripheral blood monoclonal population $< 0.5 \times 10^9$/L, and “high-count” MBL (Swerdlow et al., 2016). Progression to CLL is very rare in low-count MBL,

whereas about 1-2% high-MBL cases progress annually (Vardi et al., 2013).

Individuals with a family history of CLL or other LPD have a 2 - 3 fold increased relative risk of MBL (Parikh et al., 2013), and MBL is reported in 13-18% of first degree relatives of F-CLL patients compared to 3-5% in the general population (Marti et al., 2003, Rawstron, 2002, Rawstron et al., 2002). These findings suggest that MBL is an early marker of an inherited predisposition to CLL that may be associated with DNA variants or non-DNA variations, including epigenetic modifications. In this present study, it was proposed that variations in gene expression are acting early in neoplastic transformation of B lymphocytes, resulting in an MBL, and subsequent somatic events further alter gene expression before overt CLL develops. The aim of this Chapter was to investigate whether changes in mRNA levels were associated with progression from normal B-lymphocytes through pre-malignant MBL cells to malignant CLL.

HYPOTHESIS

Genes associated with the development of CLL can be identified by studying the progression of normal B lymphocytes through pre-malignant MBL cells to malignant CLL in a family with multiple affected members, considering that the development of disease phenotypes in families is partially controlled for genetic factors.

AIM

- To identify changes in mRNA levels associated with progression of normal B lymphocytes through pre-malignant MBL cells to malignant CLL.

5.2. MATERIALS

Table 5-1 List of materials used for Affymetrix GeneChip Human Transcriptome Array 2.0

	MATERIALS	COMPANY
1	Affymetrix GeneChip Human Transcriptome Array 2.0	Affymetrix Inc, Santa Clara, California, USA
2	RNA Pico 6000 chip	Agilent Technologies
3	Agilent Bioanalyser	Agilent Technologies
4	GeneChip® WT Pico Kit	Affymetrix Inc, Santa Clara, California, USA, P/N 703262 Rev.5
5	GeneChip® Whole Transcript (WT) Expression Arrays	Affymetrix Inc, Santa Clara, California, USA
6	GeneChip® Expression wash, stain and scan for Cartridge arrays kit	Affymetrix Inc, Santa Clara, California, USA
7	Gene chip scanner 3000 7G	Affymetrix Inc, Santa Clara, California, USA
8	Affymetrix Expression Console software, version 1.4	Affymetrix Inc, Santa Clara, California, USA
9	Transcriptome analysis console (TAC) 3.0 software	Affymetrix Inc, Santa Clara, California, USA

5.3. METHOD

Samples (n = 9) were collected from surviving family cases (two F-CLL, and 4 F-MBL), and unaffected family controls (n = 3) from a family with multiple cases of B-LPD (Figure 3-1). B lymphocytes were enriched using a RosetteSep™ B-Cell isolation cocktail (StemCell Technologies Inc., Vancouver, BC, Canada) as described in 2.2. Samples were amplified using the GeneChip® WT Pico Kit (Affymetrix Inc, Santa Clara, California, USA, P/N 703262 Rev.5) and processed with the Affymetrix GeneChip Human Transcriptome Array 2.0 (Affymetrix Inc, Santa Clara, California, USA) as described earlier in 3.3.5.

The expression data (.CHP format) were loaded into the Affymetrix Transcriptome analysis console (TAC 3.0) software and samples were classified into 3 groups. There were 2 F-CLL (IV-02 & IV-05), 4 F-MBL (III-10, IV-13, IV-17 & IV-18) cases and 3 family controls (IV-16, V-16, and IV-15).

5.3.1. Statistical Analysis

Expression analysis was performed using Affymetrix Transcriptome analysis console (TAC 3.0) software. One-way (between subjects) ANOVA (unpaired) was selected and the TAC analysis was performed by comparing F-CLL, F-MBL and family controls. The assessment of differences between controls, F-MBL and F-CLL was performed similarly to the analysis for F-LPD and S-CLL cases in Chapter 3. Firstly, unsupervised hierarchical clustering was performed using genes with FDR p-values < 0.05. These genes were then compared between (A) controls and MBL and (B) MBL and CLL using 2-sample t-tests, and genes with a p-value < 0.01 and log fold change > 2 were used for semi-supervised hierarchical clustering.

5.4. RESULTS

After performing TAC analysis, 1926 differentially expressed mRNAs, consisting of 1372 coding and 554 noncoding transcripts, were identified that were differentially abundant between subjects. To control for multiple testing, an FDR p-value < 0.05 was set, and 6 genes were identified that, after unsupervised hierarchical clustering, segregated normal familial controls from F-MBL and F-CLL (Figure 5-1; Table 5-2). Three of these genes, *GRASP*, *LEF1*, and *ROR1*, had been identified as differentially expressed between F-LPD and S-CLL (Table 3-4). With progression from normal familial controls through F-MBL to F-CLL, *GRASP* was downregulated, while *LEF1*, *C11orf80*, *ROR1*, *METTL8*, and *PARP3* were upregulated (Table 5-2).

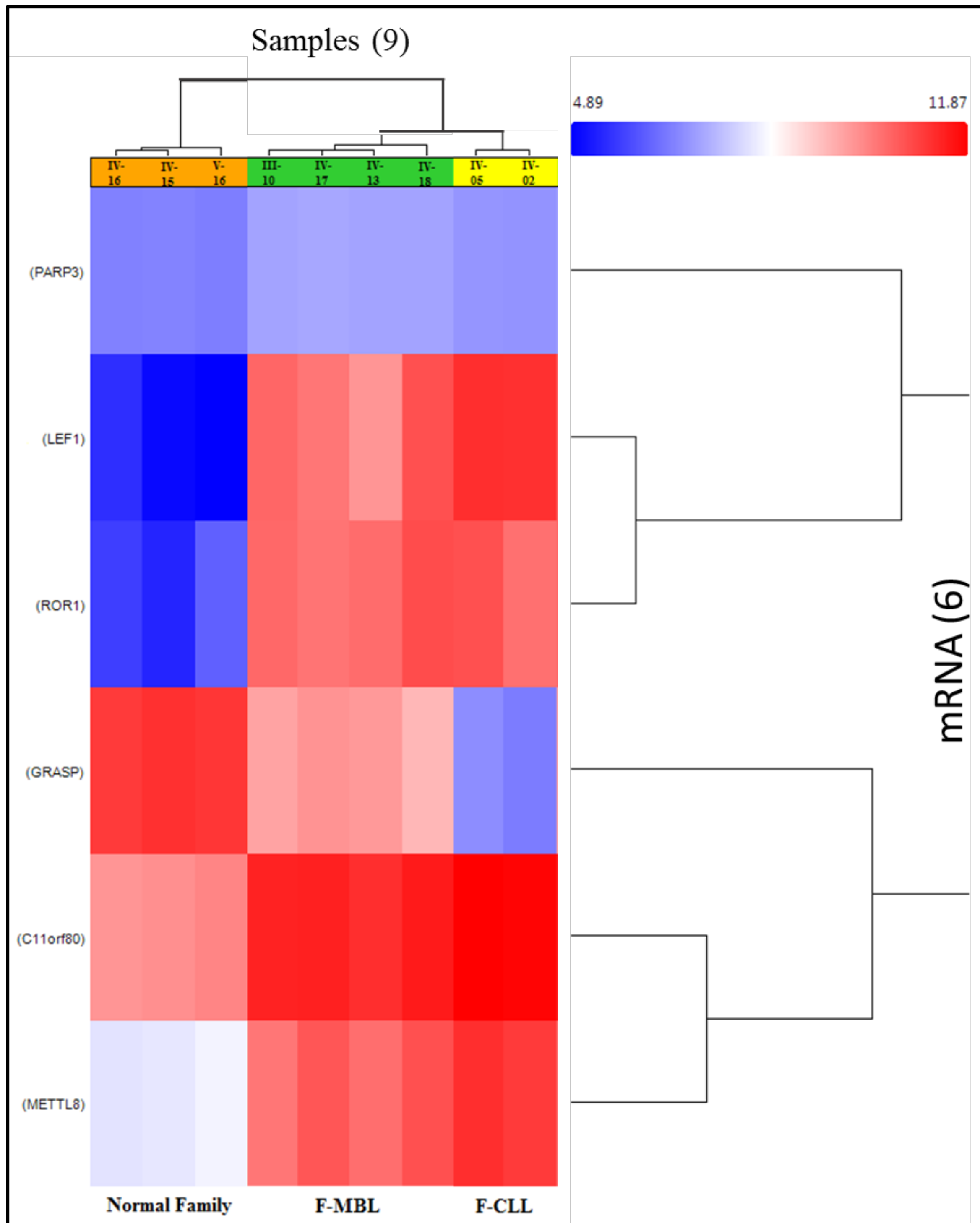

Figure 5-1 Genes differentially expressed using unsupervised hierarchical clustering in control family subjects, F-MBL and F-CLL cases

Table 5-2 Genes differentially expressed using unsupervised hierarchical clustering in control family subjects, F-MBL and F-CLL cases

Gene Symbol	Control family signal intensity*	F-MBL signal intensity	F-CLL signal intensity	FDR p-value
GRASP	11.1	9.7	6.7	**0.03**
LEF1	5.0	10.4	11.2	**0.03**
C11orf80	9.9	11.4	11.8	**0.03**
ROR1	5.8	10.4	10.6	**0.03**
METTL8	8.0	10.5	11.2	**0.04**
PARP3	6.7	7.1	6.9	**0.04**

*Signal intensity is taken as an average over the pairs of perfect match (PM) and mismatch (MM) probe spots. The statistical method used to summarize over the PM/MM pairs was the Tukey's Bi-weight average algorithm implemented in Affymetrix software. (Bolstad et al., 2003)

On the assumption that the expression of some genes becomes increasingly dysregulated with neoplastic progression, mRNAs that were differentially abundant between F-CLL and F-MBL were identified, and then compared between F-MBL and family controls. Normal controls, F-MBL and F-CLL were grouped and mean levels for mRNAs in each group were compared for significance using a Student's t-test. A 2-fold difference in up- or down-regulation of gene expression was chosen as biologically relevant and significance was set at $p < 0.01$ (Eagle et al., 2015).

Six genes (*SMAD3*, *DFNB31*, *TBC1D10C*, *INPP5F*, *GRASP*, and *RASGEF1B*) were differentially expressed for control versus F-MBL, and for comparison of F-MBL versus F-CLL (Table 5-3; Figure 5-2). Semi-supervised hierarchical clustering for each patient was then performed (Figure 5-3). With progression from MBL to F-CLL, *INPP5F* and *DFNB31/WHRN* were progressively upregulated; *SMAD3*, *GRASP* and *RASGEF1B* were progressively downregulated; and *TBC1D10C* was downregulated in F-MBL before being upregulated in F-CLL.

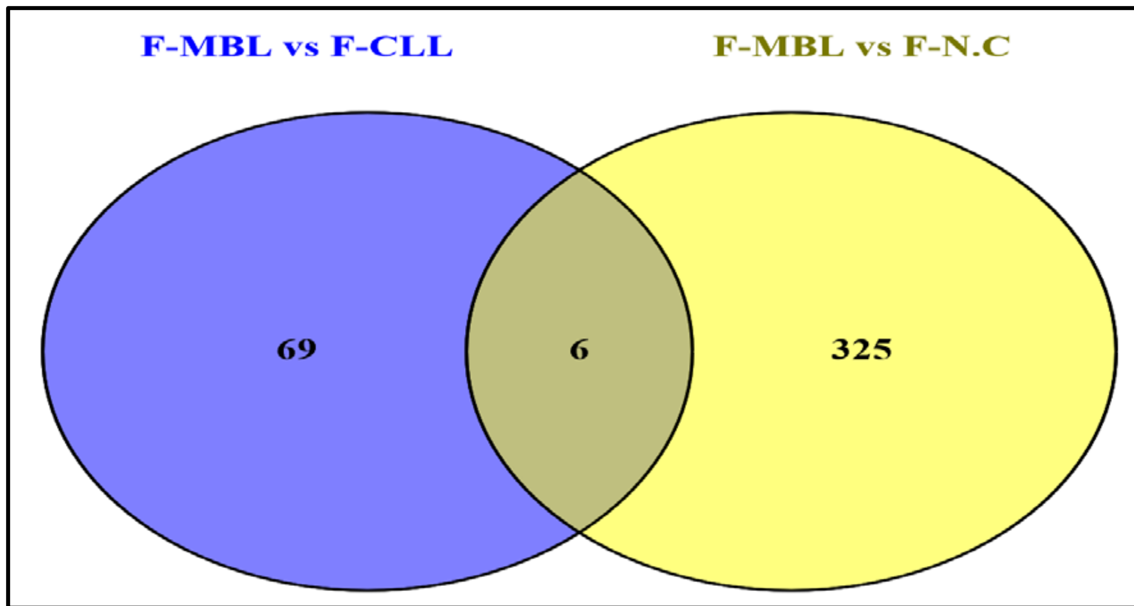

Figure 5-2. Venn diagram illustrating overlap between F-MBL vs F-CLL, and F-MBL vs Controls

Six genes were differentially abundant with t-test p < 0.01; >2-fold change in F-MBL versus F-CLL, and also differentially abundant in F-MBL versus family controls. F-CLL; familial chronic lymphocytic leukaemia, F-MBL; monoclonal B-cell lymphocytosis (F-MBL) family members, F-N.C; family controls. Retrieved from http://bioinfogp.cnb.csic.es/tools/venny/index.html on March 7th 2018.

Table 5-3 Genes differentially expressed between F-MBL and F-CLL cases that were also differentially expressed for control versus F-MBL

Gene symbol	Gene Name	Log 2-Fold Change*	p-value**	Cellular location	Function
GRASP	*general receptor for phosphoinositides 1 associated scaffold protein*	8.0	<0.0001	Plasma Membrane	Other
RASGEF1B	*RasGEF domain family member 1B*	5.3	0.005	Other	Other
SMAD3	*SMAD family member 3*	2.2	0.009	Nucleus	transcription regulator
INPP5F	*inositol polyphosphate-5-phosphatase F*	-2.4	0.007	Plasma Membrane	phosphatase
TBC1D10C	*TBC1 domain family member 10C*	-2.5	0.003	Nucleus	Other
WHRN/DFNB31	*whirlin*	-2.8	0.007	Plasma Membrane	Other

* Log2-fold change in expression between F-MBL and F-CLL

** Student's t-test.

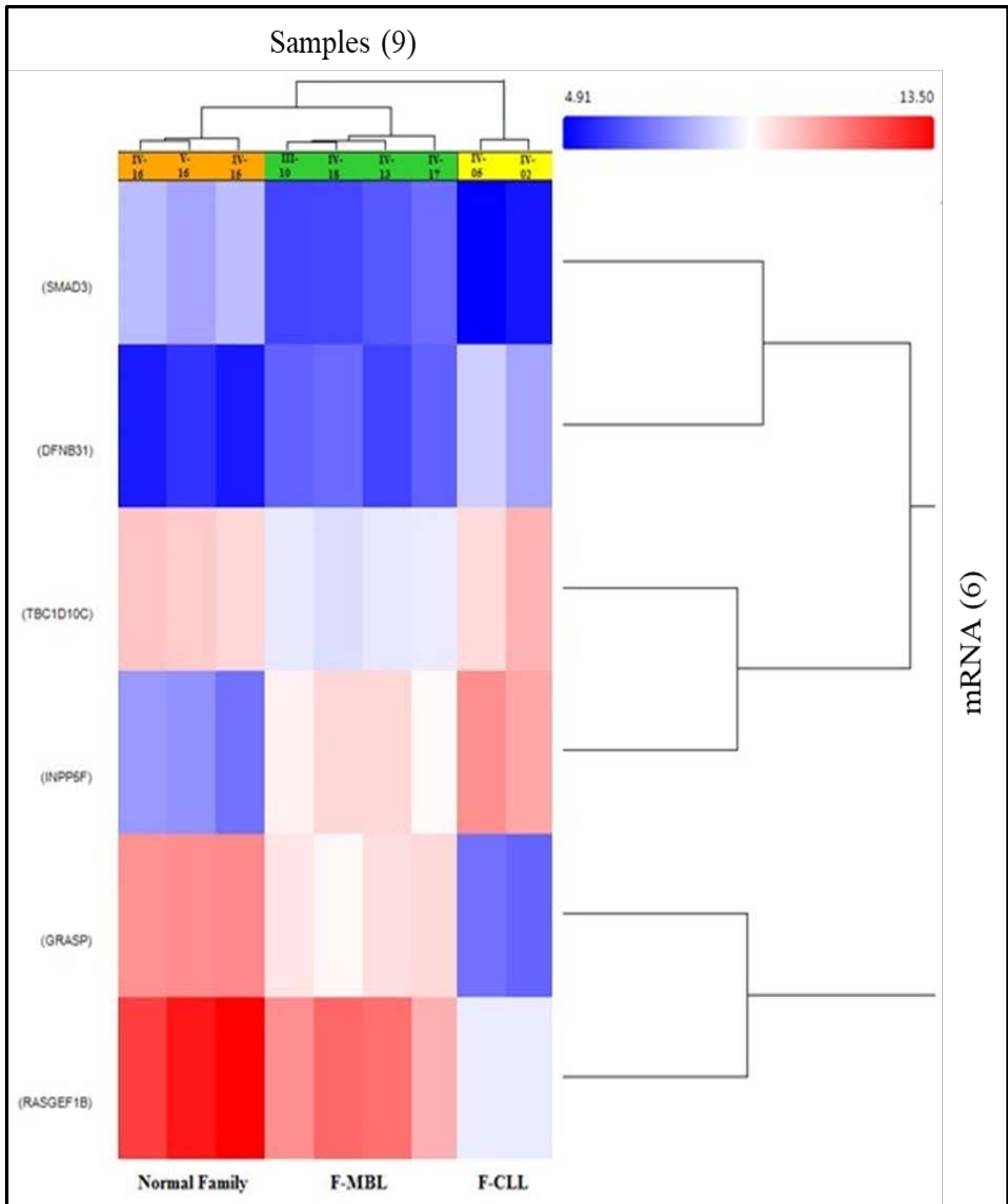

Figure 5-3. Genes differentially expressed by semi-supervised hierarchical clustering of B lymphocyte mRNA levels with progression from normal B family subjects, premalignant F-MBL to malignant F-CLL cases

5.4.1. Validation of Gene Microarrays Using qRT-PCR

Quantitative RT-PCR (qRT-PCR) was used to validate changes in gene expression associated with progression from normal B-lymphocytes to pre-malignant MBL to malignant CLL cells. For qRT-PCR, a Rotor-Gene 2000 (Corbett Research, Sydney, Australia) was used as described previously in section 2.7

Using semi-supervised clustering, 6 genes were identified as differentially associated with progression from normal B-lymphocytes to pre-malignant MBL to malignant CLL cells. qRT-PCR analyses were performed in 3 normal controls, 4 F-MBLs and 2 F-CLL cases and changes in expression were determined relative to *GAPDH* (delta C_t). The qRT-PCR results for these genes are shown in (**Error! Reference source not found.**).

The results showed that the expression levels of 4 genes (*GRASP*, *INPP5F*, *RASGEF1B,* and *SMAD3)* were significantly decreased from normal to premalignant MBL and F-CLL subjects, while *DFNB31/WHRN* was significantly increased. There were no differences in *TBC1D10C* levels between the groups. These results correlated with expression levels in the microarray analysis.

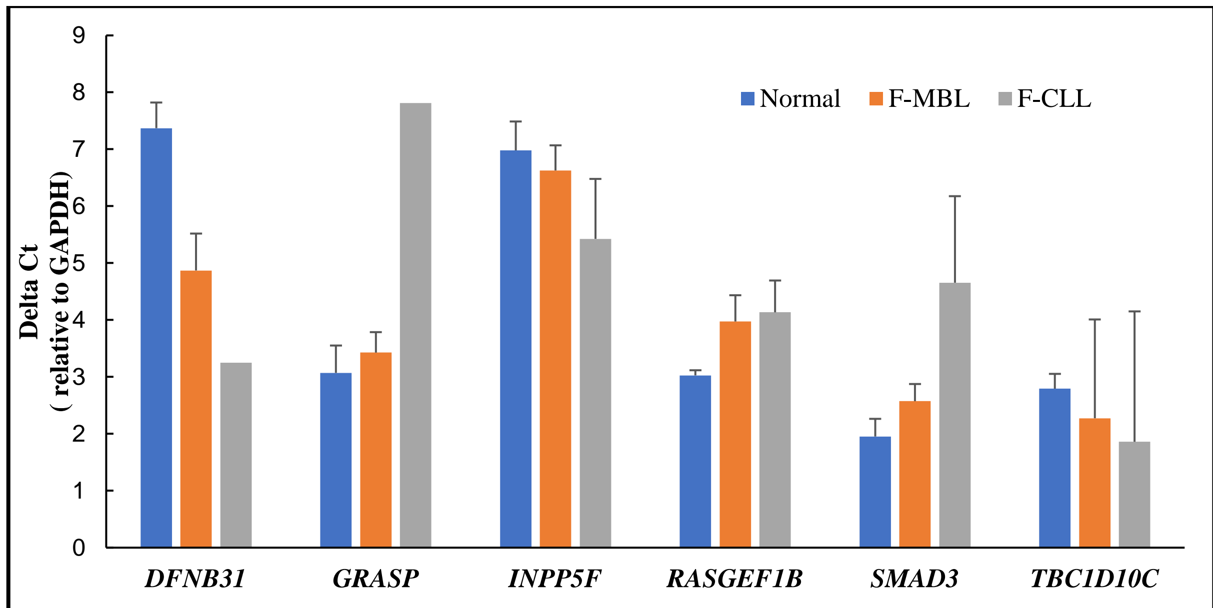

Figure 5-4. Quantitative RT-PCR analyses for 6 mRNAs in B-lymphocytes from control, F-MBL and F-CLL subjects

Delta Ct (cycle threshold) is defined by the number of cycles required for the fluorescent signal to cross threshold background fluorescence. Ct levels are inversely proportional to the amount of target mRNA, that is, the lower the Ct the greater the amount of target mRNA in the sample.

5.5. DISCUSSION

This study of differential gene expression in control B-lymphocytes, F-MBL and F-CLL from a family with multiple affected members allowed comparison of mRNA profiles associated with progression to CLL that to some extent was controlled for genetic background. Using unsupervised hierarchical clustering, with progression from normal familial controls through F-MBL to F-CLL, 6 genes were found to be dysregulated. The mRNA for *GRASP* was downregulated, while *LEF1*, *C11orf80*, *ROR1*, *METTL8*, and *PARP3* were upregulated. To further identify genes that become increasingly dysregulated with neoplastic progression, mRNAs that were differentially abundant between F-CLL and F-MBL were identified, and then compared between F-MBL and family controls. Six genes were differentially expressed in control versus F-MBL, and F-MBL versus F-CLL. With progression from F-MBL to F-CLL, *INPP5F* and *DFNB31/WHRN* were progressively upregulated and *SMAD3*, *GRASP* and *RASGEF1B* were progressively downregulated. The mRNA for *TBC1D10C* was downregulated in F-MBL before becoming upregulated in F-CLL. Of these 11 genes, *LEF1, ROR1, INPP5F,* and *SMAD3*, have previously been associated with either the development of B-CLL or progression of MBL to B-CLL.

The transcription factor, *LEF1,* is involved in the development of B lymphocytes and is highly expressed in mouse pro-B and pre-B lymphocytes but downregulated in mature B cells (Reya et al., 2000) (Gutierrez et al., 2010). *LEF1* functions in the Wnt/β-catenin signalling pathway, recruiting β-catenin to activate transcription of several target genes in response to constitutive Wnt pathway activation, which regulates B lymphocyte proliferation and survival (Gutierrez et al., 2010). B-CLL cells aberrantly express *LEF1* compared to normal B cells and *LEF1* knockdown or LEF1 inhibition by small molecules decreases CLL B-cell survival (Gutierrez

et al., 2010, Gandhirajan et al., 2010). In addition, high expression of *LEF1* in CLL has been associated with UM-CLL (Wu et al., 2016).

Increased expression of *ROR1*, a type -1 tyrosine kinase-like orphan receptor, was associated with progression from control to F-MBL and F-CLL. ROR1 signalling is involved in cell proliferation and differentiation, and embryonic development (van Genderen et al., 1994), and over-expression of *ROR1* on the surface of B-CLL has been documented in several studies (Baskar et al., 2008, Daneshmanesh et al., 2008). *ROR1* acts as a receptor for Wnt5 signalling, which increases B-CLL cell survival, proliferation and migration (Yu et al., 2016). These effects are blocked by cirmtuzumab, a humanized anti-ROR1 monoclonal antibody (Yu et al., 2016).Furthermore, inhibiting *ROR1* in CLL cells by siRNA silencing induces apoptosis of B-CLL cells but not control B cells (Choudhury et al., 2010). High level expression of *ROR1* has been associated with disease progression and may distinguish patients with aggressive from indolent disease (Cui et al., 2016). Consequently *ROR1* has been considered as a target for new CLL therapies (Aghebati-Maleki et al., 2017).

INPP5F was upregulated in association with CLL compared to MBL and MBL compared to controls. INPP5F degrades PIP2 (phosphatidylinositol 4,5-bisphosphonate) and PIP3 (phosphatidylinositol 3,4,5-trisphosphonate) regulating AKT/phosphatidylinositol 3-kinase (PI3K) signalling and PIP3 levels (Zhu et al., 2009). There may be an association between high expression of *INPP5F* and activated NF-κB, as *INPP5F* expression positively correlates with IKKb/IKBKB, an activator of NF-κB, and negatively with IKBa, an inhibitor of NF-κB (Inoue et al., 2007, Karin, 2006). In CLL treated with fludarabine-based therapies, low *INPP5F* mRNA levels are associated with better outcome compared to high levels (Palermo et al., 2015), and *INPP5F* mRNA level may be a useful prognostic biomarker.

SMAD3 is a member of the SMAD family which play an essential role in intracellular signalling of the transforming growth factor-β (TGF-β). This pathway has been shown to regulate cell proliferation, differentiation, migration and apoptosis in many different cell types (Derynck and Zhang, 2003). In this pathway, SMAD3 and SMAD2 bind together to act as a transcription factor and form receptor-regulated SMADs (R-SMADs) complex. This complex facilitates binding to a common mediator, SMAD4 protein, to form a SMAD complex, which then translocates to the nucleus. Once SMAD complex enters the nucleus, it regulates the transcription of several target genes (Derynck and Zhang, 2003). In addition to the role of SMAD3 as a transcription factor in the intracellular signalling pathway of TGF-β, it has been also reported that SMAD3 mediates upregulation of microRNA-21 to promote renal fibrosis (Zhong et al., 2011). In CLL, high microRNA-21 expression has been previously associated with poor prognosis in patients with CLL (Rossi et al., 2010). However, it is not clear how the expression of SMAD3 is involved in the biology of CLL cells, although it could be via TGF-β (Douglas et al., 1997, Matveeva et al., 2017) or microRNA-21 (Rossi et al., 2010).

In addition to validating associations between *LEF1, ROR1, INPP5F,* and *SMAD3*, this study identified 7 novel associations, including an association with *PARP3*, which facilitates the formation and maintenance of the mitotic spindle and genome integrity (Boehler et al., 2011), and is currently being investigated as a target for cancer therapy (Oplustil O'Connor et al., 2016).

Decreased expression of *GRASP* was associated with progression from control to F-MBL and F-CLL. The *GRASP* gene encodes the general receptor for phosphoinositides 1-associated scaffold protein. In neurones, GRASP interacts with scaffold proteins involved in postsynaptic organization and protein trafficking (Kitano et al., 2003), and GRASP may be involved in receptor clustering, trafficking, and intracellular signalling. GRASP has been found

to coordinate the association of ADP ribosylation factors (ARF) activating protein with the Rac-activating protein Dock180, to promote ARF to Rac signalling networks and cell migration (White et al., 2010, Attar and Santy, 2013). ARFs are members of the Ras superfamily of small GTPases. The 6 mammalian ARFs are divided into 3 classes based on sequence similarity, and regulate vesicular trafficking, cell shape, and movement (Pasqualato et al., 2002). Compared to control B lymphocytes, *GRASP* is downregulated, (Liao et al., 2015), however, whether *GRASP* plays a role in the development or progression of CLL has not been studied.

Of interest were 3 genes (*LEF1*, *ROR1* and *GRASP*) that were also differentially expressed in the F-LPD versus S-CLL comparison reported in chapter 3, a result which is consistent with the proposal that inherited upregulation or downregulation of these genes contributes to driving malignant progression of F-LPD.

5.6. CONCLUSION

The identification of differentially abundant mRNAs between normal B-lymphocytes, pre-malignant F-MBL cells and F-CLL will provide potential new biomarkers for identifying patients at increased risk of progression, and an understanding of the mechanisms of neoplastic transformation. In this chapter associations between mRNA levels and 4 genes known to be involved in either the development or progression of CLL were validated, and 7 novel associations were identified for further studies.

CHAPTER-6: A COMPARISON OF MESSENGER RNA AND PROTEIN PROFILES IN IGH MUTATED AND UNMUTATED CLL

6.1. INTRODUCTION

This Chapter will investigate whether gene and protein signature profiles in F-LPD and S-CLL differentiate between cases with mutated *IGH* genes (M-CLL) and unmutated *IGH* genes (UM-CLL). CLL cases with ≥ 98% sequence homology with germline define as UM-CLL, and those displaying homology of < 98% are designated as M-CLL (Ghia et al., 2007). In normal B cell development, antigen-activated B cells that receive appropriate T-cell help migrate into primary B-cell follicles, where they establish germinal centres (Chiorazzi et al., 2005). Proliferating germinal centre B cells, undergo *IGH* somatic hypermutation, where mutations occur specifically in *IGH* region genes. By comparing normal B cell development with B-CLL, it had been proposed that UM-CLL B cells originate from a pre-germinal centre precursor with *IGH* lacking somatic mutations, and M-CLL B cells were derived from post-germinal centre B cells that express B cell receptors (BCR) altered by somatic hypermutations (Fabbri and Dalla-Favera, 2016). However, association studies of sporadic M-CLL and UM-CLL using gene expression profiling have not provided strong evidence of these subtypes arising from separate, distinct cells of origin (Klein et al., 2001, Rosenwald et al., 2001, Ferreira et al., 2014), and suggest a common mechanism of transformation or cell origin, which is likely to be within the germinal centre (Herve et al., 2005). This proposal is supported by expression of surface markers of activation by UM-CLL cells (Chiorazzi and Ferrarini, 2003, Stevenson and Caligaris-Cappio, 2004), and telomere length measurements which show B-CLL cells have undergone a number of cell divisions prior to leukaemic transformation, most likely after exposure to antigen (Damle et al., 2004).

The membrane expression of markers of cell activation, and evidence of continued activation, underscores the importance of signaling via the B cell receptor (BCR) and the

potential presence of an unidentified foreign or self-antigen which may play a role in malignant transformation (Herve et al., 2005).

Compared to mRNA studies, protein expression studies have found higher percentages of differentially expressed proteins in M-CLL compared to UM-CLL (Eagle et al., 2015, Perrot et al., 2011, Huang et al., 2016). Isobaric tags for relative and absolute quantification (iTRAQ)-based MS followed by system biology analysis has been used to compare proteomes in 9 M-CLL and 9 UM-CLL subjects (Eagle et al., 2015). This study reported 274 proteins, or 8% of all proteins identified, were differentially abundant in the 2 subgroups, and 43 cell migration and adhesion pathways were found to be differentially active in M-CLL and UM-CLL. A high proportion of differentially under-expressed proteins in UM-CLL were involved in cell migration processes, suggesting UM-CLL cells have a defect in migration (Eagle et al., 2015). Proteins associated with cytoskeletal remodelling were also underexpressed in UM-CLL whereas proteins associated with transcriptional and translational activity were overexpressed (Eagle et al., 2015). These findings suggested that UM-CLL cells were less migratory and more adhesive than M-CLL cells, resulting in prolonged retention in the lymph nodes and exposure to proliferative and pro-survival signals (Eagle et al., 2015). In a second study that used iTRAQ analysis, an initial screen in 27 patients identified differentially abundant proteins in progressive CLL, which were then validated by elected reaction monitoring (SRM) analysis of purified CD19+ CLL cells in a second sample of 50 cases (Huang et al., 2016). Although cases were not classified according to *IGH* mutation status, proteins involved in cell proliferation, survival, DNA repair, granzyme A signalling, and apoptosis were found to be differentially expressed in stable, slow progressive and progressive disease.

Quantitative 2D-fluorescence difference Gel electrophoresis (2D-DIGE) has also been used to study 6 CLL patients (3 M-CLL and 3 UM-CLL) before and after BCR stimulation by anti-

IgM stimulation (Perrot et al., 2011). This study reported that unstimulated M- and UM-CLL cells displayed distinct proteomic profiles, which was more pronounced in UM-CLL (Perrot et al., 2011).

Although family studies of M-CLL and UM-CLL are limited by low subject numbers, the advantages of these studies are that background genetic variation is to some extent controlled, increasing the contribution of epigenetic and environmental modifiers to variation in phenotype (Borecki and Province, 2008). The study reported in this Chapter is the first gene expression and proteomic study to analyse familial cases, and in relation to protein studies, the first to normalize protein abundance in CLL cases to control samples.

HYPOTHESIS

Insights may be gained into whether there are differences in gene and protein expression in M-CLL and UM-CLL by studying F-CLL and F-MBL cases, where background genetic variation is partially controlled.

AIMS

1. Compare M-CLL and UM-CLL mRNA profiles in sporadic CLL and familial LPD cases using high-resolution microarrays and unsupervised hierarchical clustering.
2. To normalise for genes differentially expressed as a result of inter-individual B lymphocyte variation, compare mRNAs between (A) controls and M-CLL and (B) M-CLL and UM-CLL in sporadic CLL and familial LPD cases.
3. To identify mRNAs differentially expressed in family cases alone, compare mRNAs between (A) controls and M-CLL and (B) M-CLL and UM-CLL in familial cases alone.

4. Compare M-CLL and UM-CLL protein profiles corrected for normal controls in sporadic CLL and familial LPD cases using quantitative MS.

6.2. METHODS

6.2.1. M-CLL and UM-CLL mRNA Profiles

Sequence analysis of *IGH* was performed in 13 individuals with B-CLL (7 S-CLL and 2 F-CLL) and in 4 F-MBL (Table 6-1), and compared to known germline encoded *IGH* segments as described previously in section 2.6.2.5. Patients with ≥ 98% sequence homology with germline were classified as UM-CLL, and those displaying homology of < 98% were classified M-CLL (Ghia et al., 2007).

Table 6-1. List of the selected samples applied in mRNA and Protein profiles in IGH M-CLL and UM-CLL

Sample ID	Gender	Absolute lymphocyte count (x 10^9/L)	*IGH* mutational status	*IGH* region containing the mutation
S-CLL 52	67y/M	9	Mutated	*IGH1*
S-CLL 53	92y/F	41	Unmutated	NA
S-CLL 57	80y/F	29	Mutated	*IGH3*
S-CLL 58	77y/F	41	Mutated	*IGH3*
S-CLL 85 *	64y/M	9	Mutated	*IGH1*
S-CLL 87	56y/M	17	Unmutated	ND[+]
S-CLL 88	73y/F	108	Unmutated	*IGH2*
F-MBL (IV-13)	61y/F	1	Mutated	*IGH3*
F-MBL (III-10)	78y/M	1	Mutated	*IGH5*
F-MBL (IV-17)	54y/M	2	Mutated	*IGH3*
F-MBL (IV-18)	52y/M	2	Mutated	*IGH4*
F-CLL90 (IV-05)	54y/F	10	Mutated	*IGH4*
F-CLL91 (IV-02)	62y/F	17	Unmutated	NA

* S-CLL 85 was used for protein profiles in M-CLL and UM-CLL

[+] ND: not defined

To identify differentially expressed genes, 8 CLL cases were selected and grouped into 4 M-CLL and 4 UM-CLL, according to *IGH* mutational status. The groups comprised 2 F-CLL cases (1 mutated and 1 unmutated), 6 S-CLL (3 mutated, 3 unmutated) and 4 control samples (2 family members and 2 unrelated controls). Probe set analysis results (.CHP file format) for these 12 samples were loaded into the Affymetrix Transcriptome analysis console (TAC 3.0). One-way (between-subjects) ANOVA (unpaired) was selected and TAC analysis was performed comparing M-CLL, UM-CLL and controls.

To identify mRNAs that were differentially expressed based on *IGH* mutation status alone, normalised for genes differentially expressed as a result of expected interindividual B lymphocyte variation, mRNAs in M-CLL and UM-CLL cases were compared to controls. Semi-supervised clustering was then performed for those mRNAs that were differentially expressed between all 3 groups First, a 2-sample t-test was performed to compare M-CLL and UM-CLL groups. All genes that had a log fold change > 2, and FDR p-value < 0.05 were checked to ensure that they were also different on a 2-sample t-test comparison between (A) controls and M-CLL and (B) controls and UM-CLL. mRNAs that had p-values < 0.05 were considered for semi-supervised hierarchical clustering (Bair, 2013, Bair and Tibshirani, 2004).

The same method was used to identify mRNAs differentially expressed based on *IGH* mutation status using familial M-CLL and UM-CLL cases and familial controls. The subjects in this study comprised 3 related normal controls (V-16, IV-15 and IV-16), 5 F-LPD with mutated *IGH* [4 F-MBL (III-10, IV-13, IV-17 and IV-18), and 1 F-CLL (IV-05)], and 1 F-CLL (IV-02) with unmutated *IGH*.

6.2.2. Protein Profiles in M-CLL and UM-CLL

To determine if protein profiles were different in UM-CLL compared to M-CLL, purified

B-lymphocytes from UM-CLL cases (3 sporadic) were compared to 5 M-CLL cases (4 sporadic and 1 familial) and controls (one familial and one unrelated). One UM-CLL patient (IV-2) had required treatment with FCR between the times of collection of mRNA and protein samples, and this subject was not included in the proteomic analysis. In addition, F-MBL cases were not included because sufficient protein could not be extracted from low numbers of clonal B cells.

Proteins from purified B-CLL cells were precipitated and quantified, reduced and alkylated before tryptic digestion, as described in Section 4.5. Briefly, samples were purified using HLB plus short cartridges and reconstituted in 100mM HEPES buffer (pH 8.0) for TMT 10plex Mass Tag labelling (Thermo Scientific), following the manufacturer's instructions. The TMT labelled peptides were purified using HLB and fractionated by hydrophilic interaction liquid chromatography (HILIC) in offline mode using in-house packed TSK-Amide 80 HILIC columns. The TMT labelled HILIC fractions were re-suspended in MS Loading buffer and analyzed online by nano-capillary LC-MS-MS using a Dionex Ultimate 3000 HPLC system and a Thermos Scientific Orbitrap Fusion Tribrid™ Mass Spectrometer. The MS data. RAW files from each TMT experiment sample set were interrogated using the SEQUEST HT database search engine with percolator validation to perform protein identification. The Proteome Discoverer version 2.1.0.81 (Thermo Scientific) calculated abundance ratios for each sample against a normal control and protein quantification values were exported for further analysis to Excel to calculate the log fold change and p-value. Statistical significance of the difference in levels of expression of proteins between M-CLL and UM-CLL was determined using functions in the R computational environment. The heat map from protein expression data was constructed using the function "heatmap.2" in R that uses a Euclidean distance metric (Warnes et al., 2016). Functional networks of differentially abundant proteins were determined using Ingenuity Pathway Analysis (IPA) software. (IPA; ingenuity systems,

http://www.ingenuity.com).

6.3. RESULTS

6.3.1.M-CLL and UM-CLL mRNA Profiles

After performing TAC analysis, 2160 mRNA transcripts (1533 coding and 627 noncoding transcripts) were found to be differentially expressed between M-CLL, UM-CLL and controls. To minimize false positive results, an FDR p-value < 0.05 threshold was set and 582 differentially expressed mRNAs were identified. Controls and M-CLL cases clustered, however the unmutated F-CLL (IV-02) case did not cluster with sporadic UM-CLL cases (Figure 6-1). Family controls (IV-15 and IV-16) segregated from unrelated controls (1st and 3rd NC) (Figure 6-1). Based on FDR p value, the highest ranked genes downregulated in UM-CLL versus normal controls and further downregulated in M-CLL were *EBF1* ($p < 0.0001$), *OR2L1P* ($p = 0.0002$) and *PTPRK* ($p = 0.0006$). The most highly ranked upregulated genes in UM-CLL versus normal controls and further upregulated in M-CLL were *LEF1* ($p < 0.0001$), *PITPNM2* ($p = 0.0002$), *TEAD*2 ($p = 0.0005$), and *ABCA6* ($p = 0.0005$).

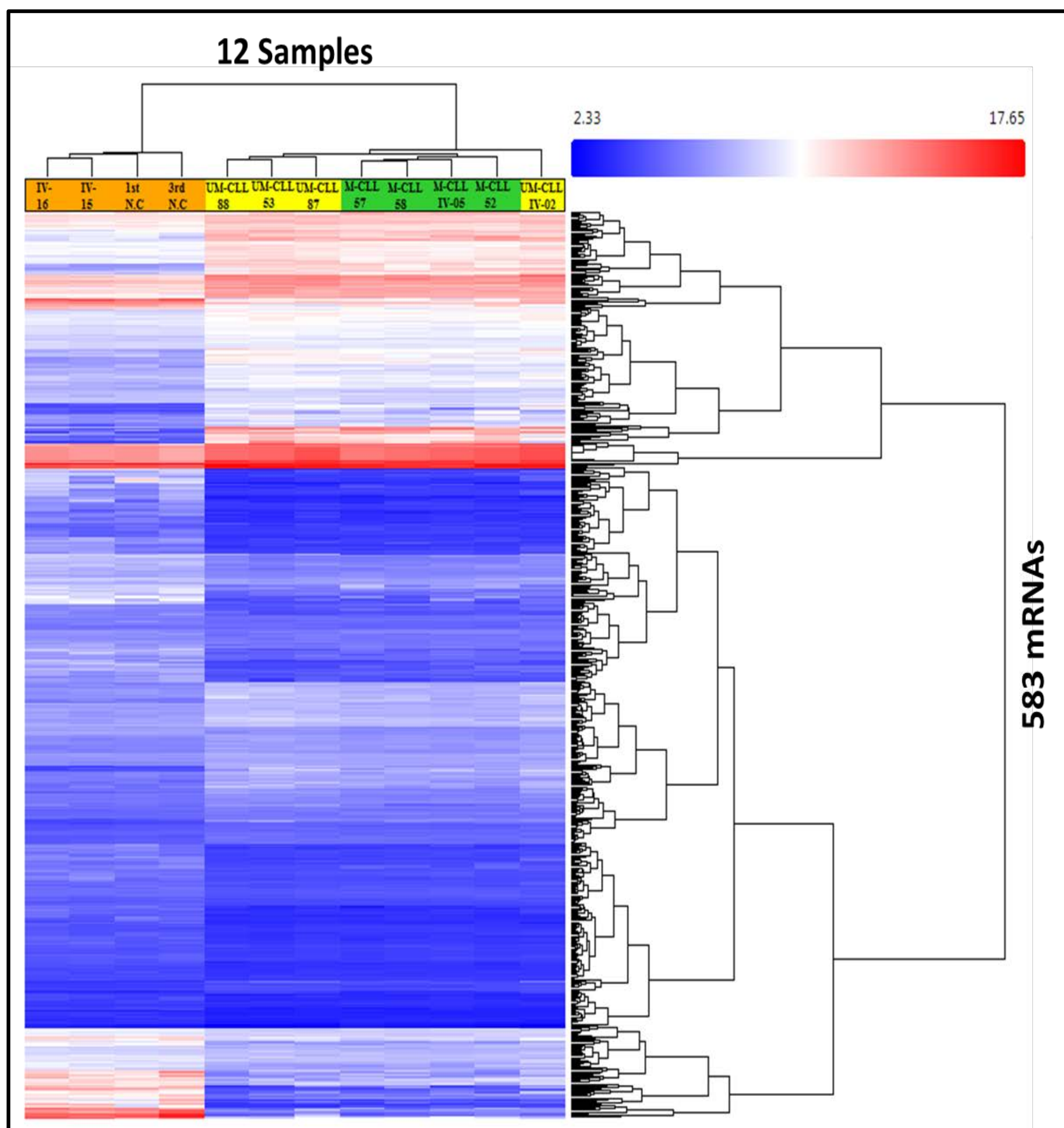

Figure 6-1 Cluster image of mRNA expression in normal, M-CLL and UM-CLL

Unsupervised hierarchical clustering of 12 M-CLL, UM-CLL and related (IV-15 and IV-16) and unrelated control subjects (1st NC and 3rd NC) showing 583 mRNAs representing ≈430 genes (FDR $p < 0.05$).

To identify mRNAs that were differentially expressed based on *IGH* mutation status alone, mRNAs in M-CLL and UM-CLL cases were compared after both groups were normalized to controls (t-test p-value < 0.05) (Figure 6-2).

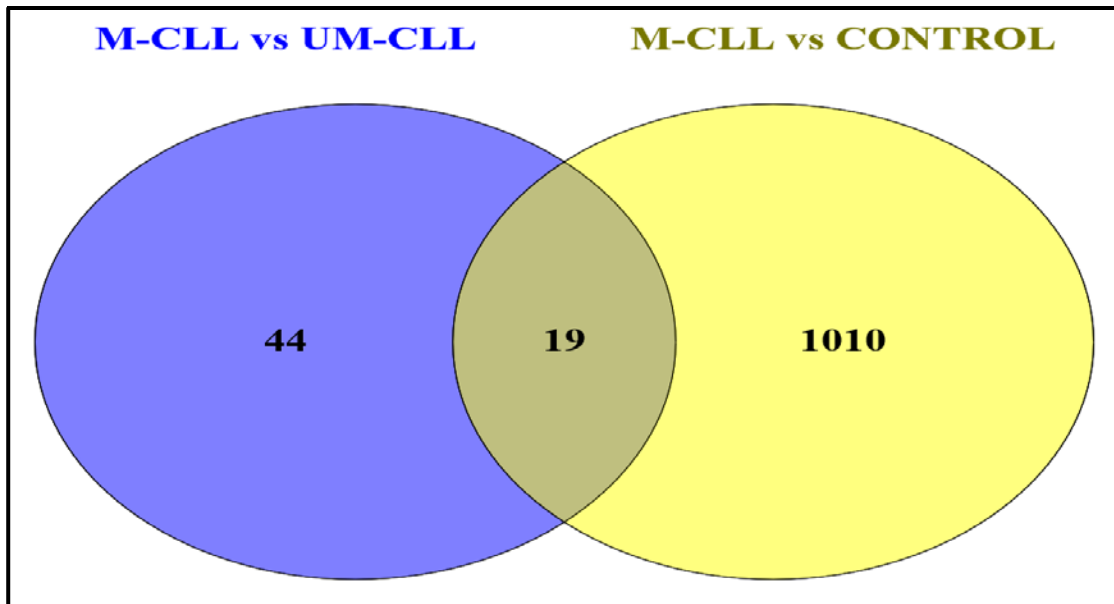

Figure 6-2. Venn diagram illustrating overlap between M-CLL vs UM-CLL, and M-CLL vs Controls

Nineteen genes were differentially abundant with t-test p < 0.05; >2-fold change in M-CLL versus UM-CLL, and also differentially abundant in M-CLL versus controls. M-CLL; chronic lymphocytic leukaemia with mutated *IGH*, UM-CLL; chronic lymphocytic leukaemia with unmutated *IGH*. Retrieved from http://bioinfogp.cnb.csic.es/tools/venny/index.html on March 7th 2018.

Compared to M-CLL, 10 mRNAs were upregulated in UM-CLL (*CD69*, *FOSB*, *CLEC2B*, *CERS6*, *SNORD3C*, *L3MBTL4*, *ME2*, *HIGD1C*, *C16orf54* and *RGCC*); while 9 were downregulated (*ADAM29*, *SFMBT2*, *LINC01224*, *SDK2*, *KLF3* and *NRIP1*). (Figure 6-3; Table 6-2). These 19 mRNAs represented 17 genes, with CD69 and FOSB each represented by 2 mRNAs. Of these mRNAs, *NRIP1* and *CERS6* segregated M-CLL, UM-CLL cases and controls (FDR < 0.05) using semi-supervised hierarchical clustering (Figure 6-4; Table 6-3).

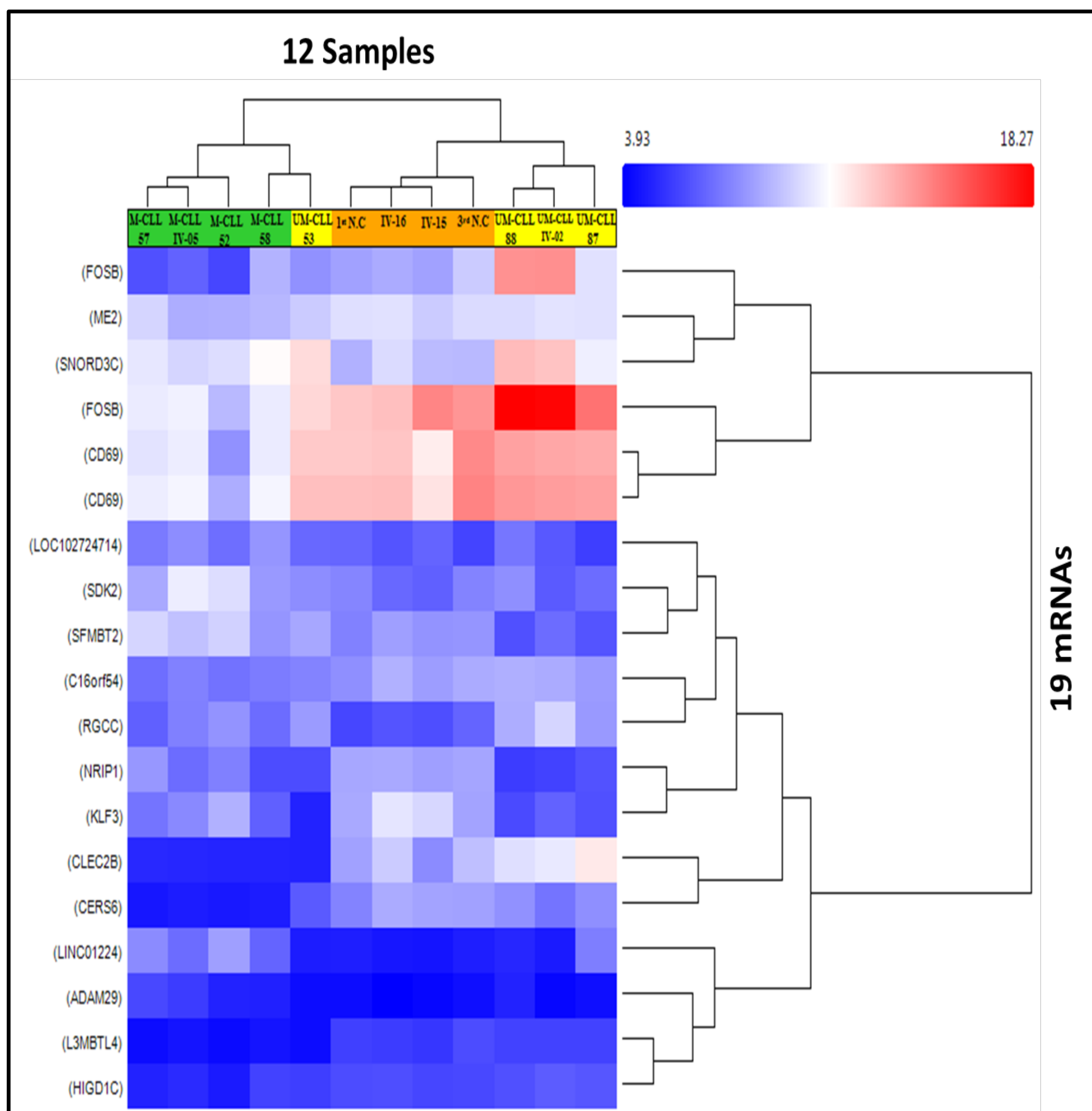

Figure 6-3.Cluster image of normal, Mutated and Unmutated CLL gene expression following correction for relatedness.

Semi-supervised hierarchical clustering of 19 cDNA elements which showed a significant (t-test $p<0.05$) 2-fold change in regulation between normal, M-CLL and UM-CLL groups. The cluster dendrogram linked UM-CLL cases with healthy control group before they linked to M-CLL cases, and one UM-CLL case (53) segregated with the M-CLL group.

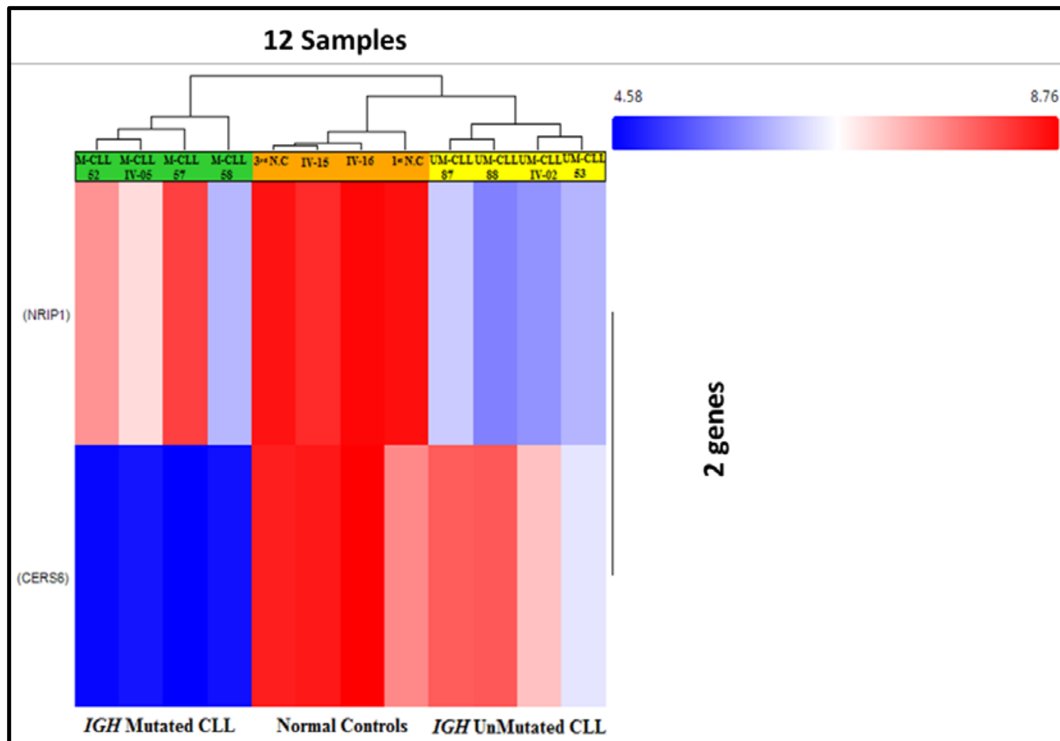

Figure 6-4. Cluster image of 2 mRNAs between M-CLL, UM-CLL and controls

Semi-supervised hierarchical clustering of 2 mRNAs which showed a significant (t-test $p<0.05$) 2-fold change in regulation between normal, M-CLL and UM-CLL and also with FDR<0.05.

Table 6-2 Genes differentially expressed between M-CLL and UM-CLL normalised for controls groups

Symbol	***Entrez Gene Name***	**Fold Change ***	**p-value**	**Location**	**Type(s)**
FOSB[+]	*FosB proto-oncogene, AP-1 transcription factor subunit*	54.4	0.009	Nucleus	transcription regulator
CLEC2B	*C-type lectin domain family 2 member B*	53.6	0.028	Plasma Membrane	other
CD69[+]	*CD69 molecule*	8.5	0.0009	Plasma Membrane	transmembrane receptor
CERS6	*ceramide synthase 6*	7.3	0.0003	Nucleus	transcription regulator
SNORD3C	*small nucleolar RNA, C/D box 3C*	4.8	0.03	Other	other
L3MBTL4	*L3MBTL4, histone methyl-lysine binding protein*	2.7	0.035	Other	other
ME2	*malic enzyme 2*	2.5	0.023	Cytoplasm	enzyme
HIGD1C	*HIG1 hypoxia inducible domain family member 1C*	2.4	0.012	Other	other
C16orf54	*chromosome 16 open reading frame 54*	2.3	0.012	Extracellular Space	other
RGCC	*regulator of cell cycle*	2.3	0.025	Cytoplasm	other
LINC01857	*long intergenic non-protein coding RNA 1857*	-2.0	0.045	Other	other
ADAM29	*ADAM metallopeptidase domain 29*	-2.1	0.03	Plasma Membrane	peptidase
NRIP1	*nuclear receptor interacting protein 1*	-2.5	0.037	Nucleus	transcription regulator
KLF3	*Kruppel like factor 3*	-2.6	0.036	Nucleus	transcription regulator
SDK2	*sidekick cell adhesion molecule 2*	-4.2	0.022	Plasma Membrane	other
LINC01224	*long intergenic non-protein coding RNA 1224*	-6.3	0.04	Other	other
SFMBT2	*Scm like with four mbt domains 2*	-9.4	0.017	Nucleus	other

Genes differentially expressed between Mutated and Un-Mutated CLL normalised for controls groups with > 2 fold change and significance level p-value < 0.05

[+] FOSB and CD69 each represented by 2 mRNAs

* Fold change M-CLL versus UM-CLL.

Table 6-3. Genes differentially expressed between normal, M-CLL and UM-CLL groups (FDR < 0.05)

Gene symbol	Gene Name	Normal Controls signal intensity *	M-CLL signal intensity	UM-CLL signal intensity	FDR p-value
CERS6	*ceramide synthase 6*	8.59	4.68	7.54	**0.005**
NRIP1	*nuclear receptor interacting protein 1*	8.63	7.24	5.95	**0.04**

*Signal intensity is taken as an average over the pairs of perfect match (PM) and mismatch (MM) probe spots. The statistical method used to summarize over the PM/MM pairs was the Tukey's Bi-weight average algorithm implemented in Affymetrix software. (Bolstad et al., 2003)

2 genes showed a significant (t-test p<0.05), 2-fold change and FDR<0.05 in regulation between normal, M-CLL and UM-CLL.

6.3.2. mRNA Profiles in *IGH* Mutated and Unmutated CLL Family Cases

To identify mRNAs differentially expressed based on *IGH* mutation status using familial M-CLL and UM-CLL cases and familial controls, unsupervised hierarchical clustering was first performed using a threshold FDR p value < 0.05. There were 523 mRNAs differentially expressed between familial M-LPD (4 F-MBL and 1 F-CLL), UM-CLL (n=1) and controls (n=3) (Figure 6-5). Based on FDR p-value, the highest ranked genes upregulated in UM-CLL versus normal controls and further upregulated in M-CLL were *HSPA1B* (p < 0.0002), *CD3G* (p = 0.0006), and *RBMS3* (p = 0.0006). The most highly ranked downregulated genes in UM-CLL versus normal controls and further downregulated in M-CLL were *RPL9* (p < 0.0005), *RPS25* (p = 0.0001), *UMPS* (p = 0.01), and *CRTC3* (p = 0.01). Unsupervised hierarchical clustering segregated these 3 groups, with M-LPD segregating with controls before UM-CLL (Figure 6-5).

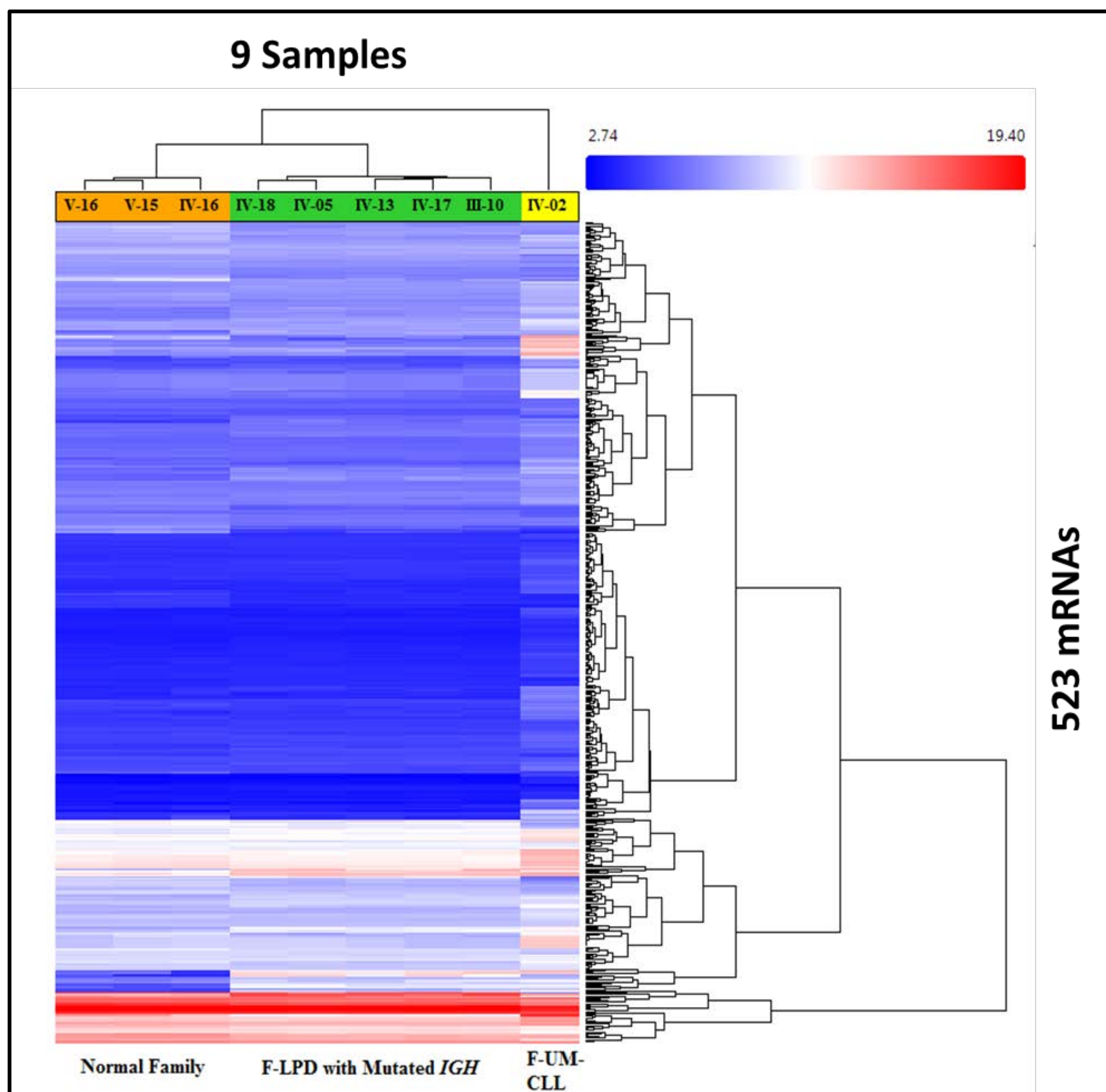

Figure 6-5. Cluster image of gene expression in familial M-CLL, UM-CLL, and familial controls

Unsupervised hierarchical clustering of 9 samples from familial M-CLL, UM-CLL and related (V-16, IV-15 and IV-16) control subjects showing 523 mRNAs representing ≈470 genes (FDR $p < 0.05$).

To identify mRNAs differentially expressed based on *IGH* mutation status alone and control for mRNAs that were differentially expressed between controls and all F-LPD cases, mRNAs were compared between (A) controls and M-LPD, and (B) M-LPD and UM-CLL (Figure 6-6).

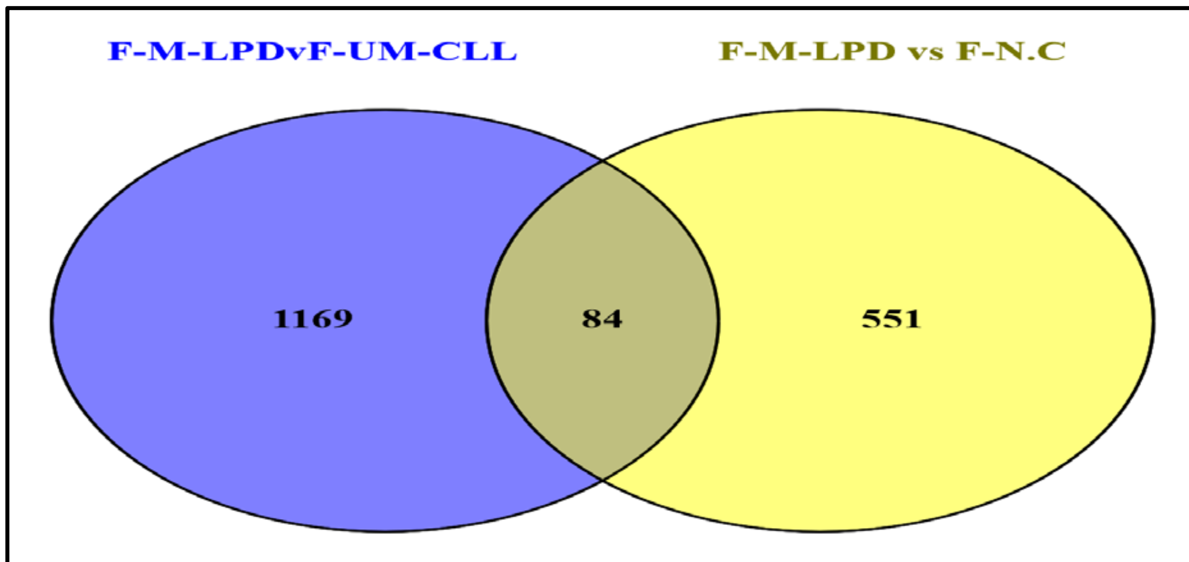

Figure 6-6. Venn diagram illustrating overlap between familial M-CLL, UM-CLL, and familial controls

Eighty-four mRNAs were differentially abundant with t-test p < 0.05 and >2-fold change in family controls and all F-LPD cases. F-M-LPD; familial-lymphoproliferative disease with mutated *IGH*, UM-CLL; familial chronic lymphocytic leukaemia with unmutated *IGH*, F-N.C; family controls. Retrieved from http://bioinfogp.cnb.csic.es/tools/venny/index.html on March 7th 2018.

There were 84 mRNAs differentially expressed between M-LPD and UM-CLL (t-test p < 0.05; fold change > 2), that did not differentiate both groups from controls (Figure 6-7). Based on t-test p value, the highest ranked genes include *CD72* (p<0.00002), *KLHL14* (p=0.0002), *SNORD14E* (p=0.0003), *FOSB* (p=0.0003), *TNF* (p=0.0007) and *IGHD3-22* (p=0.0008).

Semi-supervised hierarchical clustering using these genes segregated controls, M-LPD and UM-CLL, however in contrast to the unsupervised clustering results, UM-CLL segregated with controls before segregating with M-LPD (Figure 6-7). Genes that were more frequently upregulated in M-CLL cases include *KLHL14* (log 2-fold change = 34.1; p-value = 0.0002), *CTLA4* (log 2-fold change = 18.4; p-value = 0.02), *SESN3* (log 2-fold change = 18.1; p-value = 0.01), *WFDC21P* (log 2-fold change = 11; p-value = 0.01), and *PMAIP1* (log 2-fold change = 5.8; p-value = 0.01). Whereas for upregulated genes in UM-CLL cases, the highest expression levels were for *IGHD3-22* (log 2-fold change = 202; p-value = 0.0008), *FOSB* (log 2-fold change = 181;

p-value = 0.0003), *SNORD14E* (log 2-fold change = 80; p-value = 0.0003), *EGR1* (log 2-fold change = 18.2; p-value = 0.05) and *TNF* (log 2-fold change = 14.8; p-value = 0.0007). Of these 84 mRNAs, 18 genes segregated these 3 groups (FDR < 0.05), with M-LPD segregating with controls before UM-CLL. (Figure 6-8; Table 6-4).

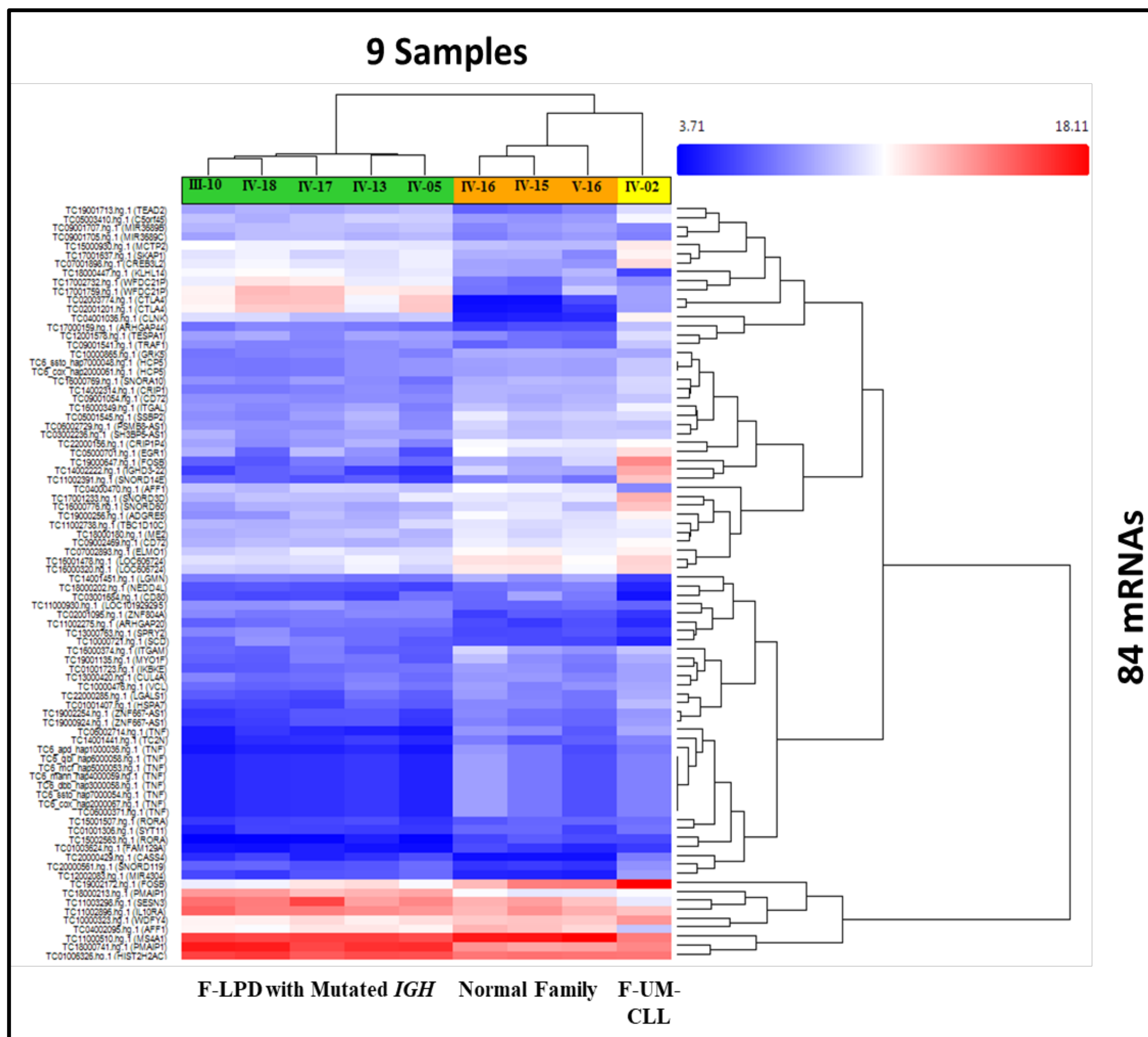

Figure 6-7. Cluster image of familial M-CLL and UM-CLL cases and family controls gene expression

Semi-supervised hierarchical clustering of 84 mRNAs which showed a 2-fold change (t-test p < 0.05) in expression between (A) controls and M-LPD, and (B) M-LPD and UM-CLL.

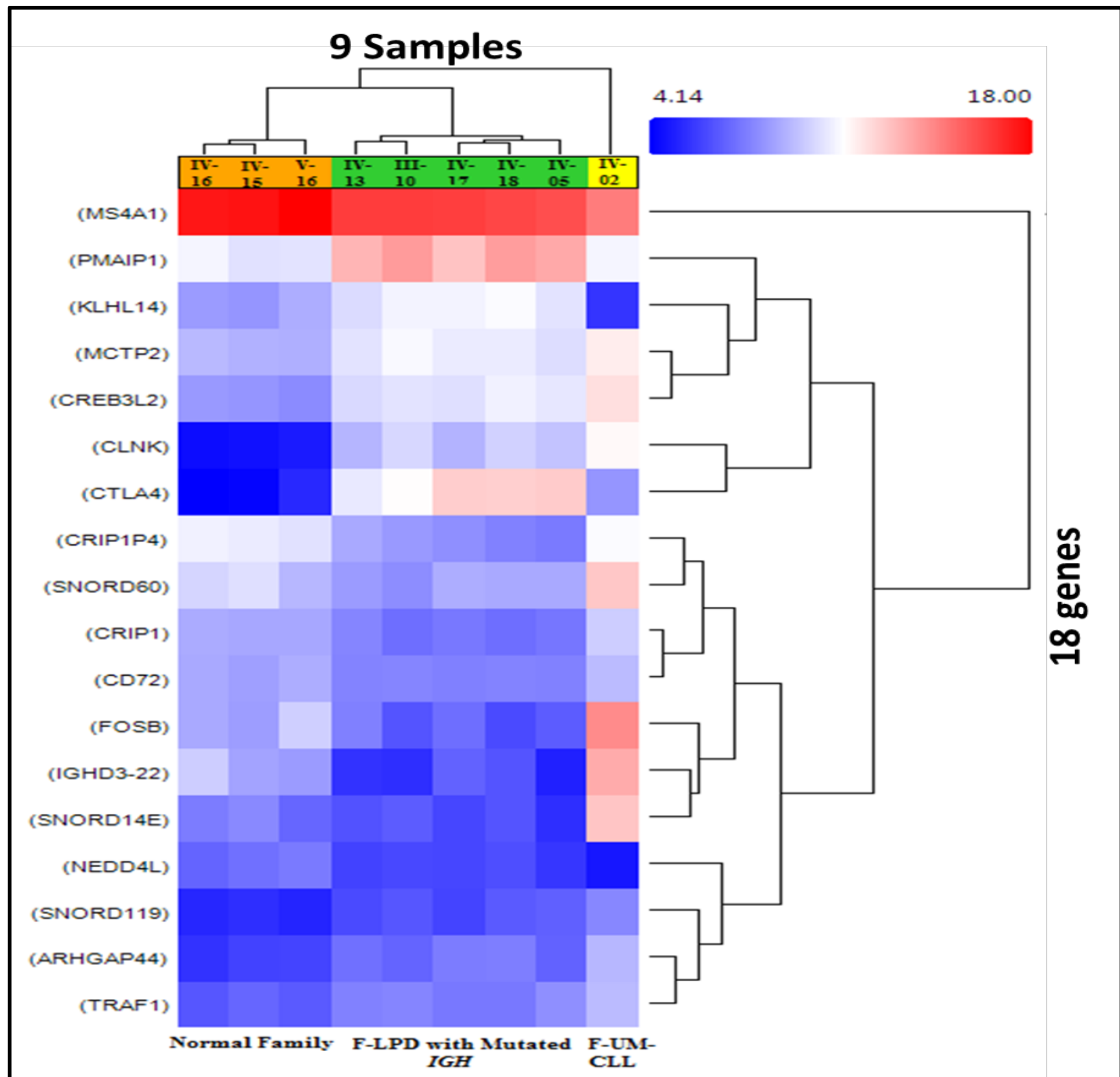

Figure 6-8. Cluster image of 18 genes between familial M-CLL, UM-CLL and familial controls (FDR < 0.05)

Semi-supervised hierarchical clustering of 18 genes which showed a significant (t-test p<0.05) 2-fold change in regulation between familial M-CLL, UM-CLL and familial normal and also with FDR<0.05.

Table 6-4. Genes differentially expressed between familial M-CLL, UM-CLL and familial normal groups (FDR < 0.05)

Gene Symbol	Gene Name	Familial Controls signal intensity *	Familial M-CLL signal intensity	Familial UM-CLL signal intensity	FDR p-value
CLNK	*cytokine-dependent hematopoietic cell linker*	4.64	9.48	11.25	**0.004522**
CREB3L2	*cAMP responsive element binding protein 3-like 2*	8.19	10.35	11.96	**0.00926**
CD72	*CD72 molecule*	8.81	7.72	9.28	**0.010823**
KLHL14	*kelch-like family member 14*	8.45	10.68	5.59	**0.014027**
SNORD14E	*small nucleolar RNA, C/D box 14E*	7.47	6.32	12.65	**0.021747**
CTLA4	*cytotoxic T-lymphocyte-associated protein 4*	4.21	12.45	8.25	**0.021747**
CRIP1	*cysteine-rich protein 1 (intestinal)*	8.74	7.35	9.74	**0.022436**
FOSB	*FBJ murine osteosarcoma viral oncogene homolog B*	8.77	6.76	14.26	**0.026734**
MS4A1	*membrane-spanning 4-domains, subfamily A, member 1*	17.48	16.33	14.62	**0.027878**
MCTP2	*multiple C2 domains, transmembrane 2*	8.97	10.51	11.58	**0.031369**
PMAIP1	*phorbol-12-myristate-13-acetate-induced protein 1*	10.3	13.39	10.84	**0.031673**
IGHD3-22	*immunoglobulin heavy diversity 3-22*	8.51	5.7	13.36	**0.033888**
ARHGAP44	*Rho GTPase activating protein 44*	5.97	7.2	9.16	**0.038423**
TRAF1	*TNF receptor-associated factor 1*	6.58	7.68	9.28	**0.039066**
SNORD119	*small nucleolar RNA, C/D box 119*	5.16	6.44	7.87	**0.044313**
NEDD4L	*neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase*	7.2	6.07	4.78	**0.04541**
CRIP1P4	*cysteine-rich protein 1 (intestinal) pseudogene 4*	10.54	8.07	10.99	**0.045846**
SNORD60	*small nucleolar RNA, C/D box 60*	9.95	8.74	12.6	**0.046455**

*Signal intensity is taken as an average over the pairs of perfect match (PM) and mismatch (MM) probe spots. The statistical method used to summarize over the PM/MM pairs was the Tukey's Bi-weight average algorithm implemented in Affymetrix software. (Bolstad et al., 2003).

18 genes showed a significant (t-test $p<0.05$), 2-fold change and FDR<0.05 in regulation between familial M-CLL, UM-CLL and familial normal controls.

6.3.3. Protein profiles in *IGH* mutated and unmutated CLL

Analysis of protein levels was performed in 5 M-CLL and 3 UM-CLL cases (n=8) using a combination of quadrupole, ion trap and Orbitrap mass spectrometer analysis, as described

in TMT10plex Mass Tag labelling 4.5.8. After normalization to 6 control samples, 5100 proteins were identified. There were 12 proteins differentially expressed between M-CLL and UM-CLL cases (log2-fold change > 2; t-test p < 0.05) (Figure 6-9; Table 6-5).

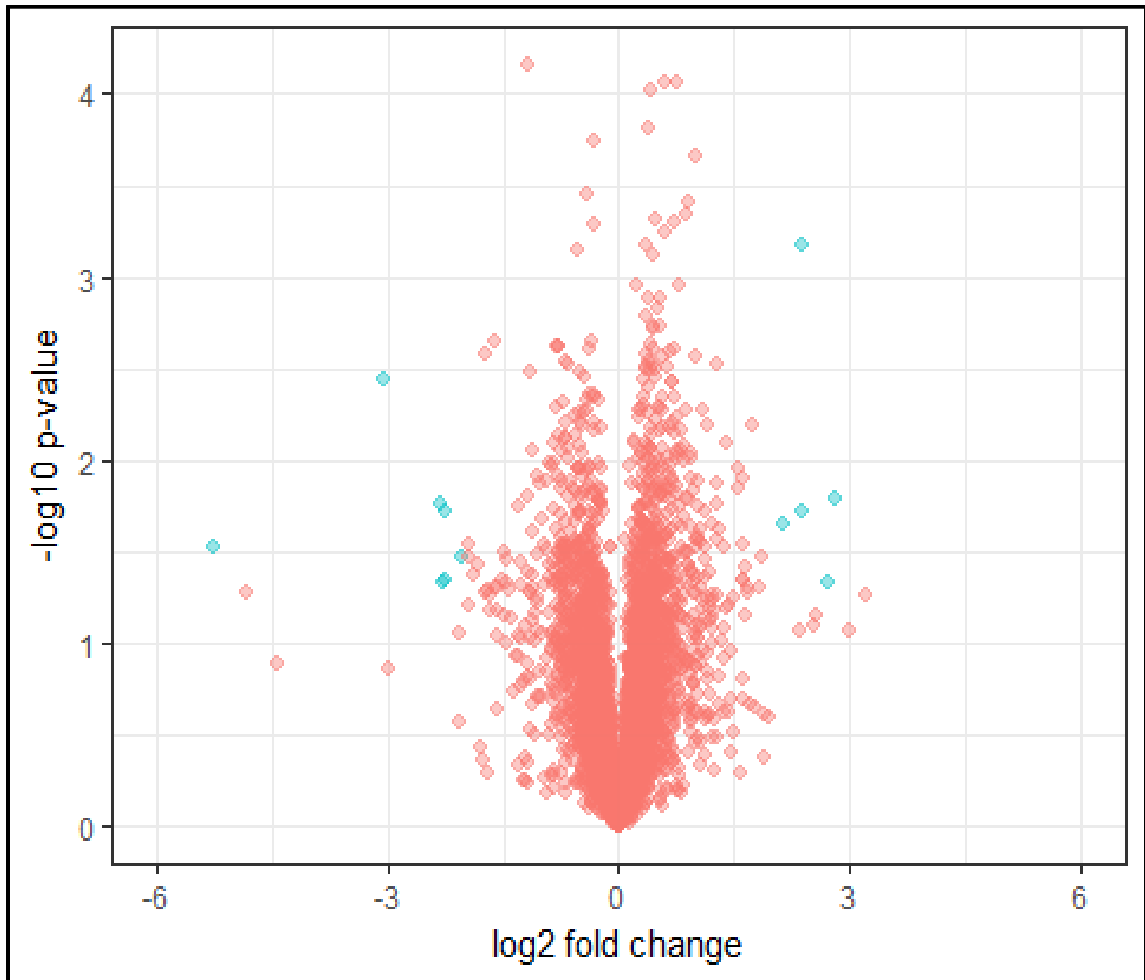

Figure 6-9. Volcano plot protein expression in M-CLL versus UM-CLL

The volcano plot shows the results of Log 2-fold change plotted against -log10 p value. The green diamond symbols represent 12 proteins that were differentially expressed between M-CLL and UM-CLL.

Of these 12 proteins, 7 proteins [MAGED4 (Fold change = -5.2; p-value = 0.03), ITPKC (Fold change = -3.0; p-value =0.004), HBD (Fold change = -2.3; p-value =0.02), PMVK (Fold change = -2.2; p-value =0.05), ALDH1A1 (Fold change = -2.3; p-value =0.04), EPB41 (Fold change = -2.2; p-value =0.02) and GYPA (Fold change = -2.04; p-value =0.03)] were under-

expressed in M-CLL compared to UM-CLL. Overexpressed proteins were PRB1/PRB2 (Fold change = 2.8; p-value =0.02), MRPL38 (Fold change = 2.7; p-value =0.05), CRIP1 (Fold change = 2.4; p-value =0.001), RPS19 (Fold change = 2.4; p-value =0.02) and ARGLU1 (Fold change = 2.1; p-value =0.02).

Table 6-5 Proteins differentially expressed between M-CLL and UM-CLL groups

Protein Symbol	Protein Name	Fold Change*	p-value	Location	Type(s)
PRB1/PRB2	proline rich protein BstNI subfamily 2	2.807	0.0163	Other	other
MRPL38	mitochondrial ribosomal protein L38	2.712	0.0468	Cytoplasm	other
CRIP1	cysteine rich protein 1	2.392	0.000665	Cytoplasm	other
RPS19	ribosomal protein S19	2.39	0.019	Cytoplasm	other
ARGLU1	arginine and glutamate-rich 1	2.141	0.0219	Other	other
GYPA	glycophorin A (MNS blood group)	-2.036	0.0333	Plasma Membrane	other
EPB41	erythrocyte membrane protein band 4.1	-2.244	0.0187	Plasma Membrane	other
ALDH1A1	aldehyde dehydrogenase 1 family member A1	-2.262	0.0454	Cytoplasm	enzyme
PMVK	phosphomevalonate kinase	-2.281	0.0468	Cytoplasm	kinase
HBD	hemoglobin subunit delta	-2.319	0.0171	Other	transporter
ITPKC	inositol-trisphosphate 3-kinase C	-3.065	0.00364	Cytoplasm	kinase
MAGED4 KIAA1859 MAGED4A MAGEE1; MAGED4B	MAGE family member D4	-5.263	0.0294	Other	other

Proteins differentially expressed between M-CLL and UM-CLL groups with > 2-fold change and significance level p-value < 0.05 (unadjusted for multiple comparisons).

*Fold change M-CLL versus UM-CLL.

However, hierarchical clustering using these 12 proteins did not segregate M-CLL from UM-CLL, with one UM-CLL case (UM-CLL 87) segregating with M-CLL (Figure 6-10).

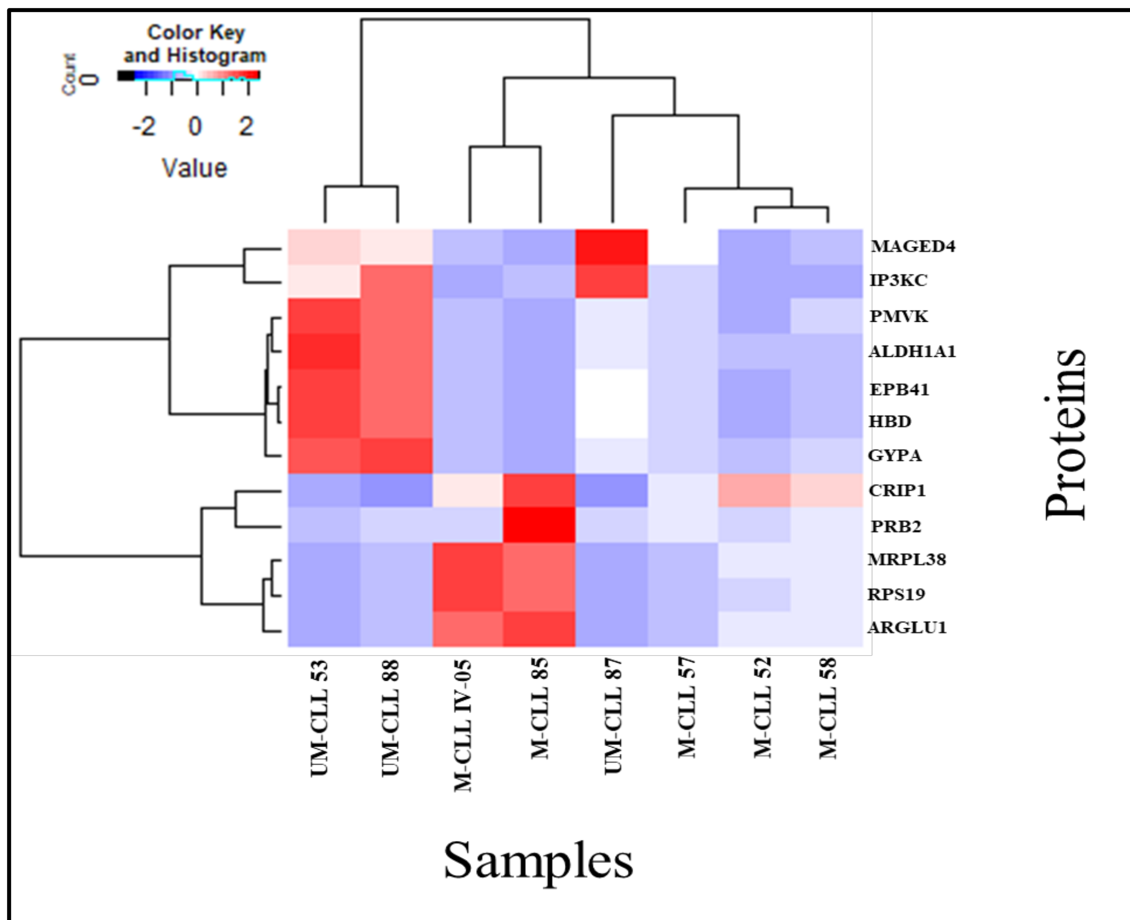

Figure 6-10. Clustering analysis of proteins differentially abundant between M-CLL and UM-CLL

Hierarchical clustering analysis using 12 proteins differentially expressed between M-CLL and UM-CLL groups (expression > 2-fold change and significance level p-value < 0.05 unadjusted for multiple comparisons). UM-CLL 87 clustered with M-CLL samples.

6.4. DISCUSSION

The presence or absence of somatic mutations in *IGH* genes has suggested that UM-CLL originates from a pre-germinal centre precursor and M-CLL B cells from post-germinal centre B cells (Damle et al., 1999, Fais et al., 1998, Hamblin et al., 1999). However, studies of sporadic M-CLL and UM-CLL using gene expression profiling have not found strong evidence of these subtypes arising from separate cells of origin (Klein et al., 2001, Rosenwald et al., 2001, Ferreira et al., 2014). Cell surface marker and telomere length studies suggest a common

mechanism of transformation or cell origin, which is likely to be within the germinal centre (Chiorazzi and Ferrarini, 2003, Stevenson and Caligaris-Cappio, 2004, Damle et al., 2004, Herve et al., 2005), and membrane expression of markers of cell activation emphasizes the importance of BCR signalling and the possibility of a foreign or self-antigen promoting malignant transformation (Herve et al., 2005). In the study reported in this Chapter, gene and protein expression were studied in a family segregating CLL, which was proposed to partially control for background DNA sequence variation and increase the likelihood of identifying epigenetic and environmental modifiers in M-CLL versus UM-CLL. Furthermore, it was proposed that mRNAs differentially expressed based on *IGH* mutation status alone may be identified by normalising gene mRNA expression in familial M-CLL and UM-CLL cases to family controls. It was proposed that mRNAs differentially expressed in M-CLL and UM-CLL, would provide evidence of the cell of origin for both CLL groups.

Using unsupervised hierarchical clustering, 582 mRNAs were differentially expressed between controls, and M-CLL and UM-CLL cases containing F-CLL and S-CLL. (FDR p-value <0.05). Although controls and M-CLL cases clustered, one UM-CLL (F-CLL IV-02) failed to segregate with other cases. To identify mRNAs that were differentially expressed based on *IGH* mutation status alone, mRNAs in M-CLL and UM-CLL cases were compared after both groups were normalized to controls (t-test p-value < 0.05). Compared to M-CLL, 10 mRNAs were upregulated in UM-CLL (*CD69*, *FOSB*, *CLEC2B*, *CERS6*, *SNORD3C*, *L3MBTL4*, *ME2*, *HIGD1C*, *C16orf54* and *RGCC*); while 9 were downregulated (*ADAM29*, *SFMBT2*, *LINC01224*, *SDK2*, *KLF3* and *NRIP1*). Of these mRNAs, *NRIP1* and *CERS6* segregated M-CLL, UM-CLL cases and controls (FDR < 0.05).

In agreement with previous reports, *NRIP1* was differentially expressed in M-CLL compared to UM-CLL. *NRIP1* encodes the nuclear receptor-interacting protein-1, which has

been shown to interact with a large number of nuclear receptors and transcription factors that regulate a wide range of biological process (Augereau et al., 2006). In CLL, the expression of *NRIP1* has been associated with favourable prognosis and longer overall survival (Herold et al., 2011, Lapierre et al., 2015). Furthermore, *NRIP1* expression has been found to be upregulated in M-CLL compared to UM-CLL (Oppezzo et al., 2005, Vasconcelos et al., 2005, Cornet et al., 2015).

Ceramide synthase 6 (CerS6) is a member of the ceramide synthase proteins family which play an important role in sphingolipid biosynthesis (Mullen et al., 2012), and have been implicated in the regulating cancer-cell growth, differentiation and apoptosis (Ogretmen and Hannun, 2004). It has been reported that CerS6 may have a role in autoimmune disease (Schiffmann et al., 2012), regulation of apoptosis in human head and neck squamous cell carcinoma (Senkal et al., 2010), and act as a potential transcriptional target of p53 (Fekry et al., 2016). The association of higher *CERS6* expression in UM-CLL reported in this Chapter is novel, and suggests CerS6 may be involved in the progression of CLL via its effects on p53 (Fekry et al., 2016) and sphingolipid metabolism-mediated resistance to apoptosis (Schwamb et al., 2012).

CD69 has previously been identified as a surrogate marker for *IGH* mutation status, with expression higher in UM-CLL patients compared to M-CLL cases (Olsson et al., 2008, Del Poeta et al., 2012), whereas *ADAM29* expression has been associated with higher expression in M-CLL (Oppezzo et al., 2005, Vasconcelos et al., 2005, Cornet et al., 2015). *ADAM29* encodes the disintegrin and metalloproteinase domain-containing protein 29, which belongs to a family of a membrane and secreted glycoproteins that mediate cell-cell and cell-matrix interactions (Zhao et al., 2016). This protein has been implicated in many physiological processes, and its expression has been associated with cancer development and progression

(Zhao et al., 2016). High *ADAM29* expression has been associated with Binet stage A CLL and may predict a long treatment free interval in this subset of patients (Nückel et al., 2006).

To identify mRNAs differentially expressed based on *IGH* mutation status and partially controlled for variations in germline DNA sequence, mRNAs were compared between M-CLL and UM-CLL in familial cases that had been normalised to familial controls. There were 84 mRNAs differentially expressed between M-CLL and UM-CLL that did not differentiate both groups from controls. The majority of these genes were upregulated in UM-CLL versus M-CLL and showed statistical significance $p < 0.05$ and biological fold change > 2 in regulation (Figure 6-7). Genes that were more frequently upregulated in M-CLL cases include *KLHL14, CTLA4, SESN3, WFDC21P*, and *PMAIP1;* whereas *IGHD3-22, FOSB, SNORD14E, EGR1* and *TNF* were more frequently upregulated in UM-CLL case.

KLHL14 is a member of the klech-like (KLHL) gene family that encodes a group of proteins that generally consist of a BTB (BR-C, ttk and bab) or POZ (Pox virus and Zinc finger) domain, a BACK (BTB and C-terminal Kelch) domain, and several kelch motifs (Dhanoa et al., 2013). Members of the KLHL family have previously been associated with cancer and mutations in *KLHL6,* which is involved in the formation of the germinal centre during B cell maturation, have been identified in B-CLL patients (Kroll et al., 2005). *KLHL6* has also been identified as a target of somatic hypermutation in mutated CLL patients (Puente et al., 2011). *KLHL14* is over-expressed in B lymphocytes (http://www.proteomicsdb.org), however the finding reported in the present study is the first report of an association between *KLHL14* and M-CLL.

The cytotoxic T-lymphocyte antigen 4 (CTLA4) was also overexpressed in M-CLL cases. CTLA-4 acts as a negative regulator of T cell activation by interacting with B7 ligands

(CD80/CD86) on antigen presenting cells to inhibit cell proliferation, cytokines production, and cell cycle progression (Ciszak et al., 2016b). It has been shown that higher expression of CTLA-4 is associated with a good prognosis and lower expression with a shorter time to treatment and poor prognosis (Joshi et al., 2007, Ciszak et al., 2016a). Furthermore, downregulation of CTLA4 *in vitro* led to a significant increase in the proliferation and survival of CLL cells (Mittal et al., 2013).

In addition to IGH mutation status, *IGH* gene usage further refines prognosis. One of the most frequently used genes is *IGHV4-34*, which is associated with M-CLL and an indolent course (Murray et al., 2008, Xochelli et al., 2017). Although *IGHV4-34* was not found to be over-represented in the M-CLL cases in this present study, in agreement with previous reports, *IGHD3-22* was more frequent in M-CLL compared to UM-CLL (Xochelli et al., 2017).

The present study found tumour necrosis factor (TNF) was overexpressed in UM-CLL. TNF is a multifunctional cytokine that plays important roles in several cellular process including cell proliferation, differentiation and apoptosis (Wang and Lin, 2008). It has been reported that TNF has an essential role in the activation, growth and apoptosis of malignant lymphocytes in B-CLL cells. A higher concentration of plasma TNF in patients with B-CLL has been associated with more aggressive disease, suggesting a role for TNF in B-CLL progression (Ferrajoli et al., 2002, Bojarska-Junak et al., 2008).

Compared to mRNA studies, protein expression studies have found higher percentages of differentially expressed proteins in M-CLL and UM-CLL (Eagle et al., 2015, Perrot et al., 2011, Huang et al., 2016), however these studies have not normalized protein abundance in CLL cases to control samples, or have been enriched with familial cases. In the study reported in this Chapter, MS analysis identified 5100 proteins which were normalized to the 6 control

samples. Twelve proteins were differentially expressed between M-CLL and UM-CLL cases normalized to controls, however at an individual case level, these proteins did not completely segregate M-CLL from UM-CLL, with one UM-CLL case segregating with M-CLL cases. Of the 12 differentially expressed proteins, 7 were under-expressed in M-CLL compared to UM-CLL (MAGED4, ITPKC, HBD, PMVK, ALDH1A1, EPB41 and GYPA), whilst 5 were overexpressed in M-CLL compared to UM-CLL (PRB1/PRB2, MRPL38, CRIP1, RPS19 and ARGLU1). One of these proteins, HBD (hemoglobin subunit delta) have been previously reported to be differentially expressed between M-CLL and UM-CLL cases (Eagle et al., 2015, Barnidge et al., 2005), and 2 proteins (MRPL38 and CRIP1) have been associated with other malignancies.

Mitochondrial ribosomal protein L38 (MRPL38) abundance was significantly higher in M-CLL compared to UM-CLL cases. Overexpression of *MRPL38* gene, which is involved in protein synthesis, has been reported to be > 4-fold higher in precursor T-cell lymphoblastic leukemia/lymphoma (pre-T LBL) compared to normal thymus (Lin and Aplan, 2007), and may be a potential target for treatment of pre-T LBL. The MRPL38 protein, along with other mitochondrial ribosomal proteins, has also been detected in a human lymphoma B-cell line (Diez et al., 2015).

Cysteine-rich protein 1 (CRIP1) was also found at higher levels in M-CLL and UM-CLL. This protein and other family members interact with the actin-binding domain of α-actinin to directly bundle actin microfilaments (Tran et al., 2005). The expression of CRIP1 has been reported in several tumour types including breast, cervical, prostate, pancreatic, and colorectal cancers (Ma et al., 2003, Ludyga et al., 2013, Chen et al., 2003, Wang et al., 2007, Terris et al., 2002, Groene et al., 2006).

In conclusion, although this family study of mRNA and protein levels of M-CLL and UM-

CLL reported in this Chapter was limited by low subject numbers, the results have validated findings from previous studies and identified a number of new genes and proteins that may be used for prognostication or as targets for novel therapies.

CHAPTER-7: DISCUSSION AND FINAL COMMENTS

Genetic linkage, candidate gene and genome-wide association studies have identified a large number of mutations that may be associated with CLL. However, these mutations account for only ≈19% of familial risk for developing CLL (Law et al., 2017). Of note, GWAS haplotypes are enriched in regulatory elements including key B-cell transcription binding factor sites (Law et al., 2017). Therefore, it is likely that a proportion of genetic susceptibility to CLL results from mutations that affect gene regulation, including transcription factor binding and epigenetic modification, rather than changes in DNA sequences that affect protein function. To identify differences in gene regulation, the present study used high-resolution expression profiling microarrays and mass spectrometry to identify differentially expressed genes and proteins in purified normal, premalignant and malignant B lymphocytes. To partially normalise for inter-individual variation in gene expression, a family with multiple cases of CLL and MBL was studied. For this family, my supervisor, A/Professor Fuller, had previously reported a genetic linkage scan, which showed weak linkage to 14q24.1 and 14q31.2 (non-parametric linkage statistic = 2.24; p = 0.03). (Fuller et al., 2008).

The first aims of this study were to determine if family CLL/MBL (familial lymphoproliferative disease; F-LPD) B lymphocytes contain unique mRNA and protein profiles compared to B lymphocytes from unaffected family subjects and sporadic CLL (S-CLL) cases. The mRNA profiles between F-LPD and S-CLL groups were performed after removing genes that were differentially expressed as a result of genetic relatedness within the F-LPD group. The first hypothesis was that inherited mutations or epimutations affecting the expression of imprinted genes can be inferred by differences in mRNA levels in controls, F-CLL cases, F-MBL, and S-CLL. The 2nd hypothesis was that Identification of differentially abundant proteins extracted from CLL cells from F-LPD and S-CLL will provide protein profiles that can be used as disease signatures and will identify proteins that may act as ‘cancer drivers’ for different

subtypes of CLL.

The results reported in Chapter 3 show that gene expression profiles in B lymphocytes from F-CLL and F-MBL, combined as F-LPD cases, were unique compared to B lymphocytes from unaffected subjects and sporadic CLL cases. Furthermore, gene expression profiles segregated family normal controls and unrelated controls. To identify genes differentially expressed between S-CLL and F-LPD, we removed genes that were similarly expressed for normal familial and F-LPD. The resultant panel of differentially expressed genes were then studied at an individual level using semi-supervised hierarchical clustering to determine if they segregated normal B-lymphocytes from F-LPD and S-CLL. Twelve genes and 1 microRNA (miRNA) showed a >2-fold change in regulation. The cluster dendrogram segregated healthy controls, F-LPD, and S-CLL. The 2nd aim of this study was to use a combination of quadrupole, ion trap and Orbitrap mass spectrometer analysis, and unsupervised hierarchical clustering, to determine if F-LPD B lymphocytes contain unique protein signatures compared to B lymphocytes from controls and S-CLL cases. It was hypothesised that proteins extracted from CLL cells would identify protein profiles unique to F-LPD and S-CLL and identify proteins that may act as 'cancer drivers'. Analysis of protein levels using advanced proteomic techniques identified 4672 proteins that after normalisation to 6 control samples, segregated F-LPD, S-CLL and healthy control groups using unsupervised hierarchical clustering based on protein level patterns. Assessment for differences in normalised protein abundance between F-LPD and S-CLL patients was performed using independent two-sample t-tests. A total of 30 proteins were identified with log fold changes > 2 and t-test p-values < 0.05. Of these proteins, 22 were overexpressed, and 8 proteins were underexpressed in F-LPD compared to S-CLL. A number of proteins were identified that are involved in the regulation of various cellular processes such as cell cycle progression and cellular proliferation. These proteins were

localized to the cytoplasm, nucleus, plasma membrane, and extracellular space, and have different molecular functions such as cytokine and kinase activities, transmembrane receptors and transcription regulator. The identification of differentially abundant proteins for F-CLL versus S-CLL provide an understanding of the mechanism of transformation, new markers for early diagnosis, and potential targets for novel therapies.

The third aim was to investigate whether changes in mRNA profiles were associated with progression from normal B-lymphocytes through pre-malignant MBL cells to malignant CLL. Associations between mRNA levels and 4 genes known to be involved in either the development or progression of CLL were found, and 7 novel associations were identified for further studies.

The fourth aim was to study and compare gene and protein profiles in M-CLL and UM-CLL cases using high-resolution DNA microarrays and MS. In M-CLL and UM-CLL cases, gene and protein expression did not segregate the two subgroups; a finding consistent with previous publications, suggesting that mutated and unmutated CLL have common patterns of expression. Although limited by small numbers, specifically the presence of only 1 UM-CLL case, the results presented in this thesis suggest that studying M-CLL and UM-CLL mRNA expression within families may be a useful strategy to identify novel candidate genes and further characterize the cell of origin for each subset.

IN SUMMARY

Overall, the results presented in this PhD thesis identified significant candidate susceptibility genes using unbiased screening of mRNA and protein profiles in a family with multiple members affected by MBL or CLL. This family had previously been the subject of a genetic linkage study that had not found evidence for a single gene model of disease

susceptibility. Therefore, the presence of multiple affected family members, across numerous generations, suggested genes may become dysregulated as a result of inherited mutations in gene regulatory regions or epigenetic changes. Significantly, it was found that the same mRNAs and proteins were dysregulated to differing degrees in S-CLL, suggesting that studies of other F-CLL families may identify genes that contribute to the development of S-CLL.

FUTURE DIRECTIONS

Identifying dysregulated genes and proteins in F-CLL, S-CLL, MBL, and M-CLL and UM-CLL is important for further studies that will study early preventive interventions and develop targeted therapies. In this study it was hypothesised that family CLL/MBL (familial lymphoproliferative disease; F-LPD) may contain unique mRNA and/or protein profiles compared to B lymphocytes from unaffected family subjects and sporadic CLL (S-CLL) cases. It was found that mRNA and protein profiles clearly segregated clonal B lymphocytes in S-CLL from clonal B lymphocytes in F-MBL and F-CLL (combined as familial-lymphoproliferative disease; F-LPD). These profiles were distinct from those found in normal B lymphocytes in unaffected family members and unrelated controls. Furthermore, increasing upregulation or downregulation of both F-LPD specific genes and genes common to S-CLL occurred in association with progression from normal familial B lymphocytes through F-MBL to F-CLL. Thus, in the future it would be important to explore the roles of identified candidate genes and proteins in CLL development. Furthermore, it is planned to study the family presented in this study and other F-CLL families using chromatin immunoprecipitation to identify differential gene expression that results from transcriptional regulatory mechanisms and chromatin modifications.

This study has developed an extensive database and understanding of F-CLL and S-CLL mRNA profiles and proteomes, and gene regulation associated with M-CLL and UM-CLL and progression of MBL to CLL. However, there were limitations in this study which should be considered when designing future experiments. One common limitation for familial studies is the small size of families with multiple MBL and CLL cases. A solution is to screen other CLL families using a combination of mRNA and protein profiling, which would potentially identify additional candidate genes associated with the pathogenesis of S-CLL and F-CLL. Another limitation is that very low amounts of B cells from healthy controls and from early diagnosis B lymphocyte clones is a critical problem for proteomics.

One of our next aims is to study large numbers of well-characterized MBL and CLL cases using the same approaches in order to better understand the role of expressed genes and proteins in the pathogenesis of B-CLL. The data generated by IPA could be further analysed in a large number of samples to generate more information about the biological processes of the inherited dysregulation of these genes and how they contribute to driving malignant progression of F-LPD and S-CLL.

CHAPTER-8: REFERENCES

ABOUSAMRA, N. K., SALAH EL-DIN, M., HAMZA ELZAHAF, E. & ESMAEL, M. E. 2015. Ectonucleoside triphosphate diphosphohydrolase-1 (E-NTPDase1/CD39) as a new prognostic marker in chronic lymphocytic leukemia. *Leuk Lymphoma,* 56**,** 113-9.

ABU-AMERO, S., MONK, D., FROST, J., PREECE, M., STANIER, P. & MOORE, G. E. 2008. The genetic aetiology of Silver-Russell syndrome. *J Med Genet,* 45**,** 193-9.

AGHEBATI-MALEKI, L., SHABANI, M., BARADARAN, B., MOTALLEBNEZHAD, M., MAJIDI, J. & YOUSEFI, M. 2017. Receptor tyrosine kinase-like orphan receptor 1 (ROR-1): An emerging target for diagnosis and therapy of chronic lymphocytic leukemia. *Biomedicine & Pharmacotherapy,* 88**,** 814-822.

ALBRIGHT, F., TEERLINK, C., WERNER, T. L. & CANNON-ALBRIGHT, L. A. 2012. Significant evidence for a heritable contribution to cancer predisposition: a review of cancer familiality by site. *BMC Cancer,* 12**,** 138.

ALSAGABY, S. A., KHANNA, S., HART, K. W., PRATT, G., FEGAN, C., PEPPER, C., BREWIS, I. A. & BRENNAN, P. 2014. Proteomics-Based Strategies To Identify Proteins Relevant to Chronic Lymphocytic Leukemia. *J Proteome Res*.

ARBER, D. A., ORAZI, A., HASSERJIAN, R., THIELE, J., BOROWITZ, M. J., LE BEAU, M. M., BLOOMFIELD, C. D., CAZZOLA, M. & VARDIMAN, J. W. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,* 127**,** 2391-405.

ARRUGA, F., GIZDIC, B., SERRA, S., VAISITTI, T., CIARDULLO, C., COSCIA, M., LAURENTI, L., D'ARENA, G., JAKSIC, O., INGHIRAMI, G., ROSSI, D., GAIDANO, G. & DEAGLIO, S. 2014. Functional impact of NOTCH1 mutations in chronic lymphocytic leukemia. *Leukemia,* 28**,** 1060-1070.

ARTANDI, S. E. & ATTARDI, L. D. 2005. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochem Biophys Res Commun,* 331**,** 881-90.

ARTAVANIS-TSAKONAS, S., MATSUNO, K. & FORTINI, M. E. 1995. Notch signaling. *Science,* 268**,** 225-32.

ATHMA, P., RAPPAPORT, R. & SWIFT, M. 1996. Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. *Cancer Genet Cytogenet,* 92**,** 130-4.

ATTAR, M. A. & SANTY, L. C. 2013. The scaffolding protein GRASP/Tamalin directly binds to Dock180 as well as to cytohesins facilitating GTPase crosstalk in epithelial cell migration. *BMC cell biology,* 14**,** 9.

AU, W. Y., FUNG, A., WONG, K. F., CHAN, C. H. & LIANG, R. 2189. Tumor necrosis factor alpha promoter polymorphism and the risk of chronic lymphocytic leukemia and myeloma in the Chinese population. *Leuk Lymphoma,* 47**,** 2189-93.

AUER, R. L., DIGHIERO, G., GOLDIN, L. R., SYNDERCOMBE-COURT, D., JONES, C., MCELWAINE, S., NEWLAND, A. C., FEGAN, C. D., CAPORASO, N. & COTTER, F. E. 2007. Trinucleotide repeat dynamic mutation identifying susceptibility in familial and sporadic chronic lymphocytic leukaemia. *Br J Haematol,* 136**,** 73-9.

AUER, R. L., JONES, C., MULLENBACH, R. A., SYNDERCOMBE-COURT, D., MILLIGAN, D. W., FEGAN, C. D. & COTTER, F. E. 2001. Role for CCG-trinucleotide repeats in the pathogenesis of chronic lymphocytic leukemia. *Blood,* 97**,** 509-15.

AUGEREAU, P., BADIA, E., CARASCOSSA, S., CASTET, A., FRITSCH, S., HARMAND, P.-O., JALAGUIER, S. & CAVAILLÈS, V. 2006. The nuclear receptor transcriptional coregulator RIP140. *Nuclear receptor signaling,* 4.

AURELIUS, J., THOREN, F. B., AKHIANI, A. A., BRUNE, M., PALMQVIST, L., HANSSON, M., HELLSTRAND, K. & MARTNER, A. 2012. Monocytic AML cells inactivate antileukemic

lymphocytes: role of NADPH oxidase/gp91(phox) expression and the PARP-1/PAR pathway of apoptosis. *Blood,* 119**,** 5832-7.

AUSTEN, B., POWELL, J. E., ALVI, A., EDWARDS, I., HOOPER, L., STARCZYNSKI, J., TAYLOR, A. M., FEGAN, C., MOSS, P. & STANKOVIC, T. 2005. Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood,* 106**,** 3175-82.

AUSTEN, B., SKOWRONSKA, A., BAKER, C., POWELL, J. E., GARDINER, A., OSCIER, D., MAJID, A., DYER, M., SIEBERT, R., TAYLOR, A. M., MOSS, P. A. & STANKOVIC, T. 2007. Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. *J Clin Oncol,* 25**,** 5448-57.

BAIR, E. 2013. Semi-supervised clustering methods. *Wiley Interdiscip Rev Comput Stat,* 5**,** 349-361.

BAIR, E. & TIBSHIRANI, R. 2004. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol,* 2**,** E108.

BAOU, M., JEWELL, A., MUTHURANIA, A., WICKREMASINGHE, R. G., YONG, K. L., CARR, R., MARSH, P. & MURPHY, J. J. 2009. Involvement of Tis11b, an AU-rich binding protein, in induction of apoptosis by rituximab in B cell chronic lymphocytic leukemia cells. *Leukemia,* 23**,** 986-9.

BARNIDGE, D. R., JELINEK, D. F., MUDDIMAN, D. C. & KAY, N. E. 2005. Quantitative protein expression analysis of CLL B cells from mutated and unmutated IgVH subgroups using acid-cleavable isotope-coded affinity tag reagents. *Journal of proteome research,* 4**,** 1310-1317.

BASKAR, S., KWONG, K. Y., HOFER, T., LEVY, J. M., KENNEDY, M. G., LEE, E., STAUDT, L. M., WILSON, W. H., WIESTNER, A. & RADER, C. 2008. Unique cell surface expression of receptor tyrosine kinase ROR1 in human B-cell chronic lymphocytic leukemia. *Clin Cancer Res,* 14**,** 396-404.

BAUMANN, T., DELGADO, J., SANTACRUZ, R., MARTINEZ-TRILLOS, A., ROZMAN, M., AYMERICH, M., LOPEZ, C., COSTA, D., CARRIO, A., VILLAMOR, N. & MONTSERRAT, E. 2016. CD49d (ITGA4) expression is a predictor of time to first treatment in patients with chronic lymphocytic leukaemia and mutated IGHV status. *Br J Haematol,* 172**,** 48-55.

BEGLEITER, A., HEWITT, D., GIBSON, S. B. & JOHNSTON, J. B. 2009. Investigation of an NQO1 polymorphism as a possible risk and prognostic factor for chronic lymphocytic leukemia. *Leuk Res,* 33**,** 74-81.

BERCHUCK, A., HERON, K. A., CARNEY, M. E., LANCASTER, J. M., FRASER, E. G., VINSON, V. L., DEFFENBAUGH, A. M., MIRON, A., MARKS, J. R., FUTREAL, P. A. & FRANK, T. S. 1998. Frequency of germline and somatic BRCA1 mutations in ovarian cancer. *Clin Cancer Res,* 4**,** 2433-7.

BERNDT, S. I., SKIBOLA, C. F., JOSEPH, V., CAMP, N. J., NIETERS, A., WANG, Z., COZEN, W., MONNEREAU, A., WANG, S. S., KELLY, R. S., LAN, Q., TERAS, L. R., CHATTERJEE, N., CHUNG, C. C., YEAGER, M., BROOKS-WILSON, A. R., HARTGE, P., PURDUE, M. P., BIRMANN, B. M., ARMSTRONG, B. K., COCCO, P., ZHANG, Y., SEVERI, G., ZELENIUCH-JACQUOTTE, A., LAWRENCE, C., BURDETTE, L., YUENGER, J., HUTCHINSON, A., JACOBS, K. B., CALL, T. G., SHANAFELT, T. D., NOVAK, A. J., KAY, N. E., LIEBOW, M., WANG, A. H., SMEDBY, K. E., ADAMI, H. O., MELBYE, M., GLIMELIUS, B., CHANG, E. T., GLENN, M., CURTIN, K., CANNON-ALBRIGHT, L. A., JONES, B., DIVER, W. R., LINK, B. K., WEINER, G. J., CONDE, L., BRACCI, P. M., RIBY, J., HOLLY, E. A., SMITH, M. T.,

JACKSON, R. D., TINKER, L. F., BENAVENTE, Y., BECKER, N., BOFFETTA, P., BRENNAN, P., FORETOVA, L., MAYNADIE, M., MCKAY, J., STAINES, A., RABE, K. G., ACHENBACH, S. J., VACHON, C. M., GOLDIN, L. R., STROM, S. S., LANASA, M. C., SPECTOR, L. G., LEIS, J. F., CUNNINGHAM, J. M., WEINBERG, J. B., MORRISON, V. A., CAPORASO, N. E., NORMAN, A. D., LINET, M. S., DE ROOS, A. J., MORTON, L. M., SEVERSON, R. K., RIBOLI, E., VINEIS, P., KAAKS, R., TRICHOPOULOS, D., MASALA, G., WEIDERPASS, E., CHIRLAQUE, M. D., VERMEULEN, R. C., TRAVIS, R. C., GILES, G. G., ALBANES, D., VIRTAMO, J., WEINSTEIN, S., CLAVEL, J., ZHENG, T., HOLFORD, T. R., OFFIT, K., ZELENETZ, A., KLEIN, R. J., SPINELLI, J. J., BERTRAND, K. A., et al. 2013. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet,* 45**,** 868-76.

BERNHAGEN, J., KROHN, R., LUE, H., GREGORY, J. L., ZERNECKE, A., KOENEN, R. R., DEWOR, M., GEORGIEV, I., SCHOBER, A., LENG, L., KOOISTRA, T., FINGERLE-ROWSON, G., GHEZZI, P., KLEEMANN, R., MCCOLL, S. R., BUCALA, R., HICKEY, M. J. & WEBER, C. 2007. MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nat Med,* 13**,** 587-96.

BEVAN, S., CATOVSKY, D., MAROSSY, A., MATUTES, E., POPAT, S., ANTONOVIC, P., BELL, A., BERREBI, A., GAMINARA, E., QUABECK, K., RIBEIRO, I., MAURO, F. R., STARK, P., SYKES, H., VAN DONGEN, J., WIMPERIS, J., WRIGHT, S., YUILLE, M. R. & HOULSTON, R. S. 1999. Linkage analysis for ATM in familial B cell chronic lymphocytic leukaemia. *Leukemia,* 13**,** 1497-500.

BEVAN, S., CATOVSKY, D., MATUTES, E., ANTUNOVIC, P., AUGER, M. J., BEN-BASSAT, I., BELL, A., BERREBI, A., GAMINARA, E. J., JUNIOR, M. E., MAURO, F. R., QUABECK, K., RASSAM, S. M., REID, C., RIBEIRO, I., STARK, P., VAN DONGEN, J. J., WIMPERIS, J., WRIGHT, S., MAROSSY, A., YUILLE, M. R. & HOULSTON, R. S. 2000. Linkage analysis for major histocompatibility complex-related genetic susceptibility in familial chronic lymphocytic leukemia. *Blood,* 96**,** 3982-4.

BICHI, R., SHINTON, S. A., MARTIN, E. S., KOVAL, A., CALIN, G. A., CESARI, R., RUSSO, G., HARDY, R. R. & CROCE, C. M. 2002. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. *Proc Natl Acad Sci U S A,* 99**,** 6955-60.

BINSKY, I., HARAN, M., STARLETS, D., GORE, Y., LANTNER, F., HARPAZ, N., LENG, L., GOLDENBERG, D. M., SHVIDEL, L. & BERREBI, A. 2007. IL-8 secreted in a macrophage migration-inhibitory factor-and CD74-dependent manner regulates B cell chronic lymphocytic leukemia survival. *Proceedings of the National Academy of Sciences,* 104**,** 13408-13413.

BINSKY, I., LANTNER, F., GRABOVSKY, V., HARPAZ, N., SHVIDEL, L., BERREBI, A., GOLDENBERG, D. M., LENG, L., BUCALA, R., ALON, R., HARAN, M. & SHACHAR, I. 2010. TAp63 regulates VLA-4 expression and chronic lymphocytic leukemia cell migration to the bone marrow in a CD74-dependent manner. *J Immunol,* 184**,** 4761-9.

BLOCH, D. B., DE LA MONTE, S. M., GUIGAOURI, P., FILIPPOV, A. & BLOCH, K. D. 1996. Identification and characterization of a leukocyte-specific component of the nuclear body. *J Biol Chem,* 271**,** 29198-204.

BLUNT, M. D., CARTER, M. J., LARRAYOZ, M., SMITH, L. D., AGUILAR-HERNANDEZ, M., COX, K. L., TIPTON, T., REYNOLDS, M., MURPHY, S., LEMM, E., DIAS, S., DUNCOMBE, A., STREFFORD, J. C., JOHNSON, P. W., FORCONI, F., STEVENSON, F. K., PACKHAM, G., CRAGG, M. S. & STEELE, A. J. 2015. The PI3K/mTOR inhibitor PF-04691502 induces

apoptosis and inhibits microenvironmental signaling in CLL and the Emicro-TCL1 mouse model. *Blood,* 125**,** 4032-41.

BOEHLER, C., GAUTHIER, L. R., MORTUSEWICZ, O., BIARD, D. S., SALIOU, J. M., BRESSON, A., SANGLIER-CIANFERANI, S., SMITH, S., SCHREIBER, V., BOUSSIN, F. & DANTZER, F. 2011. Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression. *Proc Natl Acad Sci U S A,* 108**,** 2783-8.

BOGUNIA-KUBIK, K., MAZUR, G., URBANOWICZ, I., WROBEL, T., KULICZKOWSKI, K., WOZNIAK, M. & LANGE, A. 2006. Lack of association between the TNF-alpha promoter gene polymorphism and susceptibility to B-cell chronic lymphocytic leukaemia. *Int J Immunogenet,* 33**,** 21-4.

BOJARSKA-JUNAK, A., HUS, I., SZCZEPANEK, E. W., DMOSZYNSKA, A. & ROLINSKI, J. 2008. Peripheral blood and bone marrow TNF and TNF receptors in early and advanced stages of B-CLL in correlation with ZAP-70 protein and CD38 antigen. *Leuk Res,* 32**,** 225-33.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

BORECKI, I. B. & PROVINCE, M. A. 2008. Genetic and genomic discovery using family studies. *Circulation,* 118**,** 1057-63.

BRODERICK, P., SELLICK, G., FIELDING, S., CATOVSKY, D. & HOULSTON, R. 2008. Lack of a relationship between the common 18q24 variant rs12953717 and risk of chronic lymphocytic leukemia. *Leuk Lymphoma,* 49**,** 271-2.

BROK-SIMONI, F., RECHAVI, G., KATZIR, N. & BEN-BASSAT, I. 1987. Chronic lymphocytic leukaemia in twin sisters: monozygous but not identical. *Lancet,* 1**,** 329-30.

BROWN, J. R., HANNA, M., TESAR, B., POCHET, N., VARTANOV, A., FERNANDES, S. M., WERNER, L., ASH, M., RODEN, C. A., MACCONAILL, L., HAINZ, U., LONGTINE, J., WANG, Y. E., CORRELL, M., VAN DE PEER, Y., REGEV, A., WU, C., NEUBERG, D. & FREEDMAN, A. S. 2012. Germline copy number variation associated with Mendelian inheritance of CLL in two families. *Leukemia,* 26**,** 1710-3.

BULLMAN, H., LEVER, M., ROBINSON, D. O., MACKAY, D. J., HOLDER, S. E. & WAKELING, E. L. 2008. Mosaic maternal uniparental disomy of chromosome 11 in a patient with Silver-Russell syndrome. *J Med Genet,* 45**,** 396-9.

BULLRICH, F., MACLACHLAN, T. K., SANG, N., DRUCK, T., VERONESE, M. L., ALLEN, S. L., CHIORAZZI, N., KOFF, A., HEUBNER, K., CROCE, C. M. & ET AL. 1995. Chromosomal mapping of members of the cdc2 family of protein kinases, cdk3, cdk6, PISSLRE, and PITALRE, and a cdk inhibitor, p27Kip1, to regions involved in human cancer. *Cancer Res,* 55**,** 1199-205.

BUSZEWSKI, B. & NOGA, S. 2012. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Analytical and Bioanalytical Chemistry,* 402**,** 231-247.

CABRINI, G., FALZONI, S., FORCHAP, S. L., PELLEGATTI, P., BALBONI, A., AGOSTINI, P., CUNEO, A., CASTOLDI, G., BARICORDI, O. R. & DI VIRGILIO, F. 2005. A His-155 to Tyr polymorphism confers gain-of-function to the human P2X7 receptor of human leukemic lymphocytes. *J Immunol,* 175**,** 82-9.

CALANDRA, T. & ROGER, T. 2003. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol,* 3**,** 791-800.

CALIN, G. A., TRAPASSO, F., SHIMIZU, M., DUMITRU, C. D., YENDAMURI, S., GODWIN, A. K., FERRACIN, M., BERNARDI, G., CHATTERJEE, D., BALDASSARRE, G., RATTAN, S., ALDER, H., MABUCHI, H., SHIRAISHI, T., HANSEN, L. L., OVERGAARD, J., HERLEA, V., MAURO,

F. R., DIGHIERO, G., MOVSAS, B., RASSENTI, L., KIPPS, T., BAFFA, R., FUSCO, A., MORI, M., RUSSO, G., LIU, C. G., NEUBERG, D., BULLRICH, F., NEGRINI, M. & CROCE, C. M. 2005. Familial cancer associated with a polymorphism in ARLTS1. *N Engl J Med,* 352**,** 1667-76.

CAMPREGHER, P. V. & HAMERSCHLAK, N. 2014. Novel prognostic gene mutations identified in chronic lymphocytic leukemia and their impact on clinical practice. *Clin Lymphoma Myeloma Leuk,* 14**,** 271-6.

CARGNELLO, M. & ROUX, P. P. 2011. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiology and molecular biology reviews,* 75**,** 50-83.

CARTER, A., LIN, K., SHERRINGTON, P. D., ATHERTON, M., PEARSON, K., DOUGLAS, A., BURFORD, A., BRITO-BABAPULLE, V., MATUTES, E., CATOVSKY, D. & PETTITT, A. R. 2006. Imperfect correlation between p53 dysfunction and deletion of TP53 and ATM in chronic lymphocytic leukaemia. *Leukemia,* 20**,** 737-40.

CATOVSKY, D., RICHARDS, S., MATUTES, E., OSCIER, D., DYER, M. J., BEZARES, R. F., PETTITT, A. R., HAMBLIN, T., MILLIGAN, D. W., CHILD, J. A., HAMILTON, M. S., DEARDEN, C. E., SMITH, A. G., BOSANQUET, A. G., DAVIS, Z., BRITO-BABAPULLE, V., ELSE, M., WADE, R., HILLMEN, P., GROUP, U. K. N. C. R. I. H. O. C. S. & GROUP, N. C. L. L. W. 2007. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet,* 370**,** 230-9.

CAVAZZINI, F., HERNANDEZ, J. A., GOZZETTI, A., RUSSO ROSSI, A., DE ANGELI, C., TISEO, R., BARDI, A., TAMMISO, E., CRUPI, R. & LENOCI, M. P. 2008. Chromosome 14q32 translocations involving the immunoglobulin heavy chain locus in chronic lymphocytic leukaemia identify a disease subset with poor prognosis. *British journal of haematology,* 142**,** 529-537.

CERHAN, J. R., ANSELL, S. M., FREDERICKSEN, Z. S., KAY, N. E., LIEBOW, M., CALL, T. G., DOGAN, A., CUNNINGHAM, J. M., WANG, A. H., LIU-MARES, W., MACON, W. R., JELINEK, D., WITZIG, T. E., HABERMANN, T. M. & SLAGER, S. L. 2007. Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood,* 110**,** 4455-63.

CHAHROUR, O., COBICE, D. & MALONE, J. 2015. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *Journal of pharmaceutical and biomedical analysis,* 113**,** 2-20.

CHANG, E. T., SMEDBY, K. E., HJALGRIM, H., PORWIT-MACDONALD, A., ROOS, G., GLIMELIUS, B. & ADAMI, H. O. 2005. Family history of hematopoietic malignancy and risk of lymphoma. *J Natl Cancer Inst,* 97**,** 1466-74.

CHANG, L. & KARIN, M. 2001. Mammalian MAP kinase signalling cascades. *Nature,* 410**,** 37-40.

CHAPUIS, N., TAMBURINI, J., GREEN, A. S., WILLEMS, L., BARDET, V., PARK, S., LACOMBE, C., MAYEUX, P. & BOUSCARY, D. 2010. Perspectives on inhibiting mTOR as a future treatment strategy for hematological malignancies. *Leukemia,* 24**,** 1686.

CHEN, G. & PRAMANIK, B. N. 2009. Application of LC/MS to proteomics studies: current status and future prospects. *Drug discovery today,* 14**,** 465-471.

CHEN, X., WEI, S., JI, Y., GUO, X. & YANG, F. 2015. Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics,* 15**,** 3175-3192.

CHEN, Y., MILLER, C., MOSHER, R., ZHAO, X., DEEDS, J., MORRISSEY, M., BRYANT, B., YANG, D., MEYER, R., CRONIN, F., GOSTOUT, B. S., SMITH-MCCUNE, K. & SCHLEGEL, R. 2003.

Identification of cervical cancer markers by cDNA and tissue microarrays. *Cancer Res,* 63**,** 1927-35.

CHIARETTI, S., MARINELLI, M., DEL GIUDICE, I., BONINA, S., PICIOCCHI, A., MESSINA, M., VIGNETTI, M., ROSSI, D., DI MAIO, V., MAURO, F. R., GUARINI, A., GAIDANO, G. & FOA, R. 2014. NOTCH1, SF3B1, BIRC3 and TP53 mutations in patients with chronic lymphocytic leukemia undergoing first-line treatment: correlation with biological parameters and response to treatment. *Leuk Lymphoma,* 55**,** 2785-92.

CHIORAZZI, N. 2012. Implications of new prognostic markers in chronic lymphocytic leukemia. *Hematology,* 2012**,** 76-87.

CHIORAZZI, N. & FERRARINI, M. 2003. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu Rev Immunol,* 21**,** 841-94.

CHIORAZZI, N., RAI, K. R. & FERRARINI, M. 2005. Chronic lymphocytic leukemia. *New England Journal of Medicine,* 352**,** 804-15.

CHIU, B. C., KOLAR, C., GAPSTUR, S. M., LAWSON, T., ANDERSON, J. R. & WEISENBURGER, D. D. 2005. Association of NAT and GST polymorphisms with non-Hodgkin's lymphoma: a population-based case-control study. *Br J Haematol,* 128**,** 610-5.

CHOUDHURY, A., DERKOW, K., DANESHMANESH, A. H., MIKAELSSON, E., KIAII, S., KOKHAEI, P., OSTERBORG, A. & MELLSTEDT, H. 2010. Silencing of ROR1 and FMOD with siRNA results in apoptosis of CLL cells. *Br J Haematol,* 151**,** 327-35.

CISZAK, L., FRYDECKA, I., WOLOWIEC, D., SZTEBLICH, A. & KOSMACZEWSKA, A. 2016a. CTLA-4 affects expression of key cell cycle regulators of G0/G1 phase in neoplastic lymphocytes from patients with chronic lymphocytic leukaemia. *Clin Exp Med,* 16**,** 317-32.

CISZAK, L., FRYDECKA, I., WOLOWIEC, D., SZTEBLICH, A. & KOSMACZEWSKA, A. 2016b. Patients with chronic lymphocytic leukaemia (CLL) differ in the pattern of CTLA-4 expression on CLL cells: the possible implications for immunotherapy with CTLA-4 blocking antibody. *Tumour Biol,* 37**,** 4143-57.

COCHRAN, D. A., EVANS, C. A., BLINCO, D., BURTHEM, J., STEVENSON, F. K., GASKELL, S. J. & WHETTON, A. D. 2003. Proteomic analysis of chronic lymphocytic leukemia subtypes with mutated or unmutated Ig VH genes. *Molecular & Cellular Proteomics,* 2**,** 1331-1341.

CORNET, E., DEBLIQUIS, A., RIMELEN, V., CIVIC, N., DOCQUIER, M., TROUSSARD, X., DRENOU, B. & MATTHES, T. 2015. Developing Molecular Signatures for Chronic Lymphocytic Leukemia. *PLoS One,* 10**,** e0128990.

COSTINEAN, S., ZANESI, N., PEKARSKY, Y., TILI, E., VOLINIA, S., HEEREMA, N. & CROCE, C. M. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A,* 103**,** 7024-9.

CRAIG, R. & BEAVIS, R. C. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics,* 20**,** 1466-1467.

CRASSINI, K., S STEVENSON, W. & MULLIGAN, S. P. 2013. MAPK-Erk1/2 Pathway Activation and Survival Of Chronic Lymphocytic Leukemia (CLL) Cells Is Induced By In Vitro Modeling Of The Tumor Microenvironment and Can Be Effectively Inhibited Using The MEK1 Inhibitor MEK162. Am Soc Hematology.

CROWTHER-SWANEPOEL, D., BRODERICK, P., DI BERNARDO, M. C., DOBBINS, S. E., TORRES, M., MANSOURI, M., RUIZ-PONTE, C., ENJUANES, A., ROSENQUIST, R., CARRACEDO, A., JURLANDER, J., CAMPO, E., JULIUSSON, G., MONTSERRAT, E., SMEDBY, K. E., DYER, M. J. S., MATUTES, E., DEARDEN, C., SUNTER, N. J., HALL, A. G., MAINOU-FOWLER, T., JACKSON, G. H., SUMMERFIELD, G., HARRIS, R. J., PETTITT, A. R., ALLSUP, D. J., BAILEY,

J. R., PRATT, G., PEPPER, C., FEGAN, C., PARKER, A., OSCIER, D., ALLAN, J. M., CATOVSKY, D. & HOULSTON, R. S. 2010. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet,* 42**,** 132-6.

CROWTHER-SWANEPOEL, D., DI BERNARDO, M. C., JAMROZIAK, K., KARABON, L., FRYDECKA, I., DEAGLIO, S., D'ARENA, G., ROSSI, D., GAIDANO, G., OLVER, B., LLOYD, A., BRODERICK, P., LAURENTI, L., SZEMRAJ-ROGUCKA, Z., ROBAK, T., CATOVSKY, D. & HOULSTON, R. S. 2011. Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk. *Br J Haematol,* 154**,** 229-33.

CROWTHER-SWANEPOEL, D. & HOULSTON, R. S. 2010. Genetic variation and risk of chronic lymphocytic leukaemia. *Semin Cancer Biol,* 20**,** 363-9.

CROWTHER-SWANEPOEL, D., QURESHI, M., DYER, M. J., MATUTES, E., DEARDEN, C., CATOVSKY, D. & HOULSTON, R. S. 2009. Genetic variation in CXCR4 and risk of chronic lymphocytic leukemia. *Blood,* 114**,** 4843-6.

CROWTHER-SWANEPOEL, D., WILD, R., SELLICK, G., DYER, M. J., MAURO, F. R., CUTHBERT, R. J., JONSSON, V., MATUTES, E., DEARDEN, C., WILEY, J., FULLER, S., CATOVSKY, D. & HOULSTON, R. S. 2008. Insight into the pathogenesis of chronic lymphocytic leukemia (CLL) through analysis of IgVH gene usage and mutation status in familial CLL. *Blood,* 111**,** 5691-3.

CUI, B., CHEN, L., ZHANG, S., MRAZ, M., FECTEAU, J. F., YU, J., GHIA, E. M., ZHANG, L., BAO, L., RASSENTI, L. Z., MESSER, K., CALIN, G. A., CROCE, C. M. & KIPPS, T. J. 2014. MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia. *Blood,* 124**,** 546-54.

CUI, B., GHIA, E. M., CHEN, L., RASSENTI, L. Z., DEBOEVER, C., WIDHOPF, G. F., 2ND, YU, J., NEUBERG, D. S., WIERDA, W. G., RAI, K. R., KAY, N. E., BROWN, J. R., JONES, J. A., GRIBBEN, J. G., FRAZER, K. A. & KIPPS, T. J. 2016. High-level ROR1 associates with accelerated disease progression in chronic lymphocytic leukemia. *Blood,* 128**,** 2931-2940.

CUNEO, A., RIGOLIN, G. M., BIGONI, R., DE ANGELI, C., VERONESE, A., CAVAZZINI, F., BARDI, A., ROBERTI, M. G., TAMMISO, E., AGOSTINI, P., CICCONE, M., DELLA PORTA, M., TIEGHI, A., CAVAZZINI, L., NEGRINI, M. & CASTOLDI, G. 2004. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. *Leukemia,* 18**,** 476-83.

CUTTNER, J. 1992. Increased incidence of hematologic malignancies in first-degree relatives of patients with chronic lymphocytic leukemia. *Cancer Invest,* 10**,** 103-9.

CZENE, K., LICHTENSTEIN, P. & HEMMINKI, K. 2002. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer,* 99**,** 260-6.

DAMLE, R. N., BATLIWALLA, F. M., GHIOTTO, F., VALETTO, A., ALBESIANO, E., SISON, C., ALLEN, S. L., KOLITZ, J., VINCIGUERRA, V. P., KUDALKAR, P., WASIL, T., RAI, K. R., FERRARINI, M., GREGERSEN, P. K. & CHIORAZZI, N. 2004. Telomere length and telomerase activity delineate distinctive replicative features of the B-CLL subgroups defined by immunoglobulin V gene mutations. *Blood,* 103**,** 375-82.

DAMLE, R. N., WASIL, T., FAIS, F., GHIOTTO, F., VALETTO, A., ALLEN, S. L., BUCHBINDER, A., BUDMAN, D., DITTMAR, K., KOLITZ, J., LICHTMAN, S. M., SCHULMAN, P., VINCIGUERRA, V. P., RAI, K. R., FERRARINI, M. & CHIORAZZI, N. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood,* 94**,** 1840-7.

DANESHMANESH, A. H., MIKAELSSON, E., JEDDI-TEHRANI, M., BAYAT, A. A., GHODS, R., OSTADKARAMPOUR, M., AKHONDI, M., LAGERCRANTZ, S., LARSSON, C., OSTERBORG, A., SHOKRI, F., MELLSTEDT, H. & RABBANI, H. 2008. Ror1, a cell surface receptor tyrosine kinase is expressed in chronic lymphocytic leukemia and may serve as a putative target for therapy. *Int J Cancer,* 123**,** 1190-5.

DAS, I., CRAIG, C., FUNAHASHI, Y., JUNG, K. M., KIM, T. W., BYERS, R., WENG, A. P., KUTOK, J. L., ASTER, J. C. & KITAJEWSKI, J. 2004. Notch oncoproteins depend on gamma-secretase/presenilin activity for processing and function. *J Biol Chem,* 279**,** 30771-80.

DAVIES, H. D., LEUSINK, G. L., MCCONNELL, A., DEYELL, M., CASSIDY, S. B., FICK, G. H. & COPPES, M. J. 2003. Myeloid leukemia in Prader-Willi syndrome. *J Pediatr,* 142**,** 174-8.

DE ROOS, A. J., GOLD, L. S., WANG, S., HARTGE, P., CERHAN, J. R., COZEN, W., YEAGER, M., CHANOCK, S., ROTHMAN, N. & SEVERSON, R. K. 2006. Metabolic gene variants and risk of non-Hodgkin's lymphoma. *Cancer Epidemiol Biomarkers Prev,* 15**,** 1647-53.

DECKER, T., HIPP, S., RINGSHAUSEN, I., BOGNER, C., OELSNER, M., SCHNELLER, F. & PESCHEL, C. 2003. Rapamycin-induced G1 arrest in cycling B-CLL cells is associated with reduced expression of cyclin D3, cyclin E, cyclin A, and survivin. *Blood,* 101**,** 278-85.

DEISS, L. P., FEINSTEIN, E., BERISSI, H., COHEN, O. & KIMCHI, A. 1995. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes Dev,* 9**,** 15-30.

DEL GIUDICE, I., ROSSI, D., CHIARETTI, S., MARINELLI, M., TAVOLARO, S., GABRIELLI, S., LAURENTI, L., MARASCA, R., RASI, S. & FANGAZIO, M. 2012. 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,* 97**,** 437-441.

DEL POETA, G., DEL PRINCIPE, M. I., ZUCCHETTO, A., LUCIANO, F., BUCCISANO, F., ROSSI, F. M., BRUNO, A., BIAGI, A., BULIAN, P., MAURILLO, L., NERI, B., BOMBEN, R., SIMOTTI, C., COLETTA, A. M., DAL BO, M., DE FABRITIIS, P., VENDITTI, A., GATTEI, V. & AMADORI, S. 2012. CD69 is independently prognostic in chronic lymphocytic leukemia: a comprehensive clinical and biological profiling study. *Haematologica,* 97**,** 279-87.

DELGADO, J., ESPINET, B., OLIVEIRA, A. C., ABRISQUETA, P., DE LA SERNA, J., COLLADO, R., LOSCERTALES, J., LOPEZ, M., HERNANDEZ-RIVAS, J. A., FERRA, C., RAMIREZ, A., RONCERO, J. M., LOPEZ, C., AVENTIN, A., PUIGGROS, A., ABELLA, E., CARBONELL, F., COSTA, D., CARRIO, A., GONZALEZ, M., GRUPO ESPANOL DE LEUCEMIA LINFATICA, C. & GRUPO ESPANOL DE CITOGENETICA, H. 2012. Chronic lymphocytic leukaemia with 17p deletion: a retrospective analysis of prognostic factors and therapy results. *British Journal of Haematology,* 157**,** 67-74.

DEMETER, J., PORZSOLT, F., RAMISCH, S., SCHMIDT, D., SCHMID, M. & MESSER, G. 1997. Polymorphism of the tumour necrosis factor-alpha and lymphotoxin-alpha genes in chronic lymphocytic leukaemia. *Br J Haematol,* 97**,** 107-12.

DERYNCK, R., ZHANG, Y. & FENG, X. H. 1998. Smads: transcriptional activators of TGF-beta responses. *Cell,* 95**,** 737-40.

DERYNCK, R. & ZHANG, Y. E. 2003. Smad-dependent and Smad-independent pathways in TGF-β family signalling. *Nature,* 425**,** 577-584.

DHANOA, B. S., COGLIATI, T., SATISH, A. G., BRUFORD, E. A. & FRIEDMAN, J. S. 2013. Update on the Kelch-like (KLHL) gene family. *Hum Genomics,* 7**,** 13.

DI BERNARDO, M. C., BRODERICK, P., CATOVSKY, D. & HOULSTON, R. S. 2013. Common genetic variation contributes significantly to the risk of developing chronic lymphocytic leukemia. *Haematologica,* 98**,** e23-4.

DI BERNARDO, M. C., CROWTHER-SWANEPOEL, D., BRODERICK, P., WEBB, E., SELLICK, G., WILD, R., SULLIVAN, K., VIJAYAKRISHNAN, J., WANG, Y., PITTMAN, A. M., SUNTER, N. J., HALL, A. G., DYER, M. J., MATUTES, E., DEARDEN, C., MAINOU-FOWLER, T., JACKSON, G. H., SUMMERFIELD, G., HARRIS, R. J., PETTITT, A. R., HILLMEN, P., ALLSUP, D. J., BAILEY, J. R., PRATT, G., PEPPER, C., FEGAN, C., ALLAN, J. M., CATOVSKY, D. & HOULSTON, R. S. 2008. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet,* 40**,** 1204-10.

DI PALMA, S., HENNRICH, M. L., HECK, A. J. R. & MOHAMMED, S. 2012. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of Proteomics,* 75**,** 3791-3813.

DIEZ, P., DROSTE, C., DEGANO, R. M., GONZALEZ-MUNOZ, M., IBARROLA, N., PEREZ-ANDRES, M., GARIN-MUGA, A., SEGURA, V., MARKO-VARGA, G., LABAER, J., ORFAO, A., CORRALES, F. J., DE LAS RIVAS, J. & FUENTES, M. 2015. Integration of Proteomics and Transcriptomics Data Sets for the Analysis of a Lymphoma B-Cell Line in the Context of the Chromosome-Centric Human Proteome Project. *J Proteome Res,* 14**,** 3530-40.

DIEZ, P., GONGORA, R., ORFAO, A. & FUENTES, M. 2017. Functional proteomic insights in B-cell chronic lymphocytic leukemia. *Expert Rev Proteomics,* 14**,** 137-146.

DÍEZ, P., LORENZO, S., DÉGANO, R. M., IBARROLA, N., GONZÁLEZ-GONZÁLEZ, M., NIETO, W., ALMEIDA, J., GONZÁLEZ, M., ORFAO, A. & FUENTES, M. 2016. Multipronged functional proteomics approaches for global identification of altered cell signalling pathways in B-cell chronic lymphocytic leukaemia. *Proteomics,* 16**,** 1193-1203.

DIGHIERO, G. & HAMBLIN, T. J. 2008. Chronic lymphocytic leukaemia. *The Lancet,* 371**,** 1017-1029.

DOHNER, H., FISCHER, K., BENTZ, M., HANSEN, K., BENNER, A., CABOT, G., DIEHL, D., SCHLENK, R., COY, J., STILGENBAUER, S. & ET AL. 1995. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood,* 85**,** 1580-9.

DOHNER, H., STILGENBAUER, S., BENNER, A., LEUPOLT, E., KROBER, A., BULLINGER, L., DOHNER, K., BENTZ, M. & LICHTER, P. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med,* 343**,** 1910-6.

DORFER, V., PICHLER, P., STRANZL, T., STADLMANN, J., TAUS, T., WINKLER, S. & MECHTLER, K. 2014. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res,* 13**,** 3679-84.

DOUGLAS, R. S., CAPOCASALE, R. J., LAMB, R. J., NOWELL, P. C. & MOORE, J. S. 1997. Chronic lymphocytic leukemia B cells are resistant to the apoptotic effects of transforming growth factor-β. *Blood,* 89**,** 941-947.

EAGLE, G. L., ZHUANG, J., JENKINS, R. E., TILL, K. J., JITHESH, P. V., LIN, K., JOHNSON, G. G., OATES, M., PARK, K., KITTERINGHAM, N. R. & PETTITT, A. R. 2015. Total proteome analysis identifies migration defects as a major pathogenetic factor in immunoglobulin heavy chain variable region (IGHV)-unmutated chronic lymphocytic leukemia. *Mol Cell Proteomics,* 14**,** 933-45.

EFEYAN, A. & SERRANO, M. 2007. p53: guardian of the genome and policeman of the oncogenes. *Cell Cycle,* 6**,** 1006-10.

EIS, P. S., TAM, W., SUN, L., CHADBURN, A., LI, Z., GOMEZ, M. F., LUND, E. & DAHLBERG, J. E. 2005. Accumulation of miR-155 and BIC RNA in human B cell lymphomas.

Proceedings of the National Academy of Sciences of the United States of America, 102**,** 3627-3632.

ELIAS, J. E. & GYGI, S. P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods,* 4**,** 207-14.

ENG, J. K., MCCORMACK, A. L. & YATES, J. R. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom,* 5**,** 976-89.

ENJUANES, A., BENAVENTE, Y., BOSCH, F., MARTIN-GUERRERO, I., COLOMER, D., PEREZ-ALVAREZ, S., REINA, O., ARDANAZ, M. T., JARES, P., GARCIA-ORAD, A., PUJANA, M. A., MONTSERRAT, E., DE SANJOSE, S. & CAMPO, E. 2008. Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia. *Cancer Res,* 68**,** 10178-86.

ENNAS, M. G., MOORE, P. S., ZUCCA, M., ANGELUCCI, E., CABRAS, M. G., MELIS, M., GABBAS, A., SERPE, R., MADEDDU, C., SCARPA, A. & COCCO, P. 2008. Interleukin-1B (IL1B) and interleukin-6 (IL6) gene polymorphisms are associated with risk of chronic lymphocytic leukaemia. *Hematol Oncol,* 26**,** 98-103.

ERIKSSON, M. & BERGSTROM, I. 1987. Familial malignant blood disease in the country of Jamtland, Sweden. *Eur J Haematol,* 38**,** 241-5.

ESSAKALI, S., CARNEY, D., WESTERMAN, D., GAMBELL, P., SEYMOUR, J. F. & DOBROVIC, A. 2008. Negative selection of chronic lymphocytic leukaemia cells using a bifunctional rosette-based antibody cocktail. *BMC Biotechnol,* 8**,** 6.

ESTELLER, M., FRAGA, M. F., GUO, M., GARCIA-FONCILLAS, J., HEDENFALK, I., GODWIN, A. K., TROJAN, J., VAURS-BARRIERE, C., BIGNON, Y. J., RAMUS, S., BENITEZ, J., CALDES, T., AKIYAMA, Y., YUASA, Y., LAUNONEN, V., CANAL, M. J., RODRIGUEZ, R., CAPELLA, G., PEINADO, M. A., BORG, A., AALTONEN, L. A., PONDER, B. A., BAYLIN, S. B. & HERMAN, J. G. 2001. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Hum Mol Genet,* 10**,** 3001-7.

FABBRI, G. & DALLA-FAVERA, R. 2016. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nat Rev Cancer,* 16**,** 145-62.

FAIS, F., GHIOTTO, F., HASHIMOTO, S., SELLARS, B., VALETTO, A., ALLEN, S. L., SCHULMAN, P., VINCIGUERRA, V. P., RAI, K., RASSENTI, L. Z., KIPPS, T. J., DIGHIERO, G., SCHROEDER, H. W., JR., FERRARINI, M. & CHIORAZZI, N. 1998. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest,* 102**,** 1515-25.

FALISI, E., NOVELLA, E., VISCO, C., GUERCINI, N., MAURA, F., GIARETTA, I., POMPONI, F., NICHELE, I., FINOTTO, S., MONTALDI, A., NERI, A. & RODEGHIERO, F. 2014. B-cell receptor configuration and mutational analysis of patients with chronic lymphocytic leukaemia and trisomy 12 reveal recurrent molecular abnormalities. *Hematol Oncol,* 32**,** 22-30.

FARAONI, I., ANTONETTI, F. R., CARDONE, J. & BONMASSAR, E. 2009. miR-155 gene: a typical multifunctional microRNA. *Biochim Biophys Acta,* 1792**,** 497-505.

FEKRY, B., JEFFRIES, K. A., ESMAEILNIAKOOSHKGHAZI, A., OGRETMEN, B., KRUPENKO, S. A. & KRUPENKO, N. I. 2016. CerS6 is a novel transcriptional target of p53 protein activated by non-genotoxic stress. *Journal of Biological Chemistry,* 291**,** 16586-16596.

FERRAJOLI, A., KEATING, M. J., MANSHOURI, T., GILES, F. J., DEY, A., ESTROV, Z., KOLLER, C. A., KURZROCK, R., THOMAS, D. A., FADERL, S., LERNER, S., O'BRIEN, S. & ALBITAR, M. 2002. The clinical significance of tumor necrosis factor-alpha plasma level in patients having chronic lymphocytic leukemia. *Blood,* 100**,** 1215-9.

FERREIRA, P. G., JARES, P., RICO, D., GÓMEZ-LÓPEZ, G., MARTÍNEZ-TRILLOS, A., VILLAMOR, N., ECKER, S., GONZÁLEZ-PÉREZ, A., KNOWLES, D. G. & MONLONG, J. 2014. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome research,* 24**,** 212-226.

FOA, R., DEL GIUDICE, I., GUARINI, A., ROSSI, D. & GAIDANO, G. 2013. Clinical implications of the molecular genetics of chronic lymphocytic leukemia. *Haematologica,* 98**,** 675-85.

FREY, U. H., NUCKEL, H., SELLMANN, L., SIEMER, D., KUPPERS, R., DURIG, J., DUHRSEN, U. & SIFFERT, W. 2006. The GNAS1 T393C polymorphism is associated with disease progression and survival in chronic lymphocytic leukemia. *Clin Cancer Res,* 12**,** 5686-92.

FUCHS, S. Y., ADLER, V., BUSCHMANN, T., WU, X. & RONAI, Z. 1998. Mdm2 association with p53 targets its ubiquitination. *Oncogene,* 17**,** 2543-7.

FULLER, S. J., PAPAEMMANUIL, E., MCKINNON, L., WEBB, E., SELLICK, G. S., DAO-UNG, L. P., SKARRATT, K. K., CROWTHER, D., HOULSTON, R. S. & WILEY, J. S. 2008. Analysis of a large multi-generational family provides insight into the genetics of chronic lymphocytic leukemia. *Br J Haematol,* 142**,** 238-45.

GANDHIRAJAN, R. K., STAIB, P. A., MINKE, K., GEHRKE, I., PLICKERT, G., SCHLOSSER, A., SCHMITT, E. K., HALLEK, M. & KREUZER, K. A. 2010. Small molecule inhibitors of Wnt/beta-catenin/lef-1 signaling induces apoptosis in chronic lymphocytic leukemia cells in vitro and in vivo. *Neoplasia,* 12**,** 326-35.

GANSTER, C., NEESEN, J., ZEHETMAYER, S., JAGER, U., ESTERBAUER, H., MANNHALTER, C., KLUGE, B. & FONATSCH, C. 2009. DNA repair polymorphisms associated with cytogenetic subgroups in B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer,* 48**,** 760-7.

GARRIGA, S. & CROSBY, W. H. 1959. The incidence of leukemia in families of patients with hypoplasia of the marrow. *Blood,* 14**,** 1008-14.

GARZA, E., DEL POETA, G., MARTINEZ-LOSADA, C., CATALANO, G., BORGIA, L., PIREDDA, M. L., FABIANI, E., GATTEI, V., LO-COCO, F. & NOGUERA, N. I. 2016. A case of SRSF2 mutation in chronic lymphocytic leukemia. *Leuk Res Rep,* 6**,** 11-4.

GAZZOLI, I., LODA, M., GARBER, J., SYNGAL, S. & KOLODNER, R. D. 2002. A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res,* 62**,** 3925-8.

GERMAN, J. 1980. Chromosome-breakage syndromes: different genes, different treatments, different cancers. *Basic Life Sci,* 15**,** 429-39.

GHIA, P., STAMATOPOULOS, K., BELESSI, C., MORENO, C., STILGENBAUER, S., STEVENSON, F., DAVI, F. & ROSENQUIST, R. 2007. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia,* 21**,** 1-3.

GILES, G. G., LICKISS, J. N., BAIKIE, M. J., LOWENTHAL, R. M. & PANTON, J. 1984. Myeloproliferative and lymphoproliferative disorders in Tasmania, 1972-80: occupational and familial aspects. *J Natl Cancer Inst,* 72**,** 1233-40.

GINE, E., CRESPO, M., MONTSERRAT, E. & BOSCH, F. 2004. Cytosolic Histone H1 Releasing under Different Apoptotic Stimuli in Chronic Lymphocytic Leukemia (CLL). *Blood,* 104**,** 4790-4790.

GINE, E., CRESPO, M., MUNTANOLA, A., CALPE, E., BAPTISTA, M. J., VILLAMOR, N., MONTSERRAT, E. & BOSCH, F. 2008. Induction of histone H1.2 cytosolic release in chronic lymphocytic leukemia cells after genotoxic and non-genotoxic treatment. *Haematologica,* 93**,** 75-82.

GLIBERT, P., VOSSAERT, L., VAN STEENDAM, K., LAMBRECHT, S., VAN NIEUWERBURGH, F., OFFNER, F., KIPPS, T., DHAENENS, M. & DEFORCE, D. 2014. Quantitative proteomics to characterize specific histone H2A proteolysis in chronic lymphocytic leukemia and the myeloid THP-1 cell line. *International journal of molecular sciences,* 15**,** 9407-9421.

GOLDGAR, D. E., EASTON, D. F., CANNON-ALBRIGHT, L. A. & SKOLNICK, M. H. 1994. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst,* 86**,** 1600-8.

GOLDIN, L. R., BJORKHOLM, M., KRISTINSSON, S. Y., TURESSON, I. & LANDGREN, O. 2009a. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica,* 94**,** 647-53.

GOLDIN, L. R., BJORKHOLM, M., KRISTINSSON, S. Y., TURESSON, I. & LANDGREN, O. 2009b. Highly increased familial risks for specific lymphoma subtypes. *British Journal of Haematology,* 146**,** 91-94.

GOLDIN, L. R., ISHIBE, N., SGAMBATI, M., MARTI, G. E., FONTAINE, L., LEE, M. P., KELLEY, J. M., SCHERPBIER, T., BUETOW, K. H. & CAPORASO, N. E. 2003. A genome scan of 18 families with chronic lymphocytic leukaemia. *Br J Haematol,* 121**,** 866-73.

GRA, O. A., GLOTOV, A. S., NIKITIN, E. A., GLOTOV, O. S., KUZNETSOVA, V. E., CHUDINOV, A. V., SUDARIKOV, A. B. & NASEDKINA, T. V. 2008. Polymorphisms in xenobiotic-metabolizing genes and the risk of chronic lymphocytic leukemia and non-Hodgkin's lymphoma in adult Russian patients. *Am J Hematol,* 83**,** 279-87.

GRAGERT, L., FINGERSON, S., ALBRECHT, M., MAIERS, M., KALAYCIO, M. & HILL, B. T. 2014. Fine-mapping of HLA associations with chronic lymphocytic leukemia in US populations. *Blood,* 124**,** 2657-2665.

GREBE, S. K. & SINGH, R. J. 2011. LC-MS/MS in the clinical laboratory–where to from here? *The Clinical Biochemist Reviews,* 32**,** 5.

GRIMES, D. A. & SCHULZ, K. F. 2002. Cohort studies: marching towards outcomes. *Lancet,* 359**,** 341-5.

GROENE, J., MANSMANN, U., MEISTER, R., STAUB, E., ROEPCKE, S., HEINZE, M., KLAMAN, I., BRÃ¼MMENDORF, T., HERMANN, K., LODDENKEMPER, C., PILARSKY, C., MANN, B., ADAMS, H. P., BUHR, H. J. & ROSENTHAL, A. 2006. Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III. *Int J Cancer,* 119.

GUARINI, A., MARINELLI, M., TAVOLARO, S., BELLACCHIO, E., MAGLIOZZI, M., CHIARETTI, S., DE PROPRIS, M. S., PERAGINE, N., SANTANGELO, S., PAOLONI, F., NANNI, M., DEL GIUDICE, I., MAURO, F. R., TORRENTE, I. & FOA, R. 2012. ATM gene alterations in chronic lymphocytic leukemia patients induce a distinct gene expression profile and predict disease progression. *Haematologica,* 97**,** 47-55.

GUASCH, J. 1954. [Heredity of leukemias]. *Sang,* 25**,** 384-421.

GUNZ, F. W., GUNZ, J. P., VEALE, A. M., CHAPMAN, C. J. & HOUSTON, I. B. 1975. Familial leukaemia: a study of 909 families. *Scand J Haematol,* 15**,** 117-31.

GUTIERREZ, A., JR., TSCHUMPER, R. C., WU, X., SHANAFELT, T. D., ECKEL-PASSOW, J., HUDDLESTON, P. M., 3RD, SLAGER, S. L., KAY, N. E. & JELINEK, D. F. 2010. LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. *Blood,* 116**,** 2975-83.

GUYATT, G. H., OXMAN, A. D., VIST, G., KUNZ, R., BROZEK, J., ALONSO-COELLO, P., MONTORI, V., AKL, E. A., DJULBEGOVIC, B., FALCK-YTTER, Y., NORRIS, S. L., WILLIAMS, J. W., JR.,

ATKINS, D., MEERPOHL, J. & SCHUNEMANN, H. J. 2011. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol,* 64**,** 407-15.

GUZOWSKI, D., CHANDRASEKARAN, A., GAWEL, C., PALMA, J., KOENIG, J., WANG, X. P., DOSIK, M., KAPLAN, M., CHU, C. C., CHAVAN, S., FURIE, R., ALBESIANO, E., CHIORAZZI, N. & GOODWIN, L. 2005. Analysis of single nucleotide polymorphisms in the promoter region of interleukin-10 by denaturing high-performance liquid chromatography. *J Biomol Tech,* 16**,** 154-66.

HAHN, W. C., COUNTER, C. M., LUNDBERG, A. S., BEIJERSBERGEN, R. L., BROOKS, M. W. & WEINBERG, R. A. 1999. Creation of human tumour cells with defined genetic elements. *Nature,* 400**,** 464-8.

HALLEK, M., CHESON, B. D., CATOVSKY, D., CALIGARIS-CAPPIO, F., DIGHIERO, G., DOHNER, H., HILLMEN, P., KEATING, M. J., MONTSERRAT, E., RAI, K. R. & KIPPS, T. J. 2008. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood,* 111**,** 5446-56.

HAMBLIN, T. J., DAVIS, Z., GARDINER, A., OSCIER, D. G. & STEVENSON, F. K. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood,* 94**,** 1848-54.

HAPPEL, N. & DOENECKE, D. 2009. Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene,* 431**,** 1-12.

HARSHMAN, S. W., CHEN, M. M., BRANSON, O. E., JACOB, N. K., JOHNSON, A. J., BYRD, J. C. & FREITAS, M. A. 2013. Isolation and analysis of linker histones across cellular compartments. *J Proteomics,* 91**,** 595-604.

HATCH, M. & CARDIS, E. 2017. Somatic health effects of Chernobyl: 30 years on. *Eur J Epidemiol*.

HEDENFALK, I., DUGGAN, D., CHEN, Y., RADMACHER, M., BITTNER, M., SIMON, R., MELTZER, P., GUSTERSON, B., ESTELLER, M., KALLIONIEMI, O. P., WILFOND, B., BORG, A., TRENT, J., RAFFELD, M., YAKHINI, Z., BEN-DOR, A., DOUGHERTY, E., KONONEN, J., BUBENDORF, L., FEHRLE, W., PITTALUGA, S., GRUVBERGER, S., LOMAN, N., JOHANNSSON, O., OLSSON, H. & SAUTER, G. 2001. Gene-expression profiles in hereditary breast cancer. *N Engl J Med,* 344**,** 539-48.

HEROLD, T., JURINOVIC, V., METZELER, K. H., BOULESTEIX, A. L., BERGMANN, M., SEILER, T., MULAW, M., THOENE, S., DUFOUR, A., PASALIC, Z., SCHMIDBERGER, M., SCHMIDT, M., SCHNEIDER, S., KAKADIA, P. M., FEURING-BUSKE, M., BRAESS, J., SPIEKERMANN, K., MANSMANN, U., HIDDEMANN, W., BUSKE, C. & BOHLANDER, S. K. 2011. An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. *Leukemia,* 25**,** 1639-45.

HERVE, M., XU, K., NG, Y. S., WARDEMANN, H., ALBESIANO, E., MESSMER, B. T., CHIORAZZI, N. & MEFFRE, E. 2005. Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *J Clin Invest,* 115**,** 1636-43.

HILL, D. A., WANG, S. S., CERHAN, J. R., DAVIS, S., COZEN, W., SEVERSON, R. K., HARTGE, P., WACHOLDER, S., YEAGER, M., CHANOCK, S. J. & ROTHMAN, N. 2006. Risk of non-Hodgkin lymphoma (NHL) in relation to germline variation in DNA repair and related genes. *Blood,* 108**,** 3161-7.

HOF, J., KRENTZ, S., VAN SCHEWICK, C., KORNER, G., SHALAPOUR, S., RHEIN, P., KARAWAJEW, L., LUDWIG, W. D., SEEGER, K., HENZE, G., VON STACKELBERG, A., HAGEMEIER, C., ECKERT, C. & KIRSCHNER-SCHWABE, R. 2011. Mutations and

deletions of the TP53 gene predict nonresponse to treatment and poor outcome in first relapse of childhood acute lymphoblastic leukemia. *J Clin Oncol,* 29**,** 3185-93.

HOWLADER N, N. A., KRAPCHO M, MILLER D, BISHOP K, KOSARY CL, YU M, RUHL J, TATALOVICH Z, MARIOTTO A, LEWIS DR, CHEN HS, FEUER EJ, CRONIN KA 2017. SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. Bethesda, MD. Based on November 2016 SEER data submission, posted to the SEER web site, April 2017.

HU, Q., NOLL, R. J., LI, H., MAKAROV, A., HARDMAN, M. & GRAHAM COOKS, R. 2005. The Orbitrap: a new mass spectrometer. *J Mass Spectrom,* 40**,** 430-43.

HUANG, P. Y., BEST, O. G., ALMAZI, J. G., BELOV, L., DAVIS, Z. A., MAJID, A., DYER, M. J., PASCOVICI, D., MULLIGAN, S. P. & CHRISTOPHERSON, R. I. 2014. Cell surface phenotype profiles distinguish stable and progressive chronic lymphocytic leukemia. *Leuk Lymphoma,* 55**,** 2085-92.

HUANG, P. Y., MACTIER, S., ARMACKI, N., GILES BEST, O., BELOV, L., KAUFMAN, K. L., PASCOVICI, D., MULLIGAN, S. P. & CHRISTOPHERSON, R. I. 2016. Protein profiles distinguish stable and progressive chronic lymphocytic leukemia. *Leuk Lymphoma,* 57**,** 1033-43.

HUH, Y. O., SCHWEIGHOFER, C. D., KETTERLING, R. P., KNUDSON, R. A., VEGA, F., KIM, J. E., LUTHRA, R., KEATING, M. J., MEDEIROS, L. J. & ABRUZZO, L. V. 2011. Chronic lymphocytic leukemia with t (14; 19)(q32; q13) is characterized by atypical morphologic and immunophenotypic features and distinctive genetic features. *American journal of clinical pathology,* 135**,** 686-696.

HUMPLIKOVA, L., KOLLINEROVA, S., PAPAJIK, T., PIKALOVA, Z., HOLZEROVA, M., PROCHAZKA, V., DIVOKA, M., MODRIANSKY, M., INDRAK, K. & JAROSOVA, M. 2013. Expression of miR-15a and miR-16-1 in patients with chronic lymphocytic leukemia. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub,* 157**,** 284-93.

IHARA, Y., KIHARA, Y., HAMANO, F., YANAGIDA, K., MORISHITA, Y., KUNITA, A., YAMORI, T., FUKAYAMA, M., ABURATANI, H., SHIMIZU, T. & ISHII, S. 2010. The G protein-coupled receptor T-cell death-associated gene 8 (TDAG8) facilitates tumor development by serving as an extracellular pH sensor. *Proc Natl Acad Sci U S A,* 107**,** 17309-14.

INAMDAR, K. V. & BUESO-RAMOS, C. E. 2007. Pathology of chronic lymphocytic leukemia: an update. *Ann Diagn Pathol,* 11**,** 363-89.

INOUE, J., GOHDA, J., AKIYAMA, T. & SEMBA, K. 2007. NF-kappaB activation in development and progression of cancer. *Cancer Sci,* 98**,** 268-74.

ISNARDI, I., BRUHNS, P., BISMUTH, G., FRIDMAN, W. H. & DAERON, M. 2006. The SH2 domain-containing inositol 5-phosphatase SHIP1 is recruited to the intracytoplasmic domain of human FcgammaRIIB and is mandatory for negative regulation of B cell activation. *Immunol Lett,* 104**,** 156-65.

ISONO, K., MIZUTANI-KOSEKI, Y., KOMORI, T., SCHMIDT-ZACHMANN, M. S. & KOSEKI, H. 2005. Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. *Genes Dev,* 19**,** 536-41.

ISSAQ, H. J., CHAN, K. C., JANINI, G. M., CONRADS, T. P. & VEENSTRA, T. D. 2005. Multidimensional separation of peptides for effective proteomic analysis. *Journal of Chromatography B,* 817**,** 35-47.

JACINTO, E., LOEWITH, R., SCHMIDT, A., LIN, S., RUEGG, M. A., HALL, A. & HALL, M. N. 2004. Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. *Nat Cell Biol,* 6**,** 1122-8.

JACKSON, R. S., 2ND, CHO, Y. J. & LIANG, P. 2006. TIS11D is a candidate pro-apoptotic p53 target gene. *Cell Cycle,* 5**,** 2889-93.

JAKSIK, R., IWANASZKO, M., RZESZOWSKA-WOLNY, J. & KIMMEL, M. 2015. Microarray experiments and factors which affect their reliability. *Biology direct,* 10**,** 46.

JAMROZIAK, K., BALCERCZAK, E., SMOLEWSKI, P., ROBEY, R. W., CEBULA, B., PANCZYK, M., KOWALCZYK, M., SZMIGIELSKA-KAPLON, A., MIROWSKI, M., BATES, S. E. & ROBAK, T. 2006. MDR1 (ABCB1) gene polymorphism C3435T is associated with P-glycoprotein activity in B-cell chronic lymphocytic leukemia. *Pharmacol Rep,* 58**,** 720-8.

JAMROZIAK, K., SZEMRAJ, Z., GRZYBOWSKA-IZYDORCZYK, O., SZEMRAJ, J., BIENIASZ, M., CEBULA, B., GIANNOPOULOS, K., BALCERCZAK, E., JESIONEK-KUPNICKA, D., KOWAL, M., KOSTYRA, A., CALBECKA, M., WAWRZYNIAK, E., MIROWSKI, M., KORDEK, R. & ROBAK, T. 2009. CD38 gene polymorphisms contribute to genetic susceptibility to B-cell chronic lymphocytic leukemia: evidence from two case-control studies in Polish Caucasians. *Cancer Epidemiol Biomarkers Prev,* 18**,** 945-53.

JEROMIN, S., WEISSMANN, S., HAFERLACH, C., DICKER, F., BAYER, K., GROSSMANN, V., ALPERMANN, T., ROLLER, A., KOHLMANN, A., HAFERLACH, T., KERN, W. & SCHNITTGER, S. 2014. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia,* 28**,** 108-17.

JIMENEZ-ZEPEDA, V. H., CHNG, W. J., SCHOP, R. F., BRAGGIO, E., LEIS, J. F., KAY, N. & FONSECA, R. 2013. Recurrent chromosome abnormalities define nonoverlapping unique subgroups of tumors in patients with chronic lymphocytic leukemia and known karyotypic abnormalities. *Clin Lymphoma Myeloma Leuk,* 13**,** 467-76.

JOERGER, A. C. & FERSHT, A. R. 2008. Structural biology of the tumor suppressor p53. *Annual Review of Biochemistry,* 77**,** 557-82.

JOSHI, A. D., HEGDE, G. V., DICKINSON, J. D., MITTAL, A. K., LYNCH, J. C., EUDY, J. D., ARMITAGE, J. O., BIERMAN, P. J., BOCIEK, R. G. & DEVETTEN, M. P. 2007. ATM, CTLA4, MNDA, and HEM1 in High versus Low CD38–Expressing B-Cell Chronic Lymphocytic Leukemia. *Clinical Cancer Research,* 13**,** 5295-5304.

JUSTUS, C. R., SANDERLIN, E. J., DONG, L., SUN, T., CHI, J. T., LERTPIRIYAPONG, K. & YANG, L. V. 2017. Contextual tumor suppressor function of T cell death-associated gene 8 (TDAG8) in hematological malignancies. *J Transl Med,* 15**,** 204.

KALL, L., CANTERBURY, J. D., WESTON, J., NOBLE, W. S. & MACCOSS, M. J. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods,* 4**,** 923-5.

KALLI, A., SMITH, G. T., SWEREDOSKI, M. J. & HESS, S. 2013. Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers. *Journal of proteome research,* 12**,** 3071-3086.

KAMIKIHARA, T., ARIMA, T., KATO, K., MATSUDA, T., KATO, H., DOUCHI, T., NAGATA, Y., NAKAO, M. & WAKE, N. 2005. Epigenetic silencing of the imprinted gene ZAC by DNA methylation is an early event in the progression of human ovarian cancer. *Int J Cancer,* 115**,** 690-700.

KANEDA, A. & FEINBERG, A. P. 2005. Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. *Cancer Res,* 65**,** 11236-40.

KARIN, M. 2006. Nuclear factor-kappaB in cancer development and progression. *Nature,* 441**,** 431-6.

KAWAUCHI, K., OGASAWARA, T. & YASUYAMA, M. 2002. Activation of extracellular signal-regulated kinase through B-cell antigen receptor in B-cell chronic lymphocytic leukemia. *International journal of hematology,* 75**,** 508-513.

KELLER, A., NESVIZHSKII, A. I., KOLKER, E. & AEBERSOLD, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem,* 74**,** 5383-92.

KERBER, R. A. & O'BRIEN, E. 2005. A cohort study of cancer risk in relation to family histories of cancer in the Utah population database. *Cancer,* 103**,** 1906-15.

KITANO, J., YAMAZAKI, Y., KIMURA, K., MASUKADO, T., NAKAJIMA, Y. & NAKANISHI, S. 2003. Tamalin is a scaffold protein that interacts with multiple neuronal proteins in distinct modes of protein-protein association. *J Biol Chem,* 278**,** 14762-8.

KLEIN, U. & DALLA-FAVERA, R. 2010. New insights into the pathogenesis of chronic lymphocytic leukemia. *Seminars in Cancer Biology,* 20**,** 377-83.

KLEIN, U., LIA, M., CRESPO, M., SIEGEL, R., SHEN, Q., MO, T., AMBESI-IMPIOMBATO, A., CALIFANO, A., MIGLIAZZA, A., BHAGAT, G. & DALLA-FAVERA, R. 2010. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell,* 17**,** 28-40.

KLEIN, U., TU, Y., STOLOVITZKY, G. A., MATTIOLI, M., CATTORETTI, G., HUSSON, H., FREEDMAN, A., INGHIRAMI, G., CRO, L., BALDINI, L., NERI, A., CALIFANO, A. & DALLA-FAVERA, R. 2001. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med,* 194**,** 1625-38.

KLINKOV, A. A., NIKITIN, E. A., MAIOROVA, O. V., IVANOV, M. A., STRELNIKOV, V. V., BABENKO, O. V., ZEMLYAKOVA, V. V., KUZNETSOVA, E. B. & ZALETAYEV, D. V. 2004. TNR/11q#1 trinucleotide (GCC)n repeat alleles and predisposition to acute and chronic leukemia. *Ann Hum Genet,* 68**,** 362-6.

KNUDSON, A. G., JR. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A,* 68**,** 820-3.

KOCHETHU, G., DELGADO, J., PEPPER, C., STARCZYNSKI, J., HOOPER, L., KRISHNAN, S., FEGAN, C. & PRATT, G. 2006. Two germ line polymorphisms of the tumour suppressor gene p53 may influence the biology of chronic lymphocytic leukaemia. *Leuk Res,* 30**,** 1113-8.

KOJIMA, K., KONOPLEVA, M., MCQUEEN, T., O'BRIEN, S., PLUNKETT, W. & ANDREEFF, M. 2006. Mdm2 inhibitor Nutlin-3a induces p53-mediated apoptosis by transcription-dependent and transcription-independent mechanisms and may overcome Atm-mediated resistance to fludarabine in chronic lymphocytic leukemia. *Blood,* 108**,** 993-1000.

KROLL, J., SHI, X., CAPRIOLI, A., LIU, H. H., WASKOW, C., LIN, K. M., MIYAZAKI, T., RODEWALD, H. R. & SATO, T. N. 2005. The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol Cell Biol,* 25**,** 8531-40.

KRYSOV, S., DIAS, S., PATERSON, A., MOCKRIDGE, C. I., POTTER, K. N., SMITH, K.-A., ASHTON-KEY, M., STEVENSON, F. K. & PACKHAM, G. 2012. Surface IgM stimulation induces MEK1/2-dependent MYC expression in chronic lymphocytic leukemia cells. *Blood,* 119**,** 170-179.

KULKARNI, A. B., HUH, C. G., BECKER, D., GEISER, A., LYGHT, M., FLANDERS, K. C., ROBERTS, A. B., SPORN, M. B., WARD, J. M. & KARLSSON, S. 1993. Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death. *Proc Natl Acad Sci U S A,* 90**,** 770-4.

KYRTSONIS, M. C., REPA, C., DEDOUSSIS, G. V., MOUZAKI, A., SIMEONIDIS, A., STAMATELOU, M. & MANIATIS, A. 1998. Serum transforming growth factor-beta 1 is related to the degree of immunoparesis in patients with multiple myeloma. *Med Oncol,* 15**,** 124-8.

LAGNEAUX, L., DELFORGE, A., BRON, D., BOSMANS, E. & STRYCKMANS, P. 1995. Comparative analysis of cytokines released by bone marrow stromal cells from normal donors and B-cell chronic lymphocytic leukemic patients. *Leukemia & lymphoma,* 17**,** 127-133.

LAGNEAUX, L., DELFORGE, A., BRON, D., MASSY, M., BERNIER, M. & STRYCKMANS, P. 1997. Heterogenous response of B lymphocytes to transforming growth factor-beta in B-cell chronic lymphocytic leukaemia: correlation with the expression of TGF-β receptors. *British journal of haematology,* 97**,** 612-620.

LAN, Q., MORTON, L. M., ARMSTRONG, B., HARTGE, P., MENASHE, I., ZHENG, T., PURDUE, M. P., CERHAN, J. R., ZHANG, Y., GRULICH, A., COZEN, W., YEAGER, M., HOLFORD, T. R., VAJDIC, C. M., DAVIS, S., LEADERER, B., KRICKER, A., SCHENK, M., ZAHM, S. H., CHATTERJEE, N., CHANOCK, S. J., ROTHMAN, N. & WANG, S. S. 2009. Genetic variation in caspase genes and risk of non-Hodgkin lymphoma: a pooled analysis of 3 population-based case-control studies. *Blood,* 114**,** 264-7.

LAN, Q., ZHENG, T., SHEN, M., ZHANG, Y., WANG, S. S., ZAHM, S. H., HOLFORD, T. R., LEADERER, B., BOYLE, P. & CHANOCK, S. 2007. Genetic polymorphisms in the oxidative stress pathway and susceptibility to non-Hodgkin lymphoma. *Hum Genet,* 121**,** 161-8.

LANASA, M. C., ALLGOOD, S. D., SLAGER, S. L., DAVE, S. S., LOVE, C., MARTI, G. E., KAY, N. E., HANSON, C. A., RABE, K. G., ACHENBACH, S. J., GOLDIN, L. R., CAMP, N. J., GOODMAN, B. K., VACHON, C. M., SPECTOR, L. G., RASSENTI, L. Z., LEIS, J. F., GOCKERMAN, J. P., STROM, S. S., CALL, T. G., GLENN, M., CERHAN, J. R., LEVESQUE, M. C., WEINBERG, J. B. & CAPORASO, N. E. 2011. Immunophenotypic and gene expression analysis of monoclonal B-cell lymphocytosis shows biologic characteristics associated with good prognosis CLL. *Leukemia,* 25**,** 1459-66.

LANDGREN, O., ALBITAR, M., MA, W., ABBASI, F., HAYES, R. B., GHIA, P., MARTI, G. E. & CAPORASO, N. E. 2009. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med,* 360**,** 659-67.

LAPIERRE, M., CASTET-NICOLAS, A., GITENAY, D., JALAGUIER, S., TEYSSIER, C., BRET, C., CARTRON, G., MOREAUX, J. & CAVAILLES, V. 2015. Expression and role of RIP140/NRIP1 in chronic lymphocytic leukemia. *J Hematol Oncol,* 8**,** 20.

LAPLANTE, M. & SABATINI, D. M. 2012. mTOR signaling in growth control and disease. *Cell,* 149**,** 274-93.

LAW, P. J., SUD, A., MITCHELL, J. S., HENRION, M., ORLANDO, G., LENIVE, O., BRODERICK, P., SPEEDY, H. E., JOHNSON, D. C., KAISER, M., WEINHOLD, N., COOKE, R., SUNTER, N. J., JACKSON, G. H., SUMMERFIELD, G., HARRIS, R. J., PETTITT, A. R., ALLSUP, D. J., CARMICHAEL, J., BAILEY, J. R., PRATT, G., RAHMAN, T., PEPPER, C., FEGAN, C., VON STRANDMANN, E. P., ENGERT, A., FORSTI, A., CHEN, B., FILHO, M. I., THOMSEN, H., HOFFMANN, P., NOETHEN, M. M., EISELE, L., JOCKEL, K. H., ALLAN, J. M., SWERDLOW, A. J., GOLDSCHMIDT, H., CATOVSKY, D., MORGAN, G. J., HEMMINKI, K. & HOULSTON, R. S. 2017. Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci. *Sci Rep,* 7**,** 41071.

LEE, S. H., WRAY, N. R., GODDARD, M. E. & VISSCHER, P. M. 2011. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet,* 88**,** 294-305.

LEMOS, M. C., CABRITA, F. J., SILVA, H. A., VIVAN, M., PLACIDO, F. & REGATEIRO, F. J. 1999. Genetic polymorphism of CYP2D6, GSTM1 and NAT2 and susceptibility to haematological neoplasias. *Carcinogenesis,* 20**,** 1225-9.

LENG, L., METZ, C. N., FANG, Y., XU, J., DONNELLY, S., BAUGH, J., DELOHERY, T., CHEN, Y., MITCHELL, R. A. & BUCALA, R. 2003. MIF signal transduction initiated by binding to CD74. *J Exp Med,* 197**,** 1467-76.

LI, F. P. & FRAUMENI, J. F., JR. 1969. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J Natl Cancer Inst,* 43**,** 1365-73.

LI, Z., DONG, L., DEAN, E. & YANG, L. V. 2013. Acidosis decreases c-Myc oncogene expression in human lymphoma cells: a role for the proton-sensing G protein-coupled receptor TDAG8. *Int J Mol Sci,* 14**,** 20236-55.

LIA, M., CARETTE, A., TANG, H., SHEN, Q., MO, T., BHAGAT, G., DALLA-FAVERA, R. & KLEIN, U. 2012. Functional dissection of the chromosome 13q14 tumor-suppressor locus using transgenic mouse lines. *Blood,* 119**,** 2981-90.

LIANG, X. S., CAPORASO, N., MCMASTER, M. L., NG, D., LANDGREN, O., YEAGER, M., CHANOCK, S. & GOLDIN, L. R. 2009. Common genetic variants in candidate genes and risk of familial lymphoid malignancies. *Br J Haematol,* 146**,** 418-23.

LIAO, W., JORDAAN, G., NHAM, P., PHAN, R. T., PELEGRINI, M. & SHARMA, S. 2015. Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens. *BMC cancer,* 15**,** 1.

LICHTENSTEIN, P., HOLM, N. V., VERKASALO, P. K., ILIADOU, A., KAPRIO, J., KOSKENVUO, M., PUKKALA, E., SKYTTHE, A. & HEMMINKI, K. 2000. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med,* 343**,** 78-85.

LIM, U., WANG, S. S., HARTGE, P., COZEN, W., KELEMEN, L. E., CHANOCK, S., DAVIS, S., BLAIR, A., SCHENK, M., ROTHMAN, N. & LAN, Q. 2007. Gene-nutrient interactions among determinants of folate and one-carbon metabolism on the risk of non-Hodgkin lymphoma: NCI-SEER case-control study. *Blood,* 109**,** 3050-9.

LIN, K., LANE, B., CARTER, A., JOHNSON, G. G., ONWUAZOR, O., OATES, M., ZENZ, T., STILGENBAUER, S., ATHERTON, M., DOUGLAS, A., EBRAHIMI, B., SHERRINGTON, P. D. & PETTITT, A. R. 2013. The gene expression signature associated with TP53 mutation/deletion in chronic lymphocytic leukaemia is dominated by the under-expression of TP53 and other genes on chromosome 17p. *Br J Haematol,* 160**,** 53-62.

LIN, Y.-W. & APLAN, P. D. 2007. Gene expression profiling of precursor T-cell lymphoblastic leukemia/lymphoma identifies oncogenic pathways that are potential therapeutic targets. *Leukemia,* 21**,** 1276.

LIU, P., RAMACHANDRAN, S., SEYED, M. A., SCHARER, C. D., LAYCOCK, N., DALTON, W. B., WILLIAMS, H., KARANAM, S., DATTA, M. W. & JAYE, D. L. 2006. Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer research,* 66**,** 4011-4019.

LIU, Y., GRANDER, D., SODERHALL, S., JULIUSSON, G., GAHRTON, G. & EINHORN, S. 1992. Retinoblastoma gene deletions in B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer,* 4**,** 250-6.

LOZANSKI, G., HEEREMA, N. A., FLINN, I. W., SMITH, L., HARBISON, J., WEBB, J., MORAN, M., LUCAS, M., LIN, T., HACKBARTH, M. L., PROFFITT, J. H., LUCAS, D., GREVER, M. R. & BYRD, J. C. 2004. Alemtuzumab is an effective therapy for chronic lymphocytic leukemia with p53 mutations and deletions. *Blood,* 103**,** 3278-81.

LUDYGA, N., ENGLERT, S., PFLIEGER, K., RAUSER, S., BRASELMANN, H., WALCH, A., AUER, G., HÖFLER, H. & AUBELE, M. 2013. The impact of Cysteine-Rich Intestinal Protein 1 (CRIP1) in human breast cancer. *Molecular Cancer,* 12**,** 28.

LUE, H., KAPURNIOTU, A., FINGERLE-ROWSON, G., ROGER, T., LENG, L., THIELE, M., CALANDRA, T., BUCALA, R. & BERNHAGEN, J. 2006. Rapid and transient activation of the ERK MAPK signalling pathway by macrophage migration inhibitory factor (MIF) and dependence on JAB1/CSN5 and Src kinase activity. *Cell Signal,* 18**,** 688-703.

LUE, H., THIELE, M., FRANZ, J., DAHL, E., SPECKGENS, S., LENG, L., FINGERLE-ROWSON, G., BUCALA, R., LUSCHER, B. & BERNHAGEN, J. 2007. Macrophage migration inhibitory factor (MIF) promotes cell survival by activation of the Akt pathway and role for CSN5/JAB1 in the control of autocrine MIF activity. *Oncogene,* 26**,** 5046-59.

LYNCH, H. T., WEISENBURGER, D. D., QUINN-LAQUER, B., WATSON, P., LYNCH, J. F. & SANGER, W. G. 2002. Hereditary chronic lymphocytic leukemia: an extended family study and literature review. *Am J Med Genet,* 115**,** 113-7.

MA, X. J., SALUNGA, R., TUGGLE, J. T., GAUDET, J., ENRIGHT, E., MCQUARY, P., PAYETTE, T., PISTONE, M., STECKER, K., ZHANG, B. M., ZHOU, Y. X., VARNHOLT, H., SMITH, B., GADD, M., CHATFIELD, E., KESSLER, J., BAER, T. M., ERLANDER, M. G. & SQROI, D. C. 2003. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A,* 100.

MACHULLA, H. K., MULLER, L. P., SCHAAF, A., KUJAT, G., SCHONERMARCK, U. & LANGNER, J. 2001. Association of chronic lymphocytic leukemia with specific alleles of the HLA-DR4:DR53:DQ8 haplotype in German patients. *Int J Cancer,* 92**,** 203-7.

MAINOU-FOWLER, T., DICKINSON, A. M., TAYLOR, P. R., MOUNTER, P., JACK, F., PROCTOR, S. J., NORDON, J. & MIDDLETON, P. G. 2000. Tumour necrosis factor gene polymorphisms in lymphoproliferative disease. *Leuk Lymphoma,* 38**,** 547-52.

MAJID, A., TSOULAKIS, O., WALEWSKA, R., GESK, S., SIEBERT, R., KENNEDY, D. B. & DYER, M. J. 2008. BCL2 expression in chronic lymphocytic leukemia: lack of association with the BCL2 938A>C promoter single nucleotide polymorphism. *Blood,* 111**,** 874-7.

MALKIN, D., LI, F. P., STRONG, L. C., FRAUMENI, J. F., JR., NELSON, C. E., KIM, D. H., KASSEL, J., GRYKA, M. A., BISCHOFF, F. Z., TAINSKY, M. A. & ET AL. 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science,* 250**,** 1233-8.

MALLAMPATI, S., SUN, B., LU, Y., MA, H., GONG, Y., WANG, D., LEE, J. S., LIN, K. & SUN, X. 2014. Integrated genetic approaches identify the molecular mechanisms of Sox4 in early B-cell development: intricate roles for RAG1/2 and CK1epsilon. *Blood,* 123**,** 4064-76.

MARASCA, R., MAFFEI, R., MARTINELLI, S., FIORCARI, S., BULGARELLI, J., DEBBIA, G., ROSSI, D., ROSSI, F. M., RIGOLIN, G. M., MARTINELLI, S., GATTEI, V., DEL POETA, G., LAURENTI, L., FORCONI, F., MONTILLO, M., GAIDANO, G. & LUPPI, M. 2013. Clinical heterogeneity of de novo 11q deletion chronic lymphocytic leukaemia: prognostic relevance of extent of 11q deleted nuclei inside leukemic clone. *Hematological Oncology,* 31**,** 88-95.

MARTI, G. E., CARTER, P., ABBASI, F., WASHINGTON, G. C., JAIN, N., ZENGER, V. E., ISHIBE, N., GOLDIN, L., FONTAINE, L., WEISSMAN, N., SGAMBATI, M., FAUGET, G., BERTIN, P., VOGT, R. F., JR., SLADE, B., NOGUCHI, P. D., STETLER-STEVENSON, M. A. & CAPORASO, N. 2003. B-cell monoclonal lymphocytosis and B-cell abnormalities in the setting of familial B-cell chronic lymphocytic leukemia. *Cytometry B Clin Cytom,* 52**,** 1-12.

MARTIN-SUBERO, J., IBBOTSON, R., KLAPPER, W., MICHAUX, L., CALLET-BAUCHU, E., BERGER, F., CALASANZ, M., DE WOLF-PEETERS, C., DYER, M. & FELMAN, P. 2007. A comprehensive genetic and histopathologic analysis identifies two subgroups of B-

cell malignancies carrying at (14; 19)(q32; q13) or variant BCL3-translocation. *Leukemia,* 21**,** 1532-1544.

MATERA, A. G. & WANG, Z. 2014. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology,* 15**,** 108-21.

MATSUZAWA, A., TSENG, P. H., VALLABHAPURAPU, S., LUO, J. L., ZHANG, W., WANG, H., VIGNALI, D. A., GALLAGHER, E. & KARIN, M. 2008. Essential cytoplasmic translocation of a cytokine receptor-assembled signaling complex. *Science,* 321**,** 663-8.

MATTHIESEN, R. 2007. Methods, algorithms and tools in computational proteomics: a practical point of view. *Proteomics,* 7**,** 2815-32.

MATVEEVA, A., KOVALEVSKA, L., KHOLODNYUK, I., IVANIVSKAYA, T. & KASHUBA, E. 2017. The TGF-beta - SMAD pathway is inactivated in cronic lymphocytic leukemia cells. *Exp Oncol,* 39**,** 286-290.

MCBRIDE, K. A., BALLINGER, M. L., KILLICK, E., KIRK, J., TATTERSALL, M. H., EELES, R. A., THOMAS, D. M. & MITCHELL, G. 2014. Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nat Rev Clin Oncol,* 11**,** 260-71.

MCCARTHY, B. A., YANCOPOULOS, S., TIPPING, M., YAN, X. J., WANG, X. P., BENNETT, F., LI, W., LESSER, M., PAUL, S., BOYLE, E., MORENO, C., CATERA, R., MESSMER, B. T., CUTRONA, G., FERRARINI, M., KOLITZ, J. E., ALLEN, S. L., RAI, K. R., RAWSTRON, A. C. & CHIORAZZI, N. 2015. A seven-gene expression panel distinguishing clonal expansions of pre-leukemic and chronic lymphocytic leukemia B cells from normal B lymphocytes. *Immunol Res,* 63**,** 90-100.

MEGHANN PALERMO, H. D., SCOTT TIGHE, JULIE DRAGON, JEFF BOND, ARTI SHUKLA, MAHESH VANGALA, JAMES VINCENT, TIM HUNTER 2014. Expression Profiling Smackdown: Human Transcriptome Array HTA 2.0 vs. RNA-Seq. *J Biomol Tech,* 25(Suppl)**,** S20–S21.

MEYER-SIEGLER, K. L., VERA, P. L., ICZKOWSKI, K. A., BIFULCO, C., LEE, A., GREGERSEN, P. K., LENG, L. & BUCALA, R. 2007. Macrophage migration inhibitory factor (MIF) gene polymorphisms are associated with increased prostate cancer incidence. *Genes Immun,* 8**,** 646-52.

MITTAL, A. K., CHATURVEDI, N. K., ROHLFSEN, R. A., GUPTA, P., JOSHI, A. D., HEGDE, G. V., BOCIEK, R. G. & JOSHI, S. S. 2013. Role of CTLA4 in the proliferation and survival of chronic lymphocytic leukemia. *PLoS One,* 8**,** e70352.

MONTES-ARES, O., MOYA-QUILES, M. R., MONTES-CASADO, M., GUERRA-PEREZ, N., CAMPILLO, J. A., GONZALEZ, C., LOPEZ-BERMEJO, A., TAMAYO, M., MAJADO, M. J., PARRADO, A., MURO, M., MARIN, L. & ALVAREZ-LOPEZ, M. R. 2006. Human leucocyte antigen-C in B chronic lymphocytic leukaemia. *Br J Haematol,* 135**,** 517-9.

MORALES, A. A., OLSSON, A., CELSING, F., OSTERBORG, A., JONDAL, M. & OSORIO, L. M. 2004. Expression and transcriptional regulation of functionally distinct Bmf isoforms in B-chronic lymphocytic leukemia cells. *Leukemia,* 18**,** 41-7.

MORIN, G. B. 1989. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell,* 59**,** 521-9.

MORTON, L. M., SCHENK, M., HEIN, D. W., DAVIS, S., ZAHM, S. H., COZEN, W., CERHAN, J. R., HARTGE, P., WELCH, R., CHANOCK, S. J., ROTHMAN, N. & WANG, S. S. 2006. Genetic variation in N-acetyltransferase 1 (NAT1) and 2 (NAT2) and risk of non-Hodgkin lymphoma. *Pharmacogenet Genomics,* 16**,** 537-45.

MORTON, L. M., SLAGER, S. L., CERHAN, J. R., WANG, S. S., VAJDIC, C. M., SKIBOLA, C. F., BRACCI, P. M., DE SANJOSE, S., SMEDBY, K. E., CHIU, B. C., ZHANG, Y., MBULAITEYE, S. M., MONNEREAU, A., TURNER, J. J., CLAVEL, J., ADAMI, H. O., CHANG, E. T.,

GLIMELIUS, B., HJALGRIM, H., MELBYE, M., CROSIGNANI, P., DI LOLLO, S., MILIGI, L., NANNI, O., RAMAZZOTTI, V., RODELLA, S., COSTANTINI, A. S., STAGNARO, E., TUMINO, R., VINDIGNI, C., VINEIS, P., BECKER, N., BENAVENTE, Y., BOFFETTA, P., BRENNAN, P., COCCO, P., FORETOVA, L., MAYNADIE, M., NIETERS, A., STAINES, A., COLT, J. S., COZEN, W., DAVIS, S., DE ROOS, A. J., HARTGE, P., ROTHMAN, N., SEVERSON, R. K., HOLLY, E. A., CALL, T. G., FELDMAN, A. L., HABERMANN, T. M., LIEBOW, M., BLAIR, A., CANTOR, K. P., KANE, E. V., LIGHTFOOT, T., ROMAN, E., SMITH, A., BROOKS-WILSON, A., CONNORS, J. M., GASCOYNE, R. D., SPINELLI, J. J., ARMSTRONG, B. K., KRICKER, A., HOLFORD, T. R., LAN, Q., ZHENG, T., ORSI, L., DAL MASO, L., FRANCESCHI, S., LA VECCHIA, C., NEGRI, E., SERRAINO, D., BERNSTEIN, L., LEVINE, A., FRIEDBERG, J. W., KELLY, J. L., BERNDT, S. I., BIRMANN, B. M., CLARKE, C. A., FLOWERS, C. R., FORAN, J. M., KADIN, M. E., PALTIEL, O., WEISENBURGER, D. D., LINET, M. S. & SAMPSON, J. N. 2014. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr,* 2014**,** 130-44.

MRUK, D. D. & CHENG, C. Y. 2011. Enhanced chemiluminescence (ECL) for routine immunoblotting: An inexpensive alternative to commercially available kits. *Spermatogenesis,* 1**,** 121-122.

MUGHAL, T. I., GOLDMAN, J. M. & MUGHAL, S. T. 2006. *Understanding leukemias, lymphomas, and myelomas,* New York, Taylor & Francis.

MULLEN, T. D., HANNUN, Y. A. & OBEID, L. M. 2012. Ceramide synthases at the centre of sphingolipid metabolism and biology. *Biochem J,* 441**,** 789-802.

MURRAY, F., DARZENTAS, N., HADZIDIMITRIOU, A., TOBIN, G., BOUDJOGRA, M., SCIELZO, C., LAOUTARIS, N., KARLSSON, K., BARAN-MARZSAK, F., TSAFTARIS, A., MORENO, C., ANAGNOSTOPOULOS, A., CALIGARIS-CAPPIO, F., VAUR, D., OUZOUNIS, C., BELESSI, C., GHIA, P., DAVI, F., ROSENQUIST, R. & STAMATOPOULOS, K. 2008. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood,* 111**,** 1524-33.

NADEU, F., DELGADO, J., ROYO, C., BAUMANN, T., STANKOVIC, T., PINYOL, M., JARES, P., NAVARRO, A., MARTIN-GARCIA, D., BEA, S., SALAVERRIA, I., OLDREIVE, C., AYMERICH, M., SUAREZ-CISNEROS, H., ROZMAN, M., VILLAMOR, N., COLOMER, D., LOPEZ-GUILLERMO, A., GONZALEZ, M., ALCOCEBA, M., TEROL, M. J., COLADO, E., PUENTE, X. S., LOPEZ-OTIN, C., ENJUANES, A. & CAMPO, E. 2016. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood,* 127**,** 2122-30.

NAZAROV, P. V., MULLER, A., KAOMA, T., NICOT, N., MAXIMO, C., BIREMBAUT, P., TRAN, N. L., DITTMAR, G. & VALLAR, L. 2017. RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. *BMC Genomics,* 18**,** 443.

NESVIZHSKII, A. I. 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics,* 73**,** 2092-123.

NG, D., MARTI, G. E., FONTAINE, L., TORO, J. R., CAPORASO, N. & GOLDIN, L. R. 2006. High-density mapping and follow-up studies on chromosomal regions 1, 3, 6, 12, 13 and 17 in 28 families with chronic lymphocytic leukaemia. *British Journal of Haematology,* 133**,** 59-61.

NG, D., TOURE, O., FONTAINE, L., MCMASTER, M. L., GOLDIN, L. R., CAPORASO, N. & TORO, J. R. 2007a. No association of ARLTS1 polymorphisms and risk for familial chronic lymphocytic leukaemia. *Br J Haematol,* 137**,** 173-5.

NG, D., TOURE, O., WEI, M. H., ARTHUR, D. C., ABBASI, F., FONTAINE, L., MARTI, G. E., FRAUMENI, J. F., JR., GOLDIN, L. R., CAPORASO, N. & TORO, J. R. 2007b. Identification of a novel chromosome region, 13q21.33-q22.2, for susceptibility genes in familial chronic lymphocytic leukemia. *Blood,* 109**,** 916-25.

NGUYEN-KHAC, F., CHEVRET, S., ROUX, C., TOUZEAU, C., COSSON, A., CYMBALISTA, F., FEUGIER, P., LEPRÊTRE, S., BENE, M.-C. & RAYNAUD, S. 2013. Mutational and cytogenetic analyses Of 177 CLL patients with trisomy 12: a retrospective study of the CLL french intergroup. *Blood,* 122**,** 4144-4144.

NIETERS, A., BECKMANN, L., DEEG, E. & BECKER, N. 2006. Gene polymorphisms in Toll-like receptors, interleukin-10, and interleukin-10 receptor alpha and lymphoma risk. *Genes Immun,* 7**,** 615-24.

NOVAK, A. J., SLAGER, S. L., FREDERICKSEN, Z. S., WANG, A. H., MANSKE, M. M., ZIESMER, S., LIEBOW, M., MACON, W. R., DILLON, S. R., WITZIG, T. E., CERHAN, J. R. & ANSELL, S. M. 2009. Genetic variation in B-cell-activating factor is associated with an increased risk of developing B-cell non-Hodgkin lymphoma. *Cancer Res,* 69**,** 4217-24.

NUCKEL, H., FREY, U. H., BAU, M., SELLMANN, L., STANELLE, J., DURIG, J., JOCKEL, K. H., DUHRSEN, U. & SIFFERT, W. 2007. Association of a novel regulatory polymorphism (-938C>A) in the BCL2 gene promoter with disease progression and survival in chronic lymphocytic leukemia. *Blood,* 109**,** 290-7.

NUCKEL, H., FREY, U. H., DURIG, J., DUHRSEN, U. & SIFFERT, W. 1816. Methylenetetrahydrofolate reductase (MTHFR) gene 677C>T and 1298A>C polymorphisms are associated with differential apoptosis of leukemic B cells in vitro and disease progression in chronic lymphocytic leukemia. *Leukemia,* 18**,** 1816-23.

NUCKEL, H., FREY, U. H., DURIG, J., DUHRSEN, U. & SIFFERT, W. 2004. 1513A/C polymorphism in the P2X7 receptor gene in chronic lymphocytic leukemia: absence of correlation with clinical outcome. *Eur J Haematol,* 72**,** 259-63.

NUCKEL, H., FREY, U. H., SELLMANN, L., BAU, M., DURIG, J., DUHRSEN, U. & SIFFERT, W. 2006. Bax gene G(-248)A promoter polymorphism and chronic lymphocytic leukemia: lack of association with incidence, disease stage and progression-free survival. *Leukemia,* 20.

NÜCKEL, H., HÜTTMANN, A., KLEIN-HITPASS, L., SCHROERS, R., FÜHRER, A., SELLMANN, L., DÜHRSEN, U. & DÜRIG, J. 2006. Lipoprotein lipase expression is a novel prognostic factor in B-cell chronic lymphocytic leukemia. *Leukemia & lymphoma,* 47**,** 1053-1061.

OGRETMEN, B. & HANNUN, Y. A. 2004. Biologically active sphingolipids in cancer pathogenesis and treatment. *Nat Rev Cancer,* 4**,** 604-16.

OLSSON, B., GRUND, S., JERNÅS, M., JACOBSSON, S., CARLSSON, L. & WADENVIK, H. 2008. CD69 as a Surrogate Marker for IgVH Gene Mutation Status in Chronic Lymphocytic Leukaemia (CLL). *Blood,* 112**,** 4160.

OPLUSTIL O'CONNOR, L., RULTEN, S. L., CRANSTON, A. N., ODEDRA, R., BROWN, H., JASPERS, J. E., JONES, L., KNIGHTS, C., EVERS, B., TING, A., BRADBURY, R. H., PAJIC, M., ROTTENBERG, S., JONKERS, J., RUDGE, D., MARTIN, N. M., CALDECOTT, K. W., LAU, A. & O'CONNOR, M. J. 2016. The PARP Inhibitor AZD2461 Provides Insights into the Role of PARP3 Inhibition for Both Synthetic Lethality and Tolerability with Chemotherapy in Preclinical Models. *Cancer Res,* 76**,** 6084-6094.

OPPEZZO, P., VASCONCELOS, Y., SETTEGRANA, C., JEANNEL, D., VUILLIER, F., LEGARFF-TAVERNIER, M., KIMURA, E. Y., BECHET, S., DUMAS, G., BRISSARD, M., MERLE-BERAL, H., YAMAMOTO, M., DIGHIERO, G. & DAVI, F. 2005. The LPL/ADAM29 expression ratio is a novel prognosis indicator in chronic lymphocytic leukemia. *Blood,* 106**,** 650-7.

OSCIER, D. G., ROSE-ZERILLI, M. J., WINKELMANN, N., GONZALEZ DE CASTRO, D., GOMEZ, B., FORSTER, J., PARKER, H., PARKER, A., GARDINER, A., COLLINS, A., ELSE, M., CROSS, N. C., CATOVSKY, D. & STREFFORD, J. C. 2013. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood,* 121**,** 468-75.

OTT, G. 2017. Aggressive B-cell lymphomas in the update of the 4th edition of the World Health Organization classification of haematopoietic and lymphatic tissues: refinements of the classification, new entities and genetic findings. *Br J Haematol,* 178**,** 871-887.

OUERTATANI-SAKOUHI, H., EL-TURK, F., FAUVET, B., CHO, M.-K., KARPINAR, D. P., LE ROY, D., DEWOR, M., ROGER, T., BERNHAGEN, J. & CALANDRA, T. 2010. Identification and characterization of novel classes of macrophage migration inhibitory factor (MIF) inhibitors with distinct mechanisms of action. *Journal of Biological Chemistry,* 285**,** 26581-26598.

PALERMO, G., MAISEL, D., BARRETT, M., SMITH, H., DUCHATEAU-NGUYEN, G., NGUYEN, T., YEH, R. F., DUFOUR, A., ROBAK, T., DORNAN, D., WEISSER, M. & INVESTIGATORS, R. 2015. Gene expression of INPP5F as an independent prognostic marker in fludarabine-based therapy of chronic lymphocytic leukemia. *Blood Cancer J,* 5**,** e353.

PAN, X., ZHAO, J., ZHANG, W. N., LI, H. Y., MU, R., ZHOU, T., ZHANG, H. Y., GONG, W. L., YU, M., MAN, J. H., ZHANG, P. J., LI, A. L. & ZHANG, X. M. 2009. Induction of SOX4 by DNA damage is critical for p53 stabilization and function. *Proc Natl Acad Sci U S A,* 106**,** 3788-93.

PANG, J. Y., COOK, L., SCHWARTZ, S. & WEISS, N. 2002. Incidence of leukemia in Asian migrants to the United States and their descendants. *Cancer Causes & Control,* 13**,** 791-795.

PARIKH, S. A., KAY, N. E. & SHANAFELT, T. D. 2013. Monoclonal B-cell lymphocytosis: update on diagnosis, clinical outcome, and counseling. *Clin Adv Hematol Oncol,* 11**,** 720-9.

PASCHE, B. 2001. Role of transforming growth factor beta in cancer. *Journal of cellular physiology,* 186**,** 153-168.

PASQUALATO, S., RENAULT, L. & CHERFILS, J. 2002. Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep,* 3**,** 1035-41.

PEMBERTON, N. C., PANEESHA, S., HILLER, L., STARCZYNSKI, J., HOOPER, L., PEPPER, C., PRATT, G. & FEGAN, C. 2006. The SDF-1 G > A polymorphism at position 801 plays no role in multiple myeloma but may contribute to an inferior cause-specific survival in chronic lymphocytic leukemia. *Leuk Lymphoma,* 47**,** 1239-44.

PEREZ-CHACON, G., CONTRERAS-MARTIN, B., CUNI, S., ROSADO, S., MARTIN-DONAIRE, T., LOSADA-FERNANDEZ, I., VARGAS, J. A., JORDA, J., ALVAREZ, N., GARCIA-MARCO, J. & PEREZ-ACIEGO, P. 2005. Polymorphism in the CD5 gene promoter in B-cell chronic lymphocytic leukemia and mantle cell lymphoma. *Am J Clin Pathol,* 123**,** 646-50.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20**,** 3551-67.

PERROT, A., PIONNEAU, C., NADAUD, S., DAVI, F., LEBLOND, V., JACOB, F., MERLE-BERAL, H., HERBRECHT, R., BENE, M. C., GRIBBEN, J. G., BAHRAM, S. & VALLAT, L. 2011. A unique proteomic profile on surface IgM ligation in unmutated chronic lymphocytic leukemia. *Blood,* 118**,** e1-15.

PETTITT, A. R., SHERRINGTON, P. D., STEWART, G., CAWLEY, J. C., TAYLOR, A. M. & STANKOVIC, T. 2001. p53 dysfunction in B-cell chronic lymphocytic leukemia: inactivation of ATM as an alternative to TP53 mutation. *Blood,* 98**,** 814-22.

PORTER, C. C. 2016. Germ line mutations associated with leukemias. *Hematology Am Soc Hematol Educ Program,* 2016**,** 302-308.

POUNDS, S. B. 2006. Estimation and control of multiple testing error rates for microarray studies. *Brief Bioinform,* 7**,** 25-36.

PUENTE, X. S., PINYOL, M., QUESADA, V., CONDE, L., ORDONEZ, G. R., VILLAMOR, N., ESCARAMIS, G., JARES, P., BEA, S., GONZALEZ-DIAZ, M., BASSAGANYAS, L., BAUMANN, T., JUAN, M., LOPEZ-GUERRA, M., COLOMER, D., TUBIO, J. M. C., LOPEZ, C., NAVARRO, A., TORNADOR, C., AYMERICH, M., ROZMAN, M., HERNANDEZ, J. M., PUENTE, D. A., FREIJE, J. M. P., VELASCO, G., GUTIERREZ-FERNANDEZ, A., COSTA, D., CARRIO, A., GUIJARRO, S., ENJUANES, A., HERNANDEZ, L., YAGUE, J., NICOLAS, P., ROMEO-CASABONA, C. M., HIMMELBAUER, H., CASTILLO, E., DOHM, J. C., DE SANJOSE, S., PIRIS, M. A., DE ALAVA, E., MIGUEL, J. S., ROYO, R., GELPI, J. L., TORRENTS, D., OROZCO, M., PISANO, D. G., VALENCIA, A., GUIGO, R., BAYES, M., HEATH, S., GUT, M., KLATT, P., MARSHALL, J., RAINE, K., STEBBINGS, L. A., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J., GUT, I., LOPEZ-GUILLERMO, A., ESTIVILL, X., MONTSERRAT, E., LOPEZ-OTIN, C. & CAMPO, E. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature,* 475**,** 101-105.

PUIGGROS, A., BLANCO, G. & ESPINET, B. 2014. Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. *Biomed Res Int,* 2014**,** 435983.

PULTE, D., OLSON, K. E., BROEKMAN, M. J., ISLAM, N., BALLARD, H. S., FURMAN, R. R., OLSON, A. E. & MARCUS, A. J. 2007a. CD39 activity correlates with stage and inhibits platelet reactivity in chronic lymphocytic leukemia. *J Transl Med,* 5**,** 23.

PULTE, E. D., BROEKMAN, M. J., OLSON, K. E., DROSOPOULOS, J. H., KIZER, J. R., ISLAM, N. & MARCUS, A. J. 2007b. CD39/NTPDase-1 activity and expression in normal leukocytes. *Thromb Res,* 121**,** 309-17.

QUESADA, V., RAMSAY, A. J. & LOPEZ-OTIN, C. 2012. Chronic lymphocytic leukemia with SF3B1 mutation. *N Engl J Med,* 366**,** 2530.

RAHMATPANAH, F. B., CARSTENS, S., HOOSHMAND, S. I., WELSH, E. C., SJAHPUTERA, O., TAYLOR, K. H., BENNETT, L. B., SHI, H., DAVIS, J. W., ARTHUR, G. L., SHANAFELT, T. D., KAY, N. E., WOOLDRIDGE, J. E. & CALDWELL, C. W. 2009. Large-scale analysis of DNA methylation in chronic lymphocytic leukemia. *Epigenomics,* 1**,** 39-61.

RAMEZANI-RAD, P., GENG, H., HURTZ, C., CHAN, L. N., CHEN, Z., JUMAA, H., MELNICK, A., PAIETTA, E., CARROLL, W. L. & WILLMAN, C. L. 2013. SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. *Blood,* 121**,** 148-155.

RAMSAY, A. J., QUESADA, V., FORONDA, M., CONDE, L., MARTINEZ-TRILLOS, A., VILLAMOR, N., RODRIGUEZ, D., KWARCIAK, A., GARABAYA, C., GALLARDO, M., LOPEZ-GUERRA, M., LOPEZ-GUILLERMO, A., PUENTE, X. S., BLASCO, M. A., CAMPO, E. & LOPEZ-OTIN, C. 2013. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat Genet,* 45**,** 526-30.

RAVAL, A., TANNER, S. M., BYRD, J. C., ANGERMAN, E. B., PERKO, J. D., CHEN, S. S., HACKANSON, B., GREVER, M. R., LUCAS, D. M., MATKOVIC, J. J., LIN, T. S., KIPPS, T. J., MURRAY, F., WEISENBURGER, D., SANGER, W., LYNCH, J., WATSON, P., JANSEN, M., YOSHINAGA, Y., ROSENQUIST, R., DE JONG, P. J., COGGILL, P., BECK, S., LYNCH, H., DE LA CHAPELLE, A. & PLASS, C. 2007. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell,* 129**,** 879-90.

RAWSTRON, A., HILLMEN, P. & HOULSTON, R. 2004. Clonal lymphocytes in persons without known chronic lymphocytic leukemia (CLL): implications of recent findings in family members of CLL patients. *Semin Hematol,* 41**,** 192-200.

RAWSTRON, A. C. 2002. Monoclonal B lymphocytes with the characteristics of "indolent" chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. *Blood,* 100**,** 635-639.

RAWSTRON, A. C., YUILLE, M. R., FULLER, J., CULLEN, M., KENNEDY, B., RICHARDS, S. J., JACK, A. S., MATUTES, E., CATOVSKY, D., HILLMEN, P. & HOULSTON, R. S. 2002. Inherited predisposition to CLL is detectable as subclinical monoclonal B-lymphocyte expansion. *Blood,* 100**,** 2289-90.

REINART, N., NGUYEN, P. H., BOUCAS, J., ROSEN, N., KVASNICKA, H. M., HEUKAMP, L., RUDOLPH, C., RISTOVSKA, V., VELMANS, T., MUELLER, C., REINERS, K. S., VON STRANDMANN, E. P., KRAUSE, G., MONTESINOS-RONGEN, M., SCHLEGELBERGER, B., HERLING, M., HALLEK, M. & FINGERLE-ROWSON, G. 2013. Delayed development of chronic lymphocytic leukemia in the absence of macrophage migration inhibitory factor. *Blood,* 121**,** 812-21.

REINER, A., YEKUTIELI, D. & BENJAMINI, Y. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics,* 19**,** 368-75.

RENDON, B. E., WILLER, S. S., ZUNDEL, W. & MITCHELL, R. A. 2009. Mechanisms of macrophage migration inhibitory factor (MIF)-dependent tumor microenvironmental adaptation. *Exp Mol Pathol,* 86**,** 180-5.

REYA, T., O'RIORDAN, M., OKAMURA, R., DEVANEY, E., WILLERT, K., NUSSE, R. & GROSSCHEDL, R. 2000. Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity,* 13**,** 15-24.

RIEMANN, K., BECKER, L., STRUWE, H., NUCKEL, H., DUHRSEN, U., ALAKUS, H., WINDE, G., NEUHAUSER, M., RUBBEN, H., SCHMITZ, K. J., WOHLSCHLAEGER, J., SCHMID, K. W. & SIFFERT, W. 2006. No association of the NFKB1 insertion/deletion promoter polymorphism with survival in colorectal and renal cell carcinoma as well as disease progression in B-cell chronic lymphocytic leukemia. *Pharmacogenet Genomics,* 16**,** 783-8.

RODRIGUEZ-VICENTE, A. E., DIAZ, M. G. & HERNANDEZ-RIVAS, J. M. 2013. Chronic lymphocytic leukemia: a clinical and molecular heterogenous disease. *Cancer Genet,* 206**,** 49-62.

ROOHI, A. & HOJJAT-FARSANGI, M. 2017. Recent advances in targeting mTOR signaling pathway using small molecule inhibitors. *J Drug Target,* 25**,** 189-201.

ROSATI, E., SABATINI, R., RAMPINO, G., TABILIO, A., DI IANNI, M., FETTUCCIARI, K., BARTOLI, A., COACCIOLI, S., SCREPANTI, I. & MARCONI, P. 2009. Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood,* 113**,** 856-65.

ROSENWALD, A., ALIZADEH, A. A., WIDHOPF, G., SIMON, R., DAVIS, R. E., YU, X., YANG, L., PICKERAL, O. K., RASSENTI, L. Z., POWELL, J., BOTSTEIN, D., BYRD, J. C., GREVER, M. R., CHESON, B. D., CHIORAZZI, N., WILSON, W. H., KIPPS, T. J., BROWN, P. O. &

STAUDT, L. M. 2001. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med,* 194**,** 1639-47.

ROSKO, A. E., MCCOLL, K. S., ZHONG, F., RYDER, C. B., CHANG, M. J., SATTAR, A., CAIMI, P. F., HILL, B. T., AL-HARBI, S., ALMASAN, A. & DISTELHORST, C. W. 2014. Acidosis Sensing Receptor GPR65 Correlates with Anti-Apoptotic Bcl-2 Family Member Expression in CLL Cells: Potential Implications for the CLL Microenvironment. *J Leuk (Los Angel),* 2.

ROSSI, D., BRUSCAGGIN, A., SPINA, V., RASI, S., KHIABANIAN, H., MESSINA, M., FANGAZIO, M., VAISITTI, T., MONTI, S., CHIARETTI, S., GUARINI, A., DEL GIUDICE, I., CERRI, M., CRESTA, S., DEAMBROGI, C., GARGIULO, E., GATTEI, V., FORCONI, F., BERTONI, F., DEAGLIO, S., RABADAN, R., PASQUALUCCI, L., FOA, R., DALLA-FAVERA, R. & GAIDANO, G. 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood,* 118**,** 6904-8.

ROSSI, D., CERRI, M., DEAMBROGI, C., SOZZI, E., CRESTA, S., RASI, S., DE PAOLI, L., SPINA, V., GATTEI, V., CAPELLO, D., FORCONI, F., LAURIA, F. & GAIDANO, G. 2009. The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness. *Clin Cancer Res,* 15**,** 995-1004.

ROSSI, D., CIARDULLO, C., SPINA, V. & GAIDANO, G. 2013. Molecular bases of chronic lymphocytic leukemia in light of new treatments. *Immunol Lett,* 155**,** 51-5.

ROSSI, D., FANGAZIO, M., RASI, S., VAISITTI, T., MONTI, S., CRESTA, S., CHIARETTI, S., DEL GIUDICE, I., FABBRI, G., BRUSCAGGIN, A., SPINA, V., DEAMBROGI, C., MARINELLI, M., FAMA, R., GRECO, M., DANIELE, G., FORCONI, F., GATTEI, V., BERTONI, F., DEAGLIO, S., PASQUALUCCI, L., GUARINI, A., DALLA-FAVERA, R., FOA, R. & GAIDANO, G. 2012. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood,* 119**,** 2854-62.

ROSSI, S., SHIMIZU, M., BARBAROTTO, E., NICOLOSO, M. S., DIMITRI, F., SAMPATH, D., FABBRI, M., LERNER, S., BARRON, L. L., RASSENTI, L. Z., JIANG, L., XIAO, L., HU, J., SECCHIERO, P., ZAULI, G., VOLINIA, S., NEGRINI, M., WIERDA, W., KIPPS, T. J., PLUNKETT, W., COOMBES, K. R., ABRUZZO, L. V., KEATING, M. J. & CALIN, G. A. 2010. microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. *Blood,* 116**,** 945-52.

RUCHLEMER, R. & POLLIACK, A. 2013. Geography, ethnicity and "roots" in chronic lymphocytic leukemia. *Leuk Lymphoma,* 54**,** 1142-50.

RUDD, M. F., SELLICK, G. S., ALLINSON, R., MATUTES, E., CATOVSKY, D. & HOULSTON, R. S. 2004. MTHFR polymorphisms and risk of chronic lymphocytic leukemia. *Cancer Epidemiol Biomarkers Prev,* 13**,** 2268-70.

RUDD, M. F., SELLICK, G. S., WEBB, E. L., CATOVSKY, D. & HOULSTON, R. S. 2006. Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia. *Blood,* 108**,** 638-44.

SAGATYS, E. M. & ZHANG, L. 2012. Clinical and laboratory prognostic indicators in chronic lymphocytic leukemia. *Cancer Control,* 19**,** 18-25.

SAKAI, A., MARTI, G. E., CAPORASO, N., PITTALUGA, S., TOUCHMAN, J. W., FEND, F. & RAFFELD, M. 2000. Analysis of expressed immunoglobulin heavy chain genes in familial B-CLL. *Blood,* 95**,** 1413-9.

SALMOIRAGHI, S., MONTALVO, M. L., UBIALI, G., TOSI, M., PERUTA, B., ZANGHI, P., OLDANI, E., BOSCHINI, C., KOHLMANN, A., BUNGARO, S., INTERMESOLI, T., TERRUZZI, E., ANGELUCCI, E., CAVATTONI, I., CICERI, F., BASSAN, R., RAMBALDI, A. & SPINELLI, O. 2016. Mutations of TP53 gene in adult acute lymphoblastic leukemia at diagnosis do

not affect the achievement of hematologic response but correlate with early relapse and very poor survival. *Haematologica,* 101**,** e245-8.

SANTOS, L. L. & MORAND, E. F. 2009. Macrophage migration inhibitory factor: a key cytokine in RA, SLE and atherosclerosis. *Clinica Chimica Acta,* 399**,** 1-7.

SARBASSOV, D. D., ALI, S. M. & SABATINI, D. M. 2005. Growing roles for the mTOR pathway. *Curr Opin Cell Biol,* 17**,** 596-603.

SCHAFFNER, C., STILGENBAUER, S., RAPPOLD, G. A., DOHNER, H. & LICHTER, P. 1999. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. *Blood,* 94**,** 748-53.

SCHIFFMANN, S., FERREIROS, N., BIROD, K., EBERLE, M., SCHREIBER, Y., PFEILSCHIFTER, W., ZIEMANN, U., PIERRE, S., SCHOLICH, K., GROSCH, S. & GEISSLINGER, G. 2012. Ceramide synthase 6 plays a critical role in the development of experimental autoimmune encephalomyelitis. *J Immunol,* 188**,** 5723-33.

SCHILHAM, M. W., OOSTERWEGEL, M. A., MOERER, P., YA, J., DE BOER, P. A., VAN DE WETERING, M., VERBEEK, S., LAMERS, W. H., KRUISBEEK, A. M. & CUMANO, A. 1996. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature,* 380**,** 711.

SCHULZ, E., VALENTIN, A., ULZ, P., BEHAM-SCHMID, C., LIND, K., RUPP, V., LACKNER, H., WOLFLER, A., ZEBISCH, A., OLIPITZ, W., GEIGL, J., BERGHOLD, A., SPEICHER, M. R. & SILL, H. 2012. Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms. *J Med Genet,* 49**,** 422-8.

SCHWAMB, J., FELDHAUS, V., BAUMANN, M., PATZ, M., BRODESSER, S., BRINKER, R., CLAASEN, J., PALLASCH, C. P., HALLEK, M., WENDTNER, C. M. & FRENZEL, L. P. 2012. B-cell receptor triggers drug sensitivity of primary CLL cells by controlling glucosylation of ceramides. *Blood,* 120**,** 3978-85.

SCOTT, R. H., DOUGLAS, J., BASKCOMB, L., HUXTER, N., BARKER, K., HANKS, S., CRAFT, A., GERRARD, M., KOHLER, J. A., LEVITT, G. A., PICTON, S., PIZER, B., RONGHE, M. D., WILLIAMS, D., FACTORS ASSOCIATED WITH CHILDHOOD TUMOURS, C., COOK, J. A., PUJOL, P., MAHER, E. R., BIRCH, J. M., STILLER, C. A., PRITCHARD-JONES, K. & RAHMAN, N. 2008a. Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. *Nat Genet,* 40**,** 1329-34.

SCOTT, R. H., DOUGLAS, J., BASKCOMB, L., NYGREN, A. O., BIRCH, J. M., COLE, T. R., CORMIER-DAIRE, V., EASTWOOD, D. M., GARCIA-MINAUR, S., LUPUNZINA, P., TATTON-BROWN, K., BLIEK, J., MAHER, E. R. & RAHMAN, N. 2008b. Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. *J Med Genet,* 45**,** 106-13.

SEGEL, G. B. & LICHTMAN, M. A. 2004. Familial (inherited) leukemia, lymphoma, and myeloma: an overview. *Blood Cells Mol Dis,* 32**,** 246-61.

SELLICK, G., FIELDING, S., QURESHI, M., CATOVSKY, D. & HOULSTON, R. 2008a. Germline mutations in RAD51, RAD51AP1, RAD51B, RAD51C,RAD51D, RAD52 and RAD54L do not contribute to familial chronic lymphocytic leukemia. *Leuk Lymphoma,* 49**,** 130-3.

SELLICK, G. S., BRODERICK, P., FIELDING, S., CATOVSKY, D. & HOULSTON, R. S. 2008b. Lack of a relationship between the common 8q24 variant rs6983267 and risk of chronic lymphocytic leukemia. *Leukemia,* 22**,** 438-439.

SELLICK, G. S., CATOVSKY, D. & HOULSTON, R. S. 2006a. Familial chronic lymphocytic leukemia. *Semin Oncol,* 33**,** 195-201.

SELLICK, G. S., CATOVSKY, D. & HOULSTON, R. S. 2006b. Relationship between ARLTS1 polymorphisms and risk of chronic lymphocytic leukemia. *Leuk Res,* 30**,** 1573-6.

SELLICK, G. S., COLEMAN, R. J., TALABAN, R. V., FLEISCHMANN, C., RUDD, M. F., ALLINSON, R., CATOVSKY, D. & HOULSTON, R. S. 2005a. Germline mutations in Dok1 do not predispose to chronic lymphocytic leukemia. *Leuk Res,* 29**,** 59-61.

SELLICK, G. S., GOLDIN, L. R., WILD, R. W., SLAGER, S. L., RESSENTI, L., STROM, S. S., DYER, M. J., MAURO, F. R., MARTI, G. E., FULLER, S., LYTTELTON, M., KIPPS, T. J., KEATING, M. J., CALL, T. G., CATOVSKY, D., CAPORASO, N. & HOULSTON, R. S. 2007. A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia. *Blood*.

SELLICK, G. S., RUDD, M., EVE, P., ALLINSON, R., MATUTES, E., CATOVSKY, D. & HOULSTON, R. S. 2004. The P2X7 receptor gene A1513C polymorphism does not contribute to risk of familial or sporadic chronic lymphocytic leukemia. *Cancer Epidemiol Biomarkers Prev,* 13**,** 1065-7.

SELLICK, G. S., WADE, R., RICHARDS, S., OSCIER, D. G., CATOVSKY, D. & HOULSTON, R. S. 2008c. Scan of 977 nonsynonymous SNPs in CLL4 trial patients for the identification of genetic variants influencing prognosis. *Blood,* 111**,** 1625-1633.

SELLICK, G. S., WEBB, E. L., ALLINSON, R., MATUTES, E., DYER, M. J., JONSSON, V., LANGERAK, A. W., MAURO, F. R., FULLER, S., WILEY, J., LYTTELTON, M., CALLEA, V., YUILLE, M., CATOVSKY, D. & HOULSTON, R. S. 2005b. A high-density SNP genomewide linkage scan for chronic lymphocytic leukemia-susceptibility loci. *Am J Hum Genet,* 77**,** 420-9.

SENKAL, C. E., PONNUSAMY, S., BIELAWSKI, J., HANNUN, Y. A. & OGRETMEN, B. 2010. Antiapoptotic roles of ceramide-synthase-6-generated C16-ceramide via selective regulation of the ATF6/CHOP arm of ER-stress-response pathways. *Faseb j,* 24**,** 296-308.

SENKO, M. W., REMES, P. M., CANTERBURY, J. D., MATHUR, R., SONG, Q., ELIUK, S. M., MULLEN, C., EARLEY, L., HARDMAN, M. & BLETHROW, J. D. 2013. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical chemistry,* 85**,** 11710-11714.

SHAHJAHANI, M., MOHAMMADIASL, J., NOROOZI, F., SEGHATOLESLAMI, M., SHAHRABI, S., SABA, F. & SAKI, N. 2015. Molecular basis of chronic lymphocytic leukemia diagnosis and prognosis. *Cellular Oncology***,** 1-17.

SHEN, A., HUMPHRIES, C., TUCKER, P. & BLATTNER, F. 1987. Human heavy-chain variable region gene family nonrandomly rearranged in familial chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A,* 84**,** 8563-7.

SHEN, M., ZHENG, T., LAN, Q., ZHANG, Y., ZAHM, S. H., WANG, S. S., HOLFORD, T. R., LEADERER, B., YEAGER, M., WELCH, R., KANG, D., BOYLE, P., ZHANG, B., ZOU, K., ZHU, Y., CHANOCK, S. & ROTHMAN, N. 2006. Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma among women in Connecticut. *Hum Genet,* 119**,** 659-68.

SHUKLA, A., SHUKLA, V. & JOSHI, S. S. 2017. Regulation of MAPK signaling and implications in chronic lymphocytic leukemia. *Leukemia & Lymphoma***,** 1-9.

SHUKLA, V., MA, S., HARDY, R. R., JOSHI, S. S. & LU, R. 2013. A role for IRF4 in the development of CLL. *Blood,* 122**,** 2848-55.

SIEGEL, R. L., MILLER, K. D. & JEMAL, A. 2015. Cancer statistics, 2015. *CA Cancer J Clin,* 65**,** 5-29.

SINGH, R., HARSHMAN, S. W., RUPPERT, A. S., MORTAZAVI, A., LUCAS, D. M., THOMAS-AHNER, J. M., CLINTON, S. K., BYRD, J. C., FREITAS, M. A. & PARTHUN, M. R. 2015. Proteomic profiling identifies specific histone species associated with leukemic and cancer cells. *Clinical proteomics,* 12**,** 22.

SINNER, D., KORDICH, J. J., SPENCE, J. R., OPOKA, R., RANKIN, S., LIN, S.-C. J., JONATAN, D., ZORN, A. M. & WELLS, J. M. 2007. Sox17 and Sox4 differentially regulate β-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Molecular and cellular biology,* 27**,** 7802-7815.

SKARRATT, K. K. & FULLER, S. J. 2014. Quantitative real-time PCR eliminates false-positives in colony screening PCR. *Journal of microbiological methods,* 96**,** 99-100.

SKOGSBERG, S., TOBIN, G., KROBER, A., KIENLE, D., THUNBERG, U., ALESKOG, A., KARLSSON, K., LAURELL, A., MERUP, M., VILPO, J., SUNDSTROM, C., ROOS, G., JERNBERG-WIKLUND, H., DOHNER, H., NILSSON, K., STILGENBAUER, S. & ROSENQUIST, R. 2006. The G(-248)A polymorphism in the promoter region of the Bax gene does not correlate with prognostic markers or overall survival in chronic lymphocytic leukemia. *Leukemia,* 20**,** 77-81.

SKOWRONSKA, A., AUSTEN, B., POWELL, J. E., WESTON, V., OSCIER, D. G., DYER, M. J., MATUTES, E., PRATT, G., FEGAN, C., MOSS, P., TAYLOR, M. A. & STANKOVIC, T. 2012. ATM germline heterozygosity does not play a role in chronic lymphocytic leukemia initiation but influences rapid disease progression through loss of the remaining ATM allele. *Haematologica,* 97**,** 142-6.

SLAGER, S. L., GOLDIN, L. R., STROM, S. S., LANASA, M. C., SPECTOR, L. G., RASSENTI, L., LEIS, J. F., CAMP, N. J., KAY, N. E., VACHON, C. M., GLENN, M., WEINBERG, J. B., RABE, K. G., CUNNINGHAM, J. M., ACHENBACH, S. J., HANSON, C. A., MARTI, G. E., CALL, T. G., CAPORASO, N. E. & CERHAN, J. R. 2010. Genetic susceptibility variants for chronic lymphocytic leukemia. *Cancer Epidemiol Biomarkers Prev,* 19**,** 1098-102.

SLAGER, S. L., RABE, K. G., ACHENBACH, S. J., VACHON, C. M., GOLDIN, L. R., STROM, S. S., LANASA, M. C., SPECTOR, L. G., RASSENTI, L. Z., LEIS, J. F., CAMP, N. J., GLENN, M., KAY, N. E., CUNNINGHAM, J. M., HANSON, C. A., MARTI, G. E., WEINBERG, J. B., MORRISON, V. A., LINK, B. K., CALL, T. G., CAPORASO, N. E. & CERHAN, J. R. 2011. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood,* 117**,** 1911-6.

SLAGER, S. L. & ZENT, C. S. 2014. Genetic risk of chronic lymphocytic leukemia: a tale of two cities. *Leuk Lymphoma,* 55**,** 735-6.

SPEEDY, H. E., DI BERNARDO, M. C., SAVA, G. P., DYER, M. J., HOLROYD, A., WANG, Y., SUNTER, N. J., MANSOURI, L., JULIUSSON, G., SMEDBY, K. E., ROOS, G., JAYNE, S., MAJID, A., DEARDEN, C., HALL, A. G., MAINOU-FOWLER, T., JACKSON, G. H., SUMMERFIELD, G., HARRIS, R. J., PETTITT, A. R., ALLSUP, D. J., BAILEY, J. R., PRATT, G., PEPPER, C., FEGAN, C., ROSENQUIST, R., CATOVSKY, D., ALLAN, J. M. & HOULSTON, R. S. 2014. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet,* 46**,** 56-60.

STANKOVIC, T., STEWART, G. S., FEGAN, C., BIGGS, P., LAST, J., BYRD, P. J., KEENAN, R. D., MOSS, P. A. & TAYLOR, A. M. 2002. Ataxia telangiectasia mutated-deficient B-cell chronic lymphocytic leukemia occurs in pregerminal center cells and results in defective damage response and unrepaired chromosome damage. *Blood,* 99**,** 300-9.

STANKOVIC, T., WEBER, P., STEWART, G., BEDENHAM, T., MURRAY, J., BYRD, P. J., MOSS, P. A. & TAYLOR, A. M. 1999. Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. *Lancet,* 353**,** 26-9.

STARCZYNSKI, J., PEPPER, C., PRATT, G., HOOPER, L., THOMAS, A., HOY, T., MILLIGAN, D., BENTLEY, P. & FEGAN, C. 2003. The P2X7 receptor gene polymorphism 1513 A-->C has no effect on clinical prognostic markers, in vitro sensitivity to fludarabine, Bcl-2 family protein expression or survival in B-cell chronic lymphocytic leukaemia. *Br J Haematol,* 123**,** 66-71.

STARCZYNSKI, J., PEPPER, C., PRATT, G., HOOPER, L., THOMAS, A., MILLIGAN, D., BENTLEY, P. & FEGAN, C. 2005. Common polymorphism G(-248)A in the promoter region of the bax gene results in significantly shorter survival in patients with chronic lymphocytic Leukemia once treatment is initiated. *J Clin Oncol,* 23**,** 1514-21.

STAROSTIK, P., MANSHOURI, T., O'BRIEN, S., FREIREICH, E., KANTARJIAN, H., HAIDAR, M., LERNER, S., KEATING, M. & ALBITAR, M. 1998. Deficiency of the ATM protein expression defines an aggressive subgroup of B-cell chronic lymphocytic leukemia. *Cancer Res,* 58**,** 4552-7.

STEVENSON, F. K. & CALIGARIS-CAPPIO, F. 2004. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood,* 103**,** 4389-95.

STRATI, P., KEATING, M. J., O'BRIEN, S. M., FERRAJOLI, A., BURGER, J., FADERL, S., TAMBARO, F. P., JAIN, N. & WIERDA, W. G. 2014. Outcomes of first-line treatment for chronic lymphocytic leukemia with 17p deletion. *Haematologica,* 99**,** 1350-5.

STRUPAT, K., SCHEIBNER, O. & BROMIRSKI, M. 2016. High-Resolution, Accurate-Mass Orbitrap Mass Spectrometry–Definitions, Opportunities, and Advantages.

SU, X., CHAKRAVARTI, D., CHO, M. S., LIU, L., GI, Y. J., LIN, Y. L., LEUNG, M. L., EL-NAGGAR, A., CREIGHTON, C. J., SURAOKAR, M. B., WISTUBA, I. & FLORES, E. R. 2010. TAp63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature,* 467**,** 986-90.

SUMMERSGILL, B., THORNTON, P., ATKINSON, S., MATUTES, E., SHIPLEY, J., CATOVSKY, D., HOULSTON, R. S. & YUILLE, M. R. 2002. Chromosomal imbalances in familial chronic lymphocytic leukaemia: a comparative genomic hybridisation analysis. *Leukemia,* 16**,** 1229-32.

SUN, B., MALLAMPATI, S., GONG, Y., WANG, D., LEFEBVRE, V. & SUN, X. 2013. Sox4 is required for the survival of pro-B cells. *The Journal of Immunology,* 190**,** 2080-2089.

SUWALSKA, K., PAWLAK, E., KARABON, L., TOMKIEWICZ, A., DOBOSZ, T., URBANIAK-KUJDA, D., KULICZKOWSKI, K., WOLOWIEC, D., JEDYNAK, A. & FRYDECKA, I. 2008. Association studies of CTLA-4, CD28, and ICOS gene polymorphisms with B-cell chronic lymphocytic leukemia in the Polish population. *Hum Immunol,* 69**,** 193-201.

SUZUKI, S., KUMATORI, A., HAAGEN, I. A., FUJII, Y., SADAT, M. A., JUN, H. L., TSUJI, Y., ROOS, D. & NAKAMURA, M. 1998. PU.1 as an essential activator for the expression of gp91(phox) gene in human peripheral neutrophils, monocytes, and B lymphocytes. *Proc Natl Acad Sci U S A,* 95**,** 6085-90.

SVEEN, A., KILPINEN, S., RUUSULEHTO, A., LOTHE, R. A. & SKOTHEIM, R. I. 2016. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene,* 35**,** 2413-27.

SWERDLOW, S. H., CAMPO, E., HARRIS, N. L., JAFFE, E. S., PILERI, S. A., STEIN, H. & THIELE, J. 2017. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, IARC.

SWERDLOW, S. H., CAMPO, E., PILERI, S. A., HARRIS, N. L., STEIN, H., SIEBERT, R., ADVANI, R., GHIELMINI, M., SALLES, G. A., ZELENETZ, A. D. & JAFFE, E. S. 2016. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood,* 127**,** 2375-90.

SWIFT, M., REITNAUER, P. J., MORRELL, D. & CHASE, C. L. 1987. Breast and other cancers in families with ataxia-telangiectasia. *N Engl J Med,* 316**,** 1289-94.

TAM, C. S., SHANAFELT, T. D., WIERDA, W. G., ABRUZZO, L. V., VAN DYKE, D. L., O'BRIEN, S., FERRAJOLI, A., LERNER, S. A., LYNN, A., KAY, N. E. & KEATING, M. J. 2009. De novo deletion 17p13.1 chronic lymphocytic leukemia shows significant clinical heterogeneity: the M. D. Anderson and Mayo Clinic experience. *Blood,* 114**,** 957-64.

TAMBURINI, J., CHAPUIS, N., BARDET, V., PARK, S., SUJOBERT, P., WILLEMS, L., IFRAH, N., DREYFUS, F., MAYEUX, P., LACOMBE, C. & BOUSCARY, D. 2008. Mammalian target of rapamycin (mTOR) inhibition activates phosphatidylinositol 3-kinase/Akt by up-regulating insulin-like growth factor-1 receptor signaling in acute myeloid leukemia: rationale for therapeutic inhibition of both pathways. *Blood,* 111**,** 379-82.

TANG, B., BOTTINGER, E. P., JAKOWLEW, S. B., BAGNALL, K. M., MARIANO, J., ANVER, M. R., LETTERIO, J. J. & WAKEFIELD, L. M. 1998. Transforming growth factor-beta1 is a new form of tumor suppressor with true haploid insufficiency. *Nat Med,* 4**,** 802-7.

TAYLOR, A. M., METCALFE, J. A., THICK, J. & MAK, Y. F. 1996. Leukemia and lymphoma in ataxia telangiectasia. *Blood,* 87**,** 423-38.

TERRIS, B., BLAVERI, E., CRNOGORAC-JURCEVIC, T., JONES, M., MISSIAGLIA, E., RUSZNIEWSKI, P., SAUVANET, A. & LEMOINE, N. R. 2002. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. *Am J Pathol,* 160**,** 1745-54.

THEODOROU, I., ABEL, L., MAURO, F., DUPREY, B., MAGNAC, C., PAYELLE-BROGARD, B., DAVI, F. & DIGHIERO, G. 2002. High occurence of DRB1 11 in chronic lymphocytic leukaemia families. *British journal of haematology,* 119**,** 713-715.

THOREEN, C. C., CHANTRANUPONG, L., KEYS, H. R., WANG, T., GRAY, N. S. & SABATINI, D. M. 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature,* 485**,** 109-13.

THUNBERG, U., TOBIN, G., JOHNSON, A., SODERBERG, O., PADYUKOV, L., HULTDIN, M., KLARESKOG, L., ENBLAD, G., SUNDSTROM, C., ROOS, G. & ROSENQUIST, R. 2002. Polymorphism in the P2X7 receptor gene and survival in chronic lymphocytic leukaemia. *Lancet,* 360**,** 1935-9.

THURGOOD, L. A., CHATAWAY, T. K., LOWER, K. M. & KUSS, B. J. 2017. From genome to proteome: Looking beyond DNA and RNA in chronic lymphocytic leukemia. *Journal of proteomics,* 155**,** 73-84.

TOBIN, G., THUNBERG, U., LAURELL, A., KARLSSON, K., ALESKOG, A., WILLANDER, K., SODERBERG, O., MERUP, M., VILPO, J., HULTDIN, M., SUNDSTROM, C., ROOS, G. & ROSENQUIST, R. 2005. Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. *Haematologica,* 90**,** 465-9.

TRAN, T. C., SINGLETON, C., FRALEY, T. S. & GREENWOOD, J. A. 2005. Cysteine-rich protein 1 (CRP1) regulates actin filament bundling. *BMC Cell Biology,* 6**,** 45.

TU, C., SHENG, Q., LI, J., MA, D., SHEN, X., WANG, X., SHYR, Y., YI, Z. & QU, J. 2015. Optimization of Search Engines and Postprocessing Approaches to Maximize Peptide and Protein Identification for High-Resolution Mass Data. *J Proteome Res,* 14**,** 4662-73.

UNGER, T., NAU, M. M., SEGAL, S. & MINNA, J. D. 1992. p53: a transdominant regulator of transcription whose function is ablated by mutations occurring in human cancer. *EMBO J,* 11**,** 1383-90.

UREN, A. G., PAKUSCH, M., HAWKINS, C. J., PULS, K. L. & VAUX, D. L. 1996. Cloning and expression of apoptosis inhibitory protein homologs that function to inhibit apoptosis and/or bind tumor necrosis factor receptor-associated factors. *Proc Natl Acad Sci U S A,* 93**,** 4974-8.

VAN DONGEN, J. J., LANGERAK, A. W., BRUGGEMANN, M., EVANS, P. A., HUMMEL, M., LAVENDER, F. L., DELABESSE, E., DAVI, F., SCHUURING, E., GARCIA-SANZ, R., VAN KRIEKEN, J. H., DROESE, J., GONZALEZ, D., BASTARD, C., WHITE, H. E., SPAARGAREN, M., GONZALEZ, M., PARREIRA, A., SMITH, J. L., MORGAN, G. J., KNEBA, M. & MACINTYRE, E. A. 2003. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia,* 17**,** 2257-317.

VAN GENDEREN, C., OKAMURA, R. M., FARINAS, I., QUO, R. G., PARSLOW, T. G., BRUHN, L. & GROSSCHEDL, R. 1994. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev,* 8**,** 2691-703.

VAN KRIEKEN, J. H., LANGERAK, A. W., MACINTYRE, E. A., KNEBA, M., HODGES, E., SANZ, R. G., MORGAN, G. J., PARREIRA, A., MOLINA, T. J., CABECADAS, J., GAULARD, P., JASANI, B., GARCIA, J. F., OTT, M., HANNSMANN, M. L., BERGER, F., HUMMEL, M., DAVI, F., BRUGGEMANN, M., LAVENDER, F. L., SCHUURING, E., EVANS, P. A., WHITE, H., SALLES, G., GROENEN, P. J., GAMEIRO, P., POTT, C. & DONGEN, J. J. 2007. Improved reliability of lymphoma diagnostics via PCR-based clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia,* 21**,** 201-6.

VARDI, A., DAGKLIS, A., SCARFO, L., JELINEK, D., NEWTON, D., BENNETT, F., ALMEIDA, J., RODRIGUEZ-CABALLERO, A., ALLGOOD, S., LANASA, M., CORTELEZZI, A., ORLANDI, E., VERONESE, S., MONTILLO, M., RAWSTRON, A., SHANAFELT, T., ORFAO, A., STAMATOPOULOS, K. & GHIA, P. 2013. Immunogenetics shows that not all MBL are equal: the larger the clone, the more similar to CLL. *Blood,* 121**,** 4521-8.

VARDIMAN, J. W., THIELE, J., ARBER, D. A., BRUNNING, R. D., BOROWITZ, M. J., PORWIT, A., HARRIS, N. L., LE BEAU, M. M., HELLSTROM-LINDBERG, E., TEFFERI, A. & BLOOMFIELD, C. D. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood,* 114**,** 937-51.

VASCONCELOS, Y., DE VOS, J., VALLAT, L., REME, T., LALANNE, A. I., WANHERDRICK, K., MICHEL, A., NGUYEN-KHAC, F., OPPEZZO, P., MAGNAC, C., MALOUM, K., AJCHENBAUM-CYMBALISTA, F., TROUSSARD, X., LEPORRIER, M., KLEIN, B., DIGHIERO, G. & DAVI, F. 2005. Gene expression profiling of chronic lymphocytic leukemia can discriminate cases with stable disease and mutated Ig genes from those with progressive disease and unmutated Ig genes. *Leukemia,* 19**,** 2002-5.

VISONE, R., RASSENTI, L. Z., VERONESE, A., TACCIOLI, C., COSTINEAN, S., AGUDA, B. D., VOLINIA, S., FERRACIN, M., PALATINI, J. & BALATTI, V. 2009. Karyotype-specific microRNA signature in chronic lymphocytic leukemia. *Blood,* 114**,** 3872-3879.

VOGELSTEIN, B., PAPADOPOULOS, N., VELCULESCU, V. E., ZHOU, S., DIAZ, L. A., JR. & KINZLER, K. W. 2013. Cancer genome landscapes. *Science,* 339**,** 1546-58.

VOUSDEN, K. H. & LANE, D. P. 2007. p53 in health and disease. *Nat Rev Mol Cell Biol,* 8**,** 275-83.

WAHL, M. C., WILL, C. L. & LÜHRMANN, R. 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell,* 136**,** 701-718.

WAN, Y. & WU, C. J. 2013. SF3B1 mutations in chronic lymphocytic leukemia. *Blood,* 121**,** 4627-34.

WANG, C. & WANG, X. 2013. The role of TP53 network in the pathogenesis of chronic lymphocytic leukemia. *International Journal of Clinical & Experimental Pathology,* 6**,** 1223-9.

WANG, L., LAWRENCE, M. S., WAN, Y., STOJANOV, P., SOUGNEZ, C., STEVENSON, K., WERNER, L., SIVACHENKO, A., DELUCA, D. S., ZHANG, L., ZHANG, W., VARTANOV, A. R., FERNANDES, S. M., GOLDSTEIN, N. R., FOLCO, E. G., CIBULSKIS, K., TESAR, B., SIEVERS, Q. L., SHEFLER, E., GABRIEL, S., HACOHEN, N., REED, R., MEYERSON, M., GOLUB, T. R., LANDER, E. S., NEUBERG, D., BROWN, J. R., GETZ, G. & WU, C. J. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *New England Journal of Medicine,* 365**,** 2497-506.

WANG, Q., WILLIAMSON, M., BOTT, S., BROOKMAN-AMISSAH, N., FREEMAN, A., NARICULAM, J., HUBANK, M. J., AHMED, A. & MASTERS, J. R. 2007. Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer. *Oncogene,* 26**,** 6560-5.

WANG, S. S., CERHAN, J. R., HARTGE, P., DAVIS, S., COZEN, W., SEVERSON, R. K., CHATTERJEE, N., YEAGER, M., CHANOCK, S. J. & ROTHMAN, N. 2006a. Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. *Cancer Res,* 66**,** 9771-80.

WANG, S. S., COZEN, W., SEVERSON, R. K., HARTGE, P., CERHAN, J. R., DAVIS, S., WELCH, R., ROTHMAN, N. & CHANOCK, S. J. 2006b. Cyclin D1 splice variant and risk for non-Hodgkin lymphoma. *Hum Genet,* 120**,** 297-300.

WANG, S. S., DAVIS, S., CERHAN, J. R., HARTGE, P., SEVERSON, R. K., COZEN, W., LAN, Q., WELCH, R., CHANOCK, S. J. & ROTHMAN, N. 2006c. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. *Carcinogenesis,* 27**,** 1828-34.

WANG, X. & LIN, Y. 2008. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin,* 29**,** 1275-88.

WARNES, G. R., BOLKER, B., BONEBAKKER, L., GENTLEMAN, R., HUBER, W., LIAW, A., LUMLEY, T., MAECHLER, M., MAGNUSSON, A., MOELLER, S., SCHWARTZ, M. & VENABLES, B. 2016. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. .

WATANABE, T., ICHIKAWA, A., SAITO, H. & HOTTA, T. 1996. Overexpression of the MDM2 oncogene in leukemia and lymphoma. *Leuk Lymphoma,* 21**,** 391-7, color plates XVI following 5.

WAWRZYNIAK, E., KOTKOWSKA, A., BLONSKI, J. Z., SIEMIENIUK-RYS, M., ZIOLKOWSKA, E., GIANNOPOULOS, K., ROBAK, T. & KORYCKA-WOLOWIEC, A. 2014. Clonal evolution in CLL patients as detected by FISH versus chromosome banding analysis, and its clinical significance. *Eur J Haematol,* 92**,** 91-101.

WELCKER, M., ORIAN, A., JIN, J., GRIM, J. E., HARPER, J. W., EISENMAN, R. N. & CLURMAN, B. E. 2004. The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proc Natl Acad Sci U S A,* 101**,** 9085-90.

WENSVEEN, F. M., DERKS, I. A., VAN GISBERGEN, K. P., DE BRUIN, A. M., MEIJERS, J. C., YIGITTOP, H., NOLTE, M. A., ELDERING, E. & VAN LIER, R. A. 2012. BH3-only protein Noxa regulates apoptosis in activated B cells and controls high-affinity antibody formation. *Blood,* 119**,** 1440-9.

WETERING, M., OOSTERWEGEL, M., NORREN, K. V. & CLEVERS, H. 1993. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *The EMBO journal,* 12**,** 3847-3854.

WHITE, D. T., MCSHEA, K. M., ATTAR, M. A. & SANTY, L. C. 2010. GRASP and IPCEF promote ARF-to-Rac signaling and cell migration by coordinating the association of ARNO/cytohesin 2 with Dock180. *Molecular biology of the cell,* 21**,** 562-571.

WIHLBORG, C., SJOBERG, J., INTAGLIETTA, M., AXDORPH, U., PISA, E. K. & PISA, P. 1999. Tumour necrosis factor-alpha cytokine promoter gene polymorphism in Hodgkin's disease and chronic lymphocytic leukaemia. *Br J Haematol,* 104**,** 346-9.

WILEY, J. S., DAO-UNG, L. P., GU, B. J., SLUYTER, R., SHEMON, A. N., LI, C., TAPER, J., GALLO, J. & MANOHARAN, A. 2002. A loss-of-function polymorphic mutation in the cytolytic P2X7 receptor gene and chronic lymphocytic leukaemia: a molecular study. *Lancet,* 359**,** 1114-9.

WOLF, S., MERTENS, D., PSCHERER, A., SCHROETER, P., WINKLER, D., GRONE, H. J., HOFELE, C., HEMMINKI, K., KUMAR, R., STEINECK, G., DOHNER, H., STILGENBAUER, S. & LICHTER, P. 2006. Ala228 variant of trail receptor 1 affecting the ligand binding site is associated with chronic lymphocytic leukemia, mantle cell lymphoma, prostate cancer, head and neck squamous cell carcinoma and bladder cancer. *Int J Cancer,* 118**,** 1831-5.

WOLFF, D. J., BAGG, A., COOLEY, L. D., DEWALD, G. W., HIRSCH, B. A., JACKY, P. B., RAO, K. W., RAO, P. N., COMMITTEE, A. F. M. P. C. P. & COMMITTEE, A. C. O. M. G. L. Q. A. 2007. Guidance for fluorescence in situ hybridization testing in hematologic disorders. *The Journal of Molecular Diagnostics,* 9**,** 134-143.

WU, W., ZHU, H., FU, Y., SHEN, W., MIAO, K., HONG, M., XU, W., FAN, L., YOUNG, K. H. & LIU, P. 2016. High LEF1 expression predicts adverse prognosis in chronic lymphocytic leukemia and may be targeted by ethacrynic acid. *Oncotarget,* 7**,** 21631.

WURTHNER, J. U., FRANK, D. B., FELICI, A., GREEN, H. M., CAO, Z., SCHNEIDER, M. D., MCNALLY, J. G., LECHLEIDER, R. J. & ROBERTS, A. B. 2001. Transforming growth factor-β receptor-associated protein 1 is a Smad4 chaperone. *Journal of Biological Chemistry,* 276**,** 19495-19502.

XOCHELLI, A., BALIAKAS, P., KAVAKIOTIS, I., AGATHANGELIDIS, A., SUTTON, L. A., MINGA, E., NTOUFA, S., TAUSCH, E., YAN, X. J., SHANAFELT, T., PLEVOVA, K., BOUDJOGRA, M., ROSSI, D., DAVIS, Z., NAVARRO, A., SANDBERG, Y., VOJDEMAN, F. J., SCARFO, L., STAVROYIANNI, N., SUDARIKOV, A., VERONESE, S., TZENOU, T., KARAN-DJURASEVIC, T., CATHERWOOD, M., KIENLE, D., CHATZOULI, M., FACCO, M., BAHLO, J., POTT, C., PEDERSEN, L. B., MANSOURI, L., SMEDBY, K. E., CHU, C. C., GIUDICELLI, V., LEFRANC, M. P., PANAGIOTIDIS, P., JULIUSSON, G., ANAGNOSTOPOULOS, A., VLAHAVAS, I., ANTIC, D., TRENTIN, L., MONTILLO, M., NIEMANN, C., DOHNER, H., LANGERAK, A. W., POSPISILOVA, S., HALLEK, M., CAMPO, E., CHIORAZZI, N., MAGLAVERAS, N., OSCIER, D., GAIDANO, G., JELINEK, D. F., STILGENBAUER, S., CHOUVARDA, I., DARZENTAS, N., BELESSI, C., DAVI, F., HADZIDIMITRIOU, A., ROSENQUIST, R., GHIA, P. & STAMATOPOULOS, K. 2017. Chronic Lymphocytic Leukemia with Mutated IGHV4-34 Receptors: Shared and Distinct Immunogenetic Features and Clinical Outcomes. *Clin Cancer Res,* 23**,** 5292-5301.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. & VISSCHER, P. M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-9.

YIN, Y., STEPHEN, C. W., LUCIANI, M. G. & FAHRAEUS, R. 2002. p53 Stability and activity is regulated by Mdm2-mediated induction of alternative p53 translation products. *Nat Cell Biol,* 4**,** 462-7.

YOSHIDA, T. 2004. Peptide separation by Hydrophilic-Interaction Chromatography: a review. *Journal of Biochemical and Biophysical Methods,* 60**,** 265-280.

YU, J., CHEN, L., CUI, B., WIDHOPF, G. F., 2ND, SHEN, Z., WU, R., ZHANG, L., ZHANG, S., BRIGGS, S. P. & KIPPS, T. J. 2016. Wnt5a induces ROR1/ROR2 heterooligomerization to enhance leukemia chemotaxis and proliferation. *J Clin Invest,* 126**,** 585-98.

YUAN, E., LI, C. M., YAMASHIRO, D. J., KANDEL, J., THAKER, H., MURTY, V. V. & TYCKO, B. 2005. Genomic profiling maps loss of heterozygosity and defines the timing and stage dependence of epigenetic and genetic events in Wilms' tumors. *Mol Cancer Res,* 3**,** 493-502.

YUILLE, M., CONDIE, A., HUDSON, C., KOTE-JARAI, Z., STONE, E., EELES, R., MATUTES, E., CATOVSKY, D. & HOULSTON, R. 2002. Relationship between glutathione S-transferase M1, T1, and P1 polymorphisms and chronic lymphocytic leukemia. *Blood,* 99**,** 4216-8.

ZENZ, T., EICHHORST, B., BUSCH, R., DENZEL, T., HABE, S., WINKLER, D., BUHLER, A., EDELMANN, J., BERGMANN, M., HOPFINGER, G., HENSEL, M., HALLEK, M., DOHNER, H. & STILGENBAUER, S. 2010a. TP53 mutation and survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology,* 28**,** 4473-9.

ZENZ, T., KROBER, A., SCHERER, K., HABE, S., BUHLER, A., BENNER, A., DENZEL, T., WINKLER, D., EDELMANN, J., SCHWANEN, C., DOHNER, H. & STILGENBAUER, S. 2008. Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. *Blood,* 112**,** 3322-9.

ZENZ, T., VOLLMER, D., TRBUSEK, M., SMARDOVA, J., BENNER, A., SOUSSI, T., HELFRICH, H., HEUBERGER, M., HOTH, P., FUGE, M., DENZEL, T., HABE, S., MALCIKOVA, J., KUGLIK, P., TRUONG, S., PATTEN, N., WU, L., OSCIER, D., IBBOTSON, R., GARDINER, A., TRACY, I., LIN, K., PETTITT, A., POSPISILOVA, S., MAYER, J., HALLEK, M., DOHNER, H., STILGENBAUER, S. & EUROPEAN RESEARCH INITIATIVE ON, C. L. L. 2010b. TP53 mutation profile in chronic lymphocytic leukemia: evidence for a disease specific profile from a comprehensive analysis of 268 mutations. *Leukemia,* 24**,** 2072-9.

ZHANG, L. Y., IBBOTSON, R. E., ORCHARD, J. A., GARDINER, A. C., SEEAR, R. V., CHASE, A. J., OSCIER, D. G. & CROSS, N. C. 2003. P2X7 polymorphism and chronic lymphocytic leukaemia: lack of correlation with incidence, survival and abnormalities of chromosome 12. *Leukemia,* 17**,** 2097-100.

ZHANG, X., FANG, A., RILEY, C. P., WANG, M., REGNIER, F. E. & BUCK, C. 2010. Multi-dimensional liquid chromatography in proteomics—A review. *Analytica chimica acta,* 664**,** 101-113.

ZHANG, Y., LAN, Q., ROTHMAN, N., ZHU, Y., ZAHM, S. H., WANG, S. S., HOLFORD, T. R., LEADERER, B., BOYLE, P., ZHANG, B., ZOU, K., CHANOCK, S. & ZHENG, T. 2005. A putative exonic splicing polymorphism in the BCL6 gene and the risk of non-Hodgkin lymphoma. *J Natl Cancer Inst,* 97**,** 1616-8.

ZHAO, M., JIA, W., JIANG, W. G., WANG, P., DU, G., CHENG, S. & SONG, M. 2016. ADAM29 Expression in Human Breast Cancer and its Effects on Breast Cancer Cells In Vitro. *Anticancer Res,* 36**,** 1251-8.

ZHONG, X., CHUNG, A. C., CHEN, H. Y., MENG, X. M. & LAN, H. Y. 2011. Smad3-mediated upregulation of miR-21 promotes renal fibrosis. *J Am Soc Nephrol,* 22**,** 1668-81.

ZHU, W., TRIVEDI, C. M., ZHOU, D., YUAN, L., LU, M. M. & EPSTEIN, J. A. 2009. Inpp5f is a polyphosphoinositide phosphatase that regulates cardiac hypertrophic responsiveness. *Circ Res,* 105**,** 1240-7.

ZUBAREV, R. A. & MAKAROV, A. 2013. Orbitrap mass spectrometry. ACS Publications.

CHAPTER-9: APPENDICES

9.1. List of CLL candidate gene association studies

Table 9-1. List of CLL candidate gene association studies

Reference	Genes	NHL (n)	CLL (n)	Control (n)	No. markers	Notes
Demeter *et al.* (1997)(Demeter et al., 1997)	*TNF,LTA*		73	117	2	Association: *TNFA* promoter SNP and CLL; p=0.006
Wihlborg *et al.* (1999)(Wihlborg et al., 1999)	*TNF*		49	51	1	No association: *TNFA* promoter SNP and CLL
Mainou-Fowler *et al.* (2000)(Mainou-Fowler et al., 2000)	*TNF, LTA, LTB*		76	40	9	No association: high TNF-alpha and TNF-beta producing alleles and CLL
Au *et al.* (2006)(Au et al., 2189)	*TNF*		92	98	1	No association: *TNFA* promoter SNP and CLL in females
Bogunia-Kubik *et al.* (2006)(Bogunia-Kubik et al., 2006)	*TNF*		61	180	1	No association: *TNFA* promoter SNP and CLL
Lemos *et al* (1999)(Lemos et al., 1999)	*CYP2D6, GSTM1, NAT2*	71	13	128	3	Association: *CYP2D6*4* combined CLL, ALL, CML, AML group; p=0.008
Auer *et al* (2001)(Auer et al., 2001)	TNR/11q22-q23		137	97		Association: TNR CCG-repeat length; p=0.036
Klinkov *et al* (2004)(Klinkov et al., 2004)	TNR/11q22-q23		82	146		Association: TNR CCG-repeat length; p=0.02
Machulla *et al* (2001)(Machulla et al., 2001)	*HLA-DR4:DR53:DQ8*		101	157		Association: *HLA-DR4*0103 ; p<0.0025*

Montes-Ares *et al* (2006)(Montes-Ares et al., 2006)	*HLA-C*	98	194		Association: HLA-Cw*16; P=0.0012
Yuille *et al* (2002)(Yuille et al., 2002)	*GSTM1, GSTT1, GSTP1*	138	280	3	Association: GSTM1, GSTT1 null alleles, GSTP1-Ile; p=0.04
Wiley *et al* (2002)(Wiley et al., 2002)	*P2RX7*	36	46	1	Association: *P2RX7* A1513C SNP; p=0.08
Thunberg *et al* (2002)(Thunberg et al., 2002)	*P2RX7*	170	200	1	Association: *P2RX7* A1513C SNP; p=0.03
Starczynski *et al* (2003)(Starczynski et al., 2003)	*P2RX7*	121	95	1	No association: *P2RX7* A1513C SNP
Zhang *et al* (2003)(Zhang et al., 2003)	*P2RX7*	144	348	1	No association: *P2RX7* A1513C SNP
Nuckel *et al* (2004)(Nuckel et al., 2004)	*P2RX7*	111	97	1	No association: P2RX7 A1513C SNP
Sellick *et al* (2004)(Sellick et al., 2004)	*P2RX7*	424	428	1	No association: P2RX7 A1513C SNP
Cabrini *et al* (2005)(Cabrini et al., 2005)	*P2RX7*	62	100	5	No association: P2RX7 nsSNPs
Rudd *et al* (2004)(Rudd et al., 2004)	*MTHFR*	832	886	2	No association: SNPs reducing MTHFR function
Nuckel *et al* (2004)(Nuckel et al.,	*MTHFR*	111	92	2	No association: SNPs

1816)						reducing MTHFR function
Guzowski *et al* (2005)(Guzowski et al., 2005)	*IL-10*		17	25	3	Association : Promoter SNPs increasing IL-10
Starczynski *et al* (2005)(Starczynski et al., 2005)	*BAX*		203	135	1	No association: Promoter SNP frequency in cases and controls. Association with shorter survival
Skogsberg *et al* (2006)(Skogsberg et al., 2006)	*BAX*		463	207	1	No association: Promoter SNP in cases and controls
Nuckel *et al* (2006)(Nuckel et al., 2006)	*BAX*		112	95	1	No association: Promoter SNP in cases and controls
Perez-Chacon *et al* (2005)(Perez-Chacon et al., 2005)	*CD5*		134	102	1	Association : Promoter Microsatellite CA, p=0.005
Zhang *et al* (2005)(Zhang et al., 2005)	*BCL-6*	461	59	535	1	Association: mRNA splicing SNP
Sellick *et al* (2005)(Sellick et al., 2005a)	*DOK1*		140	140		No association: intragenic mutation screening
Calin *et al* (2005)(Calin et al., 2005)	*ARLTS1*		17	475	5	Association: Downregulation by truncating SNP or promoter hypermethylation
Sellick *et al* (2005)(Sellick et al.,	*ARLTS1*		413	471	6	No association:

2006b)						Truncating SNPs
Ng *et al* (2007)(Ng et al., 2007a)	*ARLTS1*		31	100	8	No association: Trunc. SNPs nsSNPs
Chiu *et al.* (2005)(Chiu et al., 2005)	*NAT1 and NAT2, GSTM1, GSTT1, GSTP1*	389		535	8	NAT1 slow genotype associated with slightly increased risk in women [OR = 1.4; 95% confidence interval (CI) = 0.9-2.3], but not in men
Morton *et al.* (2006)(Morton et al., 2006)	*NAT1 and NAT2*	1136		922	10	Association: NAT1*10*/10* (OR = 1.6, 95% CI = 1.04-2.46)
Frey *et al.* (2006)(Frey et al., 2006)	*GNAS*		144		1	GNAS1 T393C status independent prognostic marker in B-CLL.
Jamroziak *et al.* (2006)(Jamroziak et al., 2006)	*ABCB1*		110	201	1	3435CT and 3435TT associated with B-CLL, (OR=1.8, 95% CI = 1.1-3.0)
Kochethu *et al.* (2006)(Kochethu et al., 2006)	*TP53*		200		2	Intron 6 SNP A2/A2 genotype associated with early stage disease, CD38 negativity and a longer time to first treatment
Pemberton *et al.* (2006)(Pemberton et al., 2006)	*SDF-1*					
Riemann *et al.* (2006)(Riemann et al., 2006)	*NFKB1*		72	307	1	Promoter SNP: no association
Wolf *et al.* (2006)(Wolf et al.,	*TNFRS10A*	32	101	137	1	A683C TNFRSF10A is more frequent in CLL, MCL

2006)					
Lan *et al.* (2007)(Lan et al., 2009)	*CASP1, CASP8, CASP9*	1946	1808	12	CASP8 rs6736233 (OR (CG) = 1.21; OR(CC) = 2.13; P trend = .011); CASP9 rs4661636 (OR(CT) = 0.89; OR(TT) = 0.77; P trend = .011); and CASP1 rs1785882 (OR(AT) = 1.12; OR(AA) = 1.30; P trend = .0054) were significantly associated with NHL risk
Nuckel *et al.* (2007)(Nuckel et al., 2007)	*BCL-2*	123	120	1	-938C>A: no association
Majid *et al.* (2008)(Majid et al., 2008)	*BCL-2*	276	100	1	No association: -938C>A SNP
Auer *et al.* (2007)(Auer et al., 2007)	*HD, ATXN1, ATXN3, FRA11B, FRA16A*	Sporadic 140 Familial 68	 90 32	10	Association: CAG and CCG-repeat loci in sporadic (HD, ATXN1, ATXN3, FRA11B) and familial (FRA16A, ATXN1) cases
Broderick *et al.* (2008)(Broderick et al., 2008)	*SMAD7*	1029	3923	1	No association: SNP rs12953717
Sellick *et al.* (2008)(Sellick et al., 2008a)	*RAD51, RAD51AP1, RAD51B, RAD51C, RAD51D, RAD52, RAD54L*	75	188	Mutation screening after PCR and DNA sequencing	No pathogenic mutations identified
Sellick *et al.* (2008)(Sellick et al.,	*8q24 locus*	984	4831	1	No association: rs6983267 SNP

2008b)						
Suwalska *et al.* (2008)(Suwalska et al., 2008)	*CTLA-4, CD28, ICOS*		173	336	5	Association: CTLA-4 –319C>T p=0.003; CD28 17+3T>C p=0.007; ICOS 1554+4(GT)n p=0.009
Begleiter *et al.* (2009)(Begleiter et al., 2009)	*NQO1*		323	299	1	No association: NQO1*2 null genotype
Crowther *et al.* (2009)(Crowther-Swanepoel et al., 2009)	*CXCR4*		188 (familial) 1058 (sporadic)	213 (familial) 1807 (sporadic)	1 SNP Sequencing and mutation screening	No association
Jamroziak *et al.* (2009)(Jamroziak et al., 2009)	*CD38*		252 (study A) 208 (study B)	249 (study A) 254 (study B)	2	Association: rs6449182 (184C>G), study A, p=6x10-12 ;B, p=3x10-13; rs1800561 (418C>T), study A, p=0.014 ; B, p=0.03
Novak *et al.* (2009)(Novak et al., 2009)	*TNFSF13B*	441		475		No association with TNFSF13B promoter SNP

9.2. List of Disease pathway and multiple gene association studies in CLL and NHL

Table 9-2. Disease pathway and multiple gene association studies in CLL and NHL

Reference	Genes	NHL (n)	CLL (n)	Control (n)	No. markers	Notes
De Roos *et al.* (2006)(De Roos et al., 2006)	Metabolic gene variants	1172	148	982	15 SNPs in 11 genes	PON1 L55M AA allele, associated with increased risk of non-Hodgkin's lymphoma (variant homozygote OR, 1.36; 95% CI, 0.96-1.95)
Hill *et al.* (2006)(Hill et al., 2006)	DNA repair and related genes	1172	148	982	34 SNPs in 19 genes	Association with RAG1 820 R/R (odds ratio [OR] = 2.7; 95% confidence interval [CI] = 1.4 to 5.0) c.f. Lys/Lys genotype. Less likely to have the LIG4 (DNA ligase IV) 9 Ile/Ile (OR = 0.5; 95% CI = 0.3 to 0.9) than T/T genotype (P trend = .03) in the non-homologous end joining (NHEJ)/V(D)J pathway.
Shen *et al.* (2006)(Shen et al., 2006)	DNA repair genes	461	59	535	32 SNPs in 18 genes	ERCC5 Asp1104His associated with increased risk of NHL (OR: 1.46; 95% CI: 1.13-1.88; P=0.004), DLBCL (OR: 1.44; 95% CI: 0.99-2.09; P=0.058), and also T cell lymphoma. WRN Cys1367Arg associated with decreased risk of NHL (OR: 0.71; 95% CI: 0.56-0.91; P=0.007)
Nieters *et al.* (2006)(Nieters et al., 2006)	Toll-like receptor, IL-10, IL10RA	710	104	710	11 SNPs in 7 genes	Association: TLR2 16933T>A, decreases risk (OR=0.61)
Wang *et al.* (2006)(Wang et al., 2006b)	Cell cycle genes	1172	148	982	12 SNPs in 7 genes	CCND1 splice variant G870A (rs603965) increased NHL risk (OR(AA) = 1.4, 95% CI = 1.1-1.8, P-trend = 0.021)
Wang *et al.*	Proinflammat	1172	148	982	57 SNPs in 36	Haplotypes in TNF-α and lymphotoxin-α

(2006)(Wang et al., 2006a)	ory and immunoregulatory				genes	(rs1800629, rs361525, rs1799724, rs909253, and rs2239704), increased non-Hodgkin lymphoma risk (OR, 1.31; 95% CI, 1.06-1.63; P = 0.01).
Wang *et al.* (2006)(Wang et al., 2006c)	Oxidative stress pathway	1172	148	982	13 SNPs in 10 genes	NOS2A Ser608Leu, rs2297518) Leu/Leu increased 2-fold risk for NHL (OR=2.2, 95% CI=1.1-4.4) (referent=Ser/Ser and Ser/Leu)
Lan *et al.* (2007)(Lan et al., 2007)	Oxidative stress pathway	461	59	535	14 SNPs in 10 genes	1.7-fold (95% CI = 1.2-2.4, P = 0.0047) increased risk of NHL for homozygous AKR1A1 (IVS5 + 282T > C) SNP. Homozygous for the CYBA (Ex4 + 11C > T) SNP: 1.6-fold (95% CI = 1.1-2.4, P = 0.019) increased risk of NHL
Lim *et al.* (2007)(Lim et al., 2007)	Folate and 1 carbon metabolism	1172	147	982	30 SNPs in 10 genes	Decreased risk of NHL with BHMTEx8+453A>T and increased risk with CBS Ex13+41C>T, FPGS Ex15-263T>C, and SHMT1 Ex12+138C>T and Ex12+236C>T
Cerhan *et al.* (2007)(Cerhan et al., 2007)	Inflammation and immune	458	126	484	7670 SNPs in 1450 genes	ITGB3 L59P (OR = 0.66; 95% CI 0.52-0.85), TLR6 V427A (OR = 5.20; CI 1.77-15.3), SELPLG M264V (OR = 3.20; CI 1.48-6.91), UNC84B G671S (OR = 1.50; CI 1.12-2.00), B3GNT3 H328R (OR = 0.74; CI 0.59-0.93), and BAT2 V1883L (OR = 0.64; CI 0.45-0.90)
Enjuanes *et al.* (2008)(Enjuanes et al., 2008)	Apoptosis and immunoregulation genes		692	738	768	Association with variants in: CCNH, APAF1, IL16, CASP8, NOS2A, CCR7, TNFRSF13B, CASO3, LTA/TNF, BAX, BCL2, CXCL12, CASP10, CCL2, BAK1, 1L1A
Ennas *et al.* (2008)(Ennas	Inflammatory cytokines		40	113	13 SNPs in 9 genes	ILIB-511T protective; IL6-174C and IL1B-511C associated with increased risk

et al., 2008)						
Gra *et al.* (2008)(Gra et al., 2008)	Xenobiotic-metabolizing enzymes	76	83	177	18 SNPs in 10 genes	Association: CYP1A1, GSTM1, CYP2C9
Ganster *et al.* (2009(Ganster et al., 2009))	DNA repair genes		461	461	7 SNPs in 5 genes	Association: ERCC2, XRCC1
Liang *et al.* (2009)(Liang et al., 2009)	Apoptosis, DNA repair, immune, oxidative stress pathways			107	1536 SNPs in 152 genes	Association: IL 10 promoter SNP rs1800890/-3575T>A; TNFSF10 SNPs
Rudd *et al.* (2006)(Rudd et al., 2006)	Multiple candidate cancer genes		992	2707	1467 nsSNPs in 865 genes	Associatition: ATM F858L (OR=2.28, P<0.0001); ATM P1054R (OR=1.68, P=0.006); CHEK2 I157T (OR=14.83, P=0.0008); BRCA2 N372H (OR=1.45, P=0.0032); BUB1B Q349R (OR=1.42, P=0.0038)

9.3. FLOW CYTOMETRY RESULTS

9.3.1. Purity check result

As described early in cell purification 2.2. All purified cells were extracted from heparinised blood by adding RossetteSep cocktail and centrifugation on Ficoll-Histopaque. Unpurified PBMC's were isolated from whole blood at the same time by the same procedure described above except that the RossetteSep antibody cocktail was not used.

To check the efficiency of the separation and the purity of the enriched B cells by flow cytometry, a minimum concentration of 2 x 10^5 cells was re-suspended in 100 µl FACS buffer and stained with fluorescent monoclonal antibodies (See flow cytometry section 2.3).

The result indicated that the purity of the enriched B-CLL cells based on CD20 and CD5 against CD3 were increased > 95% in the RosetteSep tube compared to the PBMC tube for the same CLL sample. The average purity of CD20 against CD3 was raised from 83 % to 96.60%, while the purity percentage of CD5/CD3 reached a purity of 98%. The percentage of CD20/CD5 was 97.27% in the RossetteSep tube, compared to 84.65% in the PBMC tube. These results showed the RosetteSep antibody cocktail purification kit using negative selection technique was an effective method to isolate B-CLL cells with high purity. An example shown in (Figure 9-1)

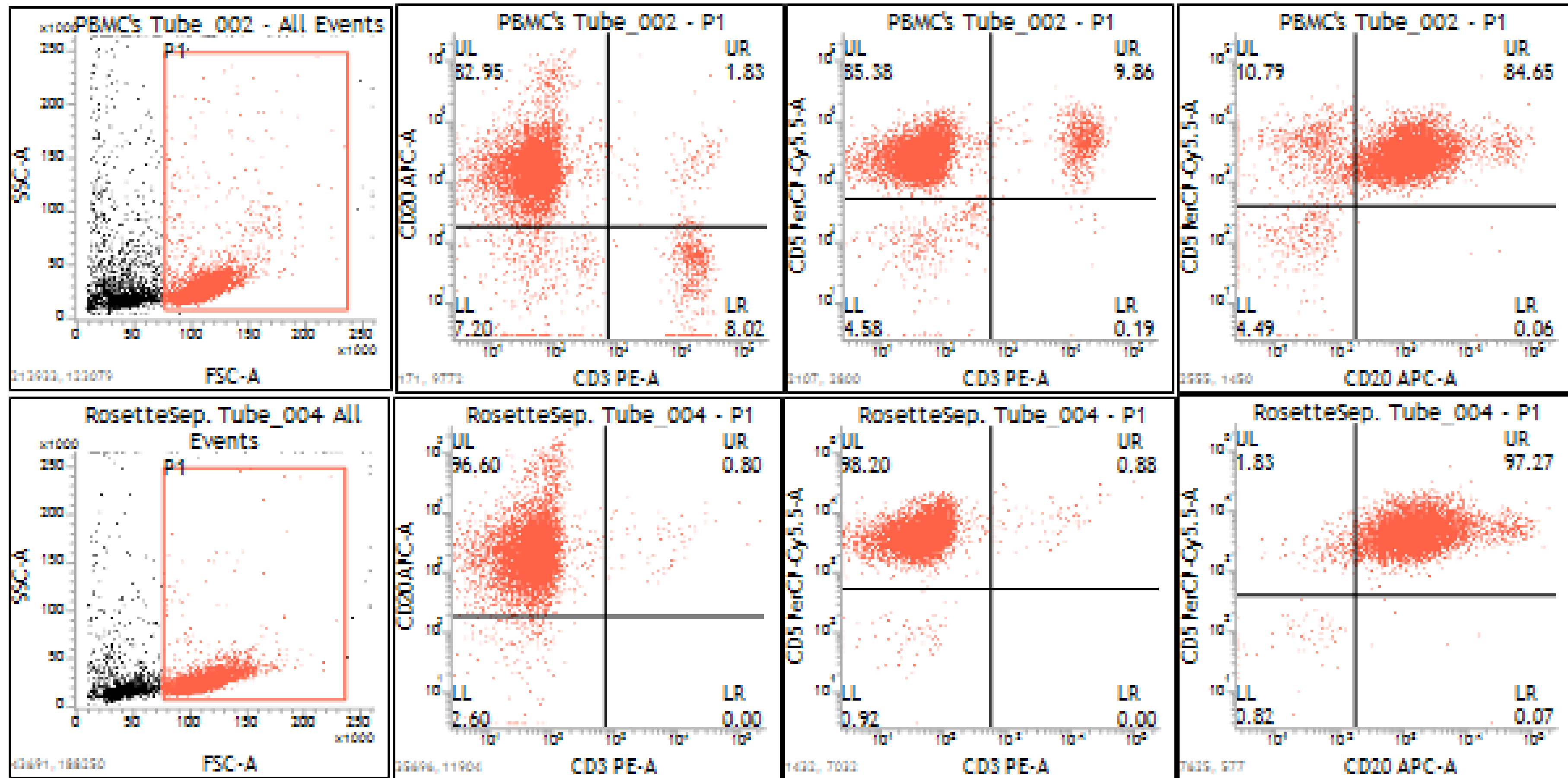

Figure 9-1. Flow cytometer analysis showing percentage B-CLL cells following RosetteSep purification

The analysed PBMC's tubes of CLL patients (Top quadrants) showed an average of 84% B-CLL cells, whereas the lower quadrants exhibited an average of 97% pure B-CLL cells after incubating the same sample with RosetteSep antibody cocktail. In addition, the average percentage of CD3 was reduced from 9% to 0.0% after RosetteSep antibody cocktail.

9.3.2.Screening for B cell lymphocytosis

For B cell lymphocytosis screening, peripheral blood samples were collected from affected and unaffected family members for analysis. Five antibody cocktails were prepared to stain 100 µL of whole blood as described in section 2.3.

Flow cytometry gating was conducted as described previously in section 2.3.2.2.2, using a Becton Dickinson FACSVerse and for each test, 50,000 cells were analysed using BD FACSuite software to identify a monoclonal B-cell lymphocytosis (MBL). Using this protocol, four subjects (III-10, IV-13, IV-17 & IV-18) showed positive expression for surface antigens CD19, CD5, CD23 and low CD20 but not CD10, CD22, CD79b or surface IgM. (Figure 9-2). The expression of light chains (kappa or lambda) could not be determined in all F-MBL cases due to low surface expression.

To evaluate the absolute lymphocytosis, full blood counts were performed for all samples at the Haematology laboratory at Nepean Hospital, Penrith, New South Wales, Australia. Patients with MBL showed normal blood cell count with B cell lymphocytosis below 5.0 $x10^9$/L and did not met the required numerical B cell cut-off to classify them as CLL. The remaining 15 cases were negative for monoclonal B-cell population and had normal blood cell counts.

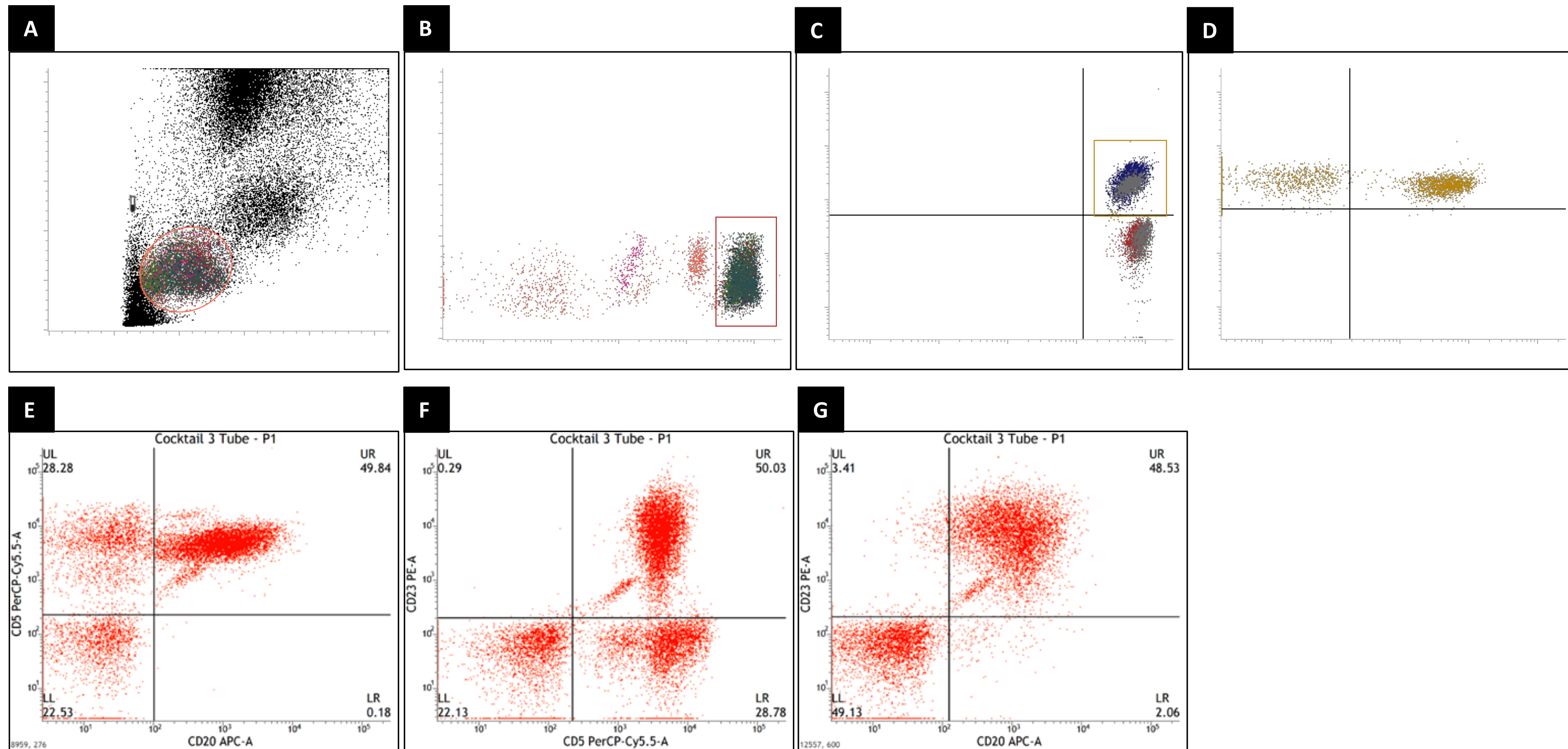

Figure 9-2. Flow cytometry gating strategy for detection of monoclonal B lymphocytosis

(A) First, the lymphocyte region for subset analysis was gated using forward light scatter (FSC) vs. side scatter (SSC). (B) Lymphocyte populations were then analysed using SSC vs. CD45 to distinguish lymphocytes from other cells. Lymphocytes were defined as CD45 bright with low SSC, and further gating on this population was performed to identify B- and T-lymphocytes and NK cells. (C) Shows B-lymphocytes were identified in the 2nd cocktail tube by analysing CD19 vs. CD45, and (D) a gate has been applied (P8) to evaluate CD19 vs. CD3. B-lymphocyte populations were studied in other cocktail tubes by gating the lymphocyte region using FSC vs. SSC, then B-CLL and MBL populations were identified using CD20 vs.CD5 and in these populations the expression of CD10, CD22, CD23, CD38, CD79b, IgM and kappa and lambda light chains were evaluated. (E) Shows the same patient expressed CD5/CD20 positive MBL cells within the lymphocyte population in cocktail 3. (F and G) These MBL cells show expressions of CD23.

9.4. FISH analysis in CLL and MBL cells

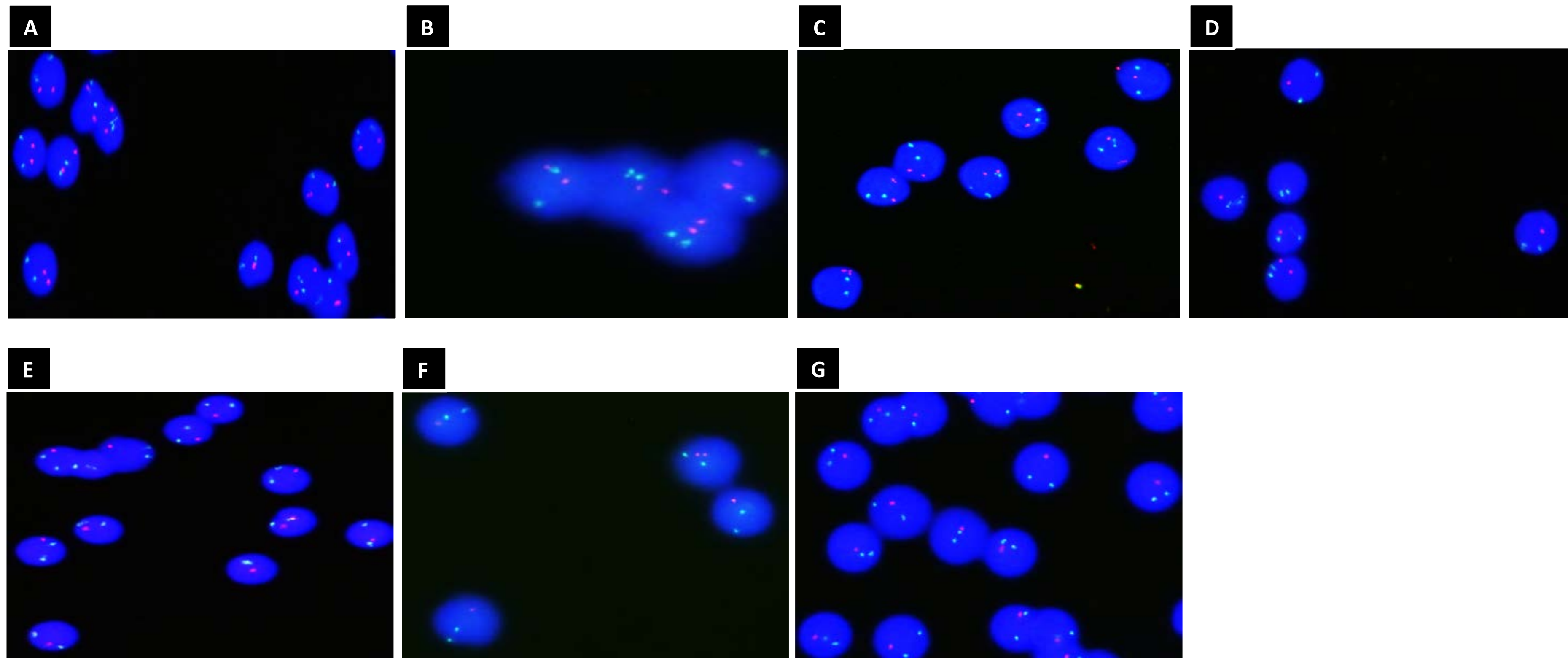

Figure 9-3. FISH analysis in CLL and MBL cells

The FISH analysis in CLL and MBL cells showing examples of normal and abnormal signal patterns for DELU (13q14), CEP12 (chromosome 12 centromere), ATM and TP53. Red signals (13q14 and TP53) are indicated as R and Green signals (chromosome 12 centromere and ATM) as G. (A) shows an example of the normal signal pattern (2R2G) for the 13q14 and centromere of chromosome 12. (B and C) show normal signal patterns (2R2G) for TP53 and ATM. (D and E) show examples of abnormal signal patterns (1R2G) for del (13q14). (F and G) show abnormal signal patterns 1R2G for TP53 deletion.

9.5. Purity of total RNA in PBMCs

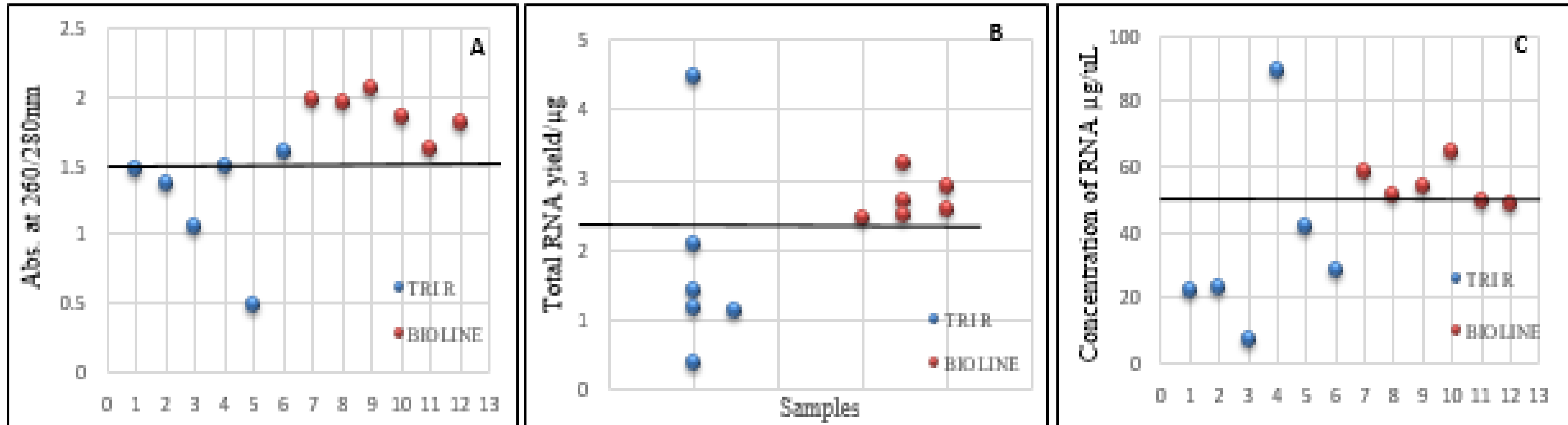

Figure 9-4. Purity of total RNA in PBMCs

UV spectrophotometer analysis shows the quantity and purity of total RNA in PBMCs**.** (A) shows the absorbance ratio at 260/280nm of RNA extracted by spin column method (Bioline) vs TRI kit. (B) The average yield of total RNA by Bioline kit was higher than the obtained yield with TRI kit. (C) Indicates that the concentrations of RNA µg/µL were also higher in samples using the spin column (Bioline) compared to those obtained with the TRI protocol.

9.6. RNA concentrations (ng/µL)

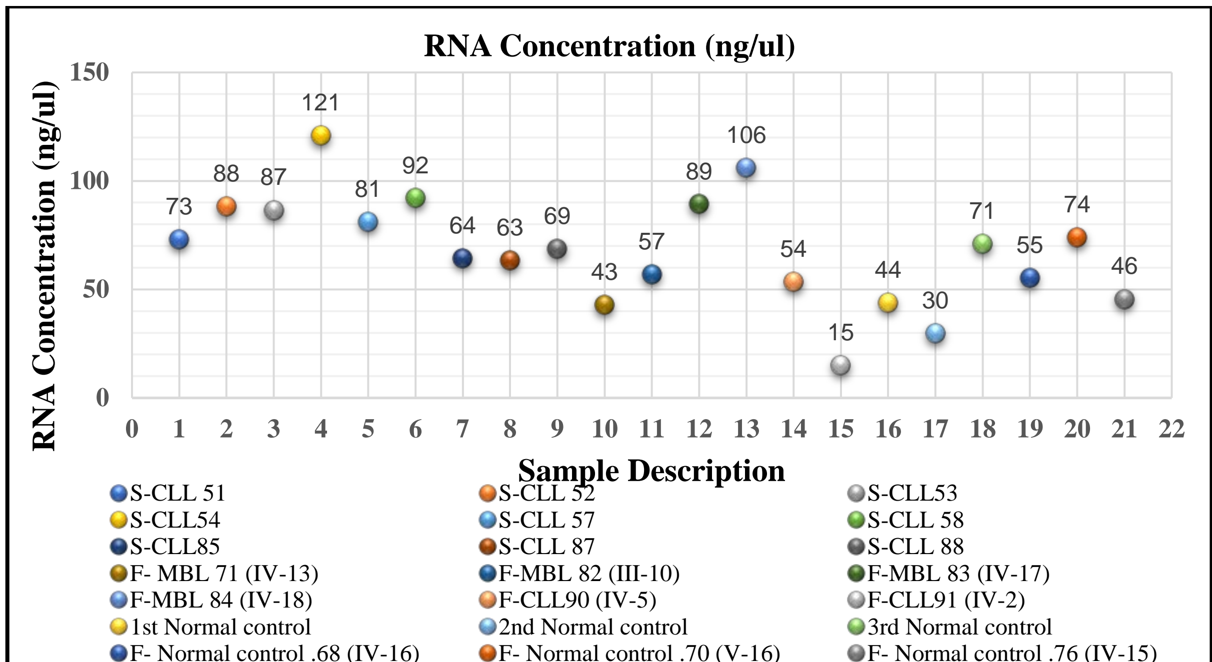

Figure 9-5. RNA concentrations (ng/µL)

9.7. A260/A280 ratio Results

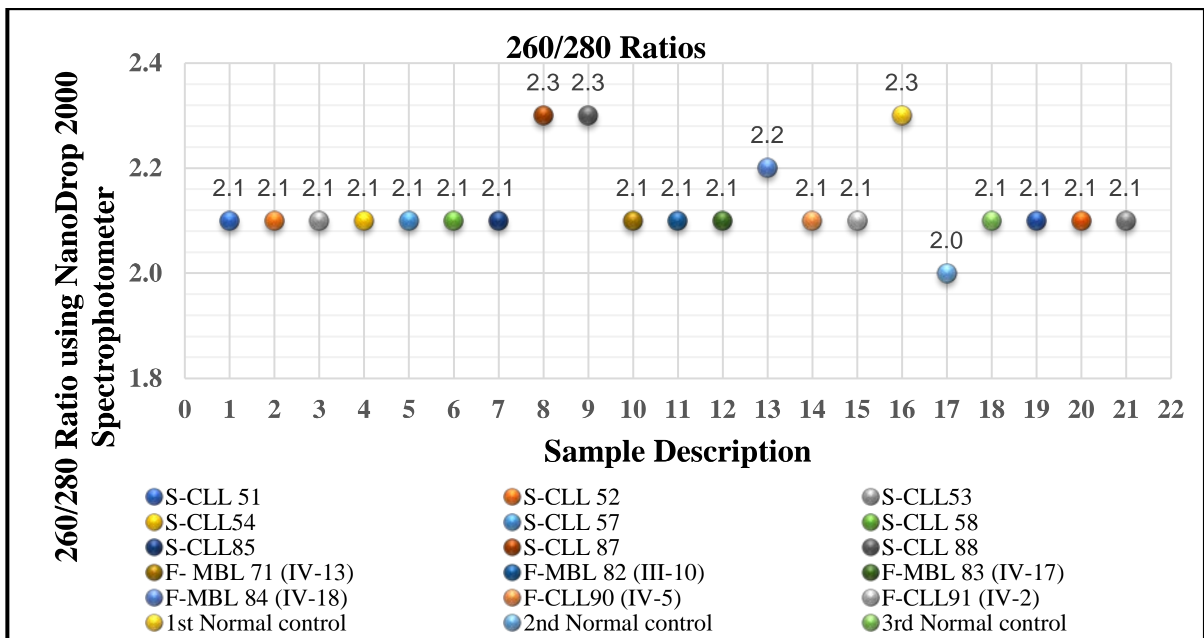

Figure 9-6. A260/A280 ratio Results

All samples in a range above 1.8 at A260/A280 ratio and indicated an acceptable RNA purity

9.8. RNA quality Assessment

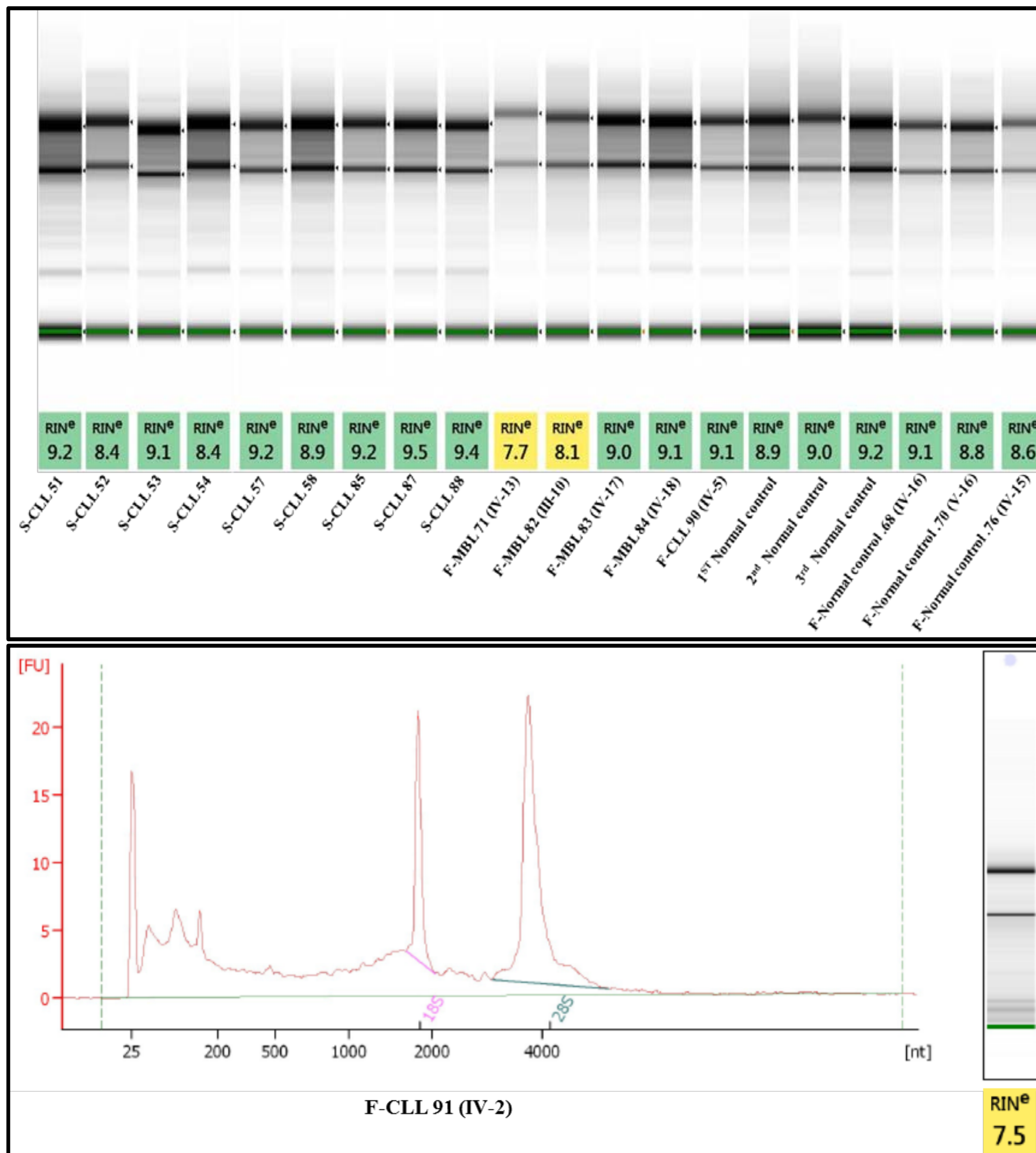

Figure 9-7. RNA quality Assessment

All RNA samples were run on an Agilent 2100 bioanalyzer to determine the ratio of two ribosomal RNAs (28S/18S) and the RNA Integrity Number (RIN). The above figures display the results of all RNA samples. Gel images from the bioanalyzer are shown in the upper figure with the intensity of 28S and 18S rRNA bands clearly indicated. The bottom figure shows the Agilent 2100 bioanalyzer electropherogram on the left side and gel image on the right side of isolated total RNA from F-CLL91 (IV-02). The wavelength of the fluorescence unit (FU) is plotted on the Y-axis, while the size of fragments [nucleotide (nt)] is displayed on the X axis. The resulting electropherogram shows two clear peaks (18S and 28S) and the RNA gel shows two bands, representing the 28S and 18S rRNAs.

9.9. Electropherogram summary results of total RNA

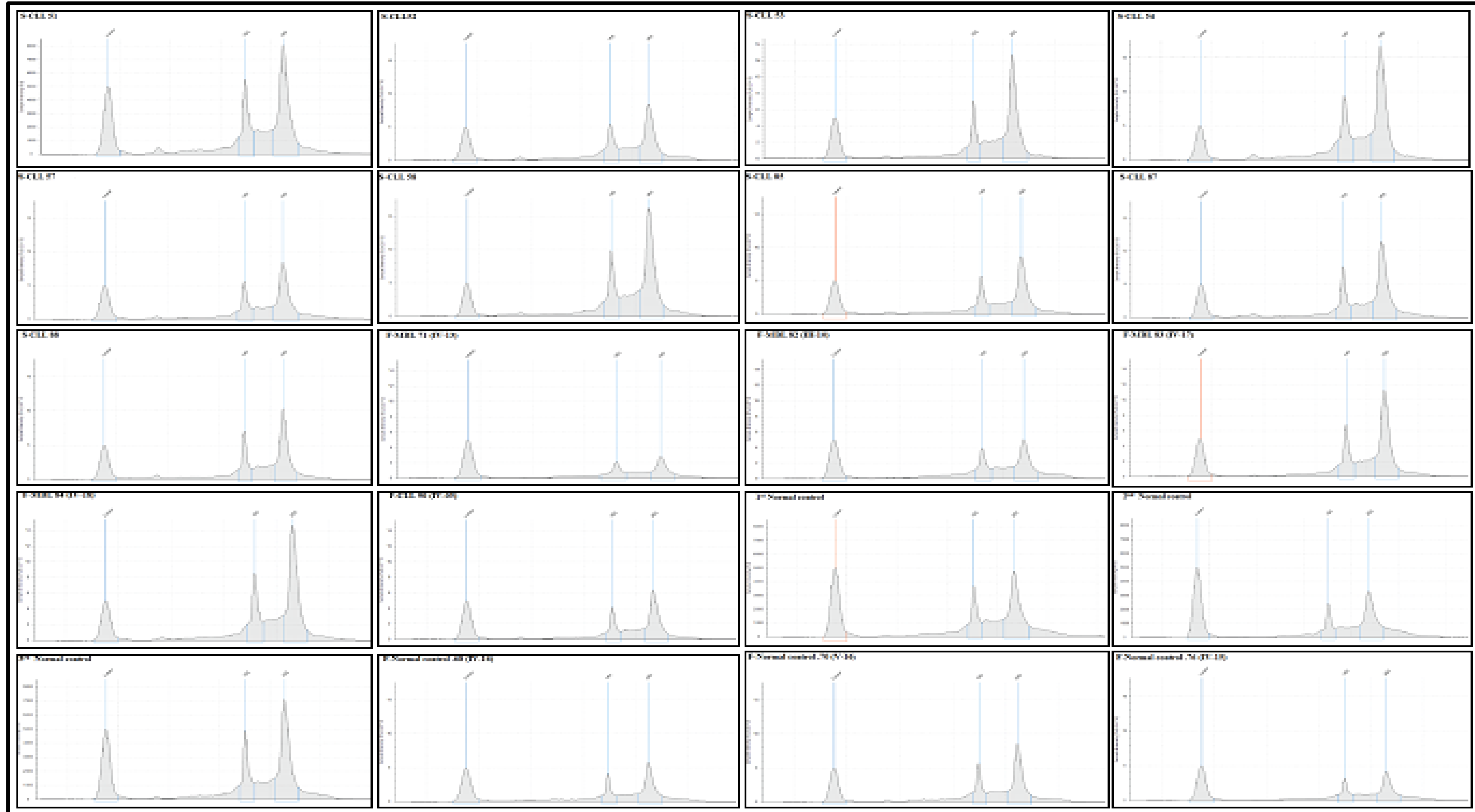

Figure 9-8. Electropherogram summary results of total RNA from all samples included in microarray assay

The 28S and 18S RNA peaks of all submitted RNA samples, showing clearly defined 18S and 28S peaks. All analysed RNA samples had a 28S/18S rRNA ratio > 2 and RIN > 7 and were included in downstream applications.

9.10. Monitoring sample quality

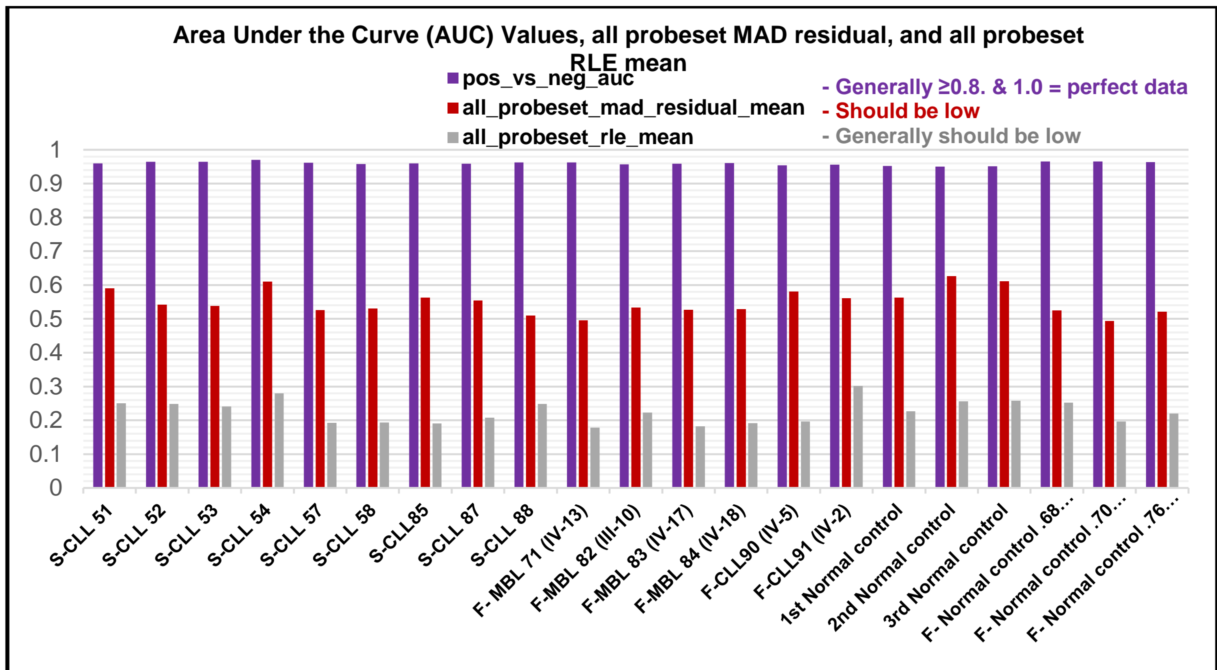

Figure 9-9. Monitoring sample quality

The QC metrics can identify outlier arrays within the data set. To begin the examination, a graph of the most important QC metrics was generated. The *pos_vs_neg_auc* (purple bars) is the area under the curve (AUC) for a receiver operating characteristic (ROC) plot comparing signal values for the positive controls to the negative controls. Typical values range between 0.8 and 0.9, with a value of 1.0 being perfect. *all_probeset_mad_residual_mean* (red bars) is the mean of the median absolute deviation of the residuals over the data set. Ideally, this value should be low. *all_probeset_rle_mean* (grey bars) is the mean over all probesets analyzed of the absolute log expression value relative to all other arrays. This metric should be consistently low, reflecting the low biological variability of the replicates.

All arrays passed Affymetrix recommended QC metrics; *pos_vs_neg_auc* > 0.8, *all_probeset_mad_residual_mea*n and *all_probeset_rle_mean* are low.

9.11. Monitoring signal intensity distributions

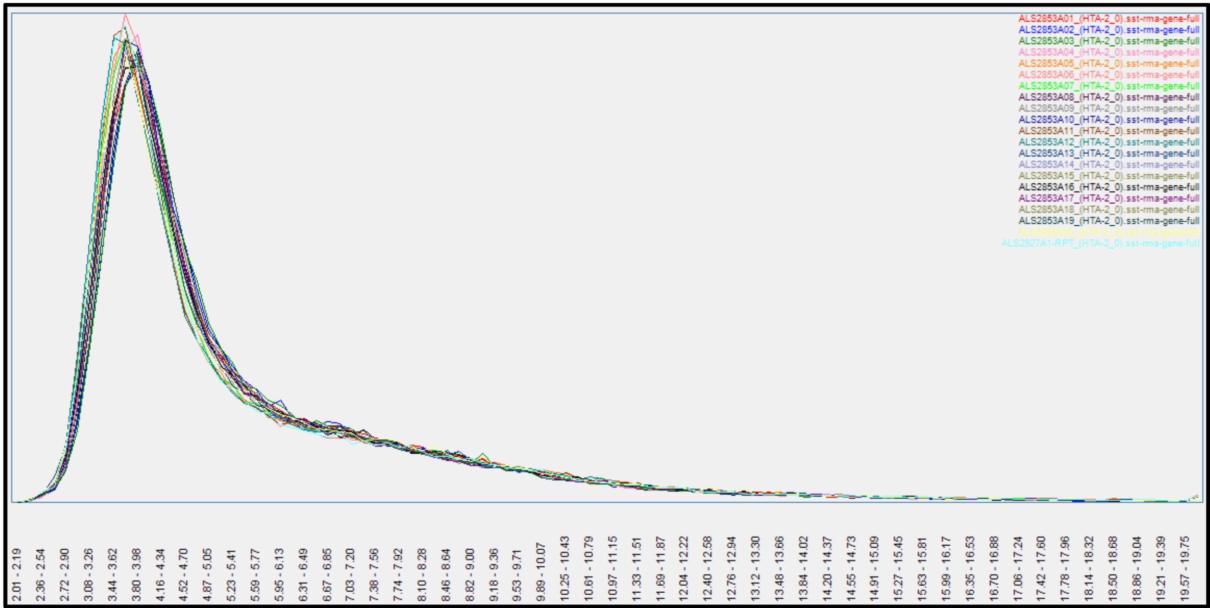

Figure 9-10. Signal intensity histogram distributions

Signal intensity histogram distributions show a normal pattern for each of 21 samples. Signal distributions are tightly clustered and have the same signal distributions.

9.12. The distributions of probe set signals before and after normalization

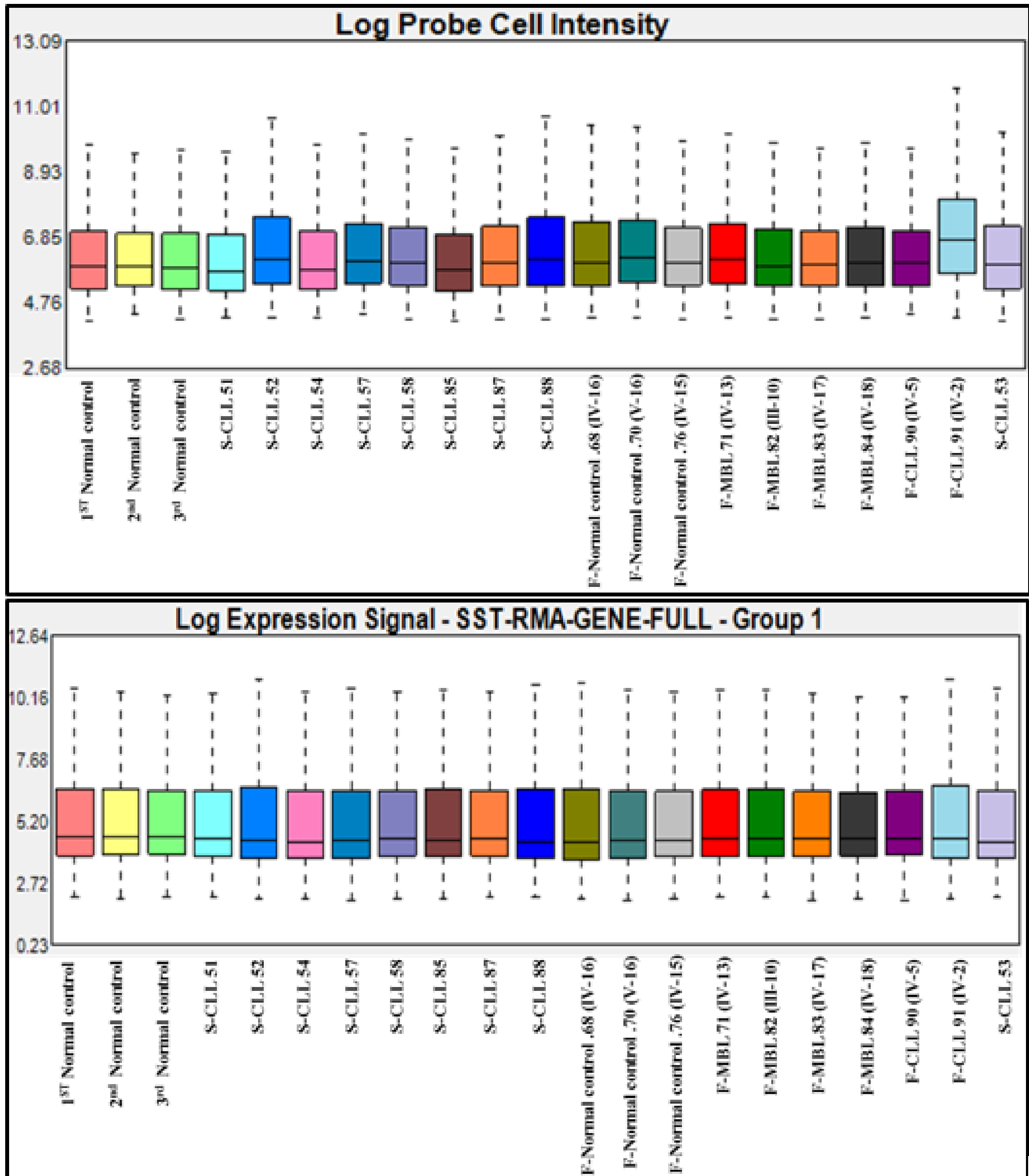

Figure 9-11. Box plots for probe set signal values distributions before and after normalization

Box plots for signal distributions before (upper) and after normalisation (lower) identified any outlier samples. The upper Log probe cell intensity box plots show the probe intensity values for each array. Minor differences in the distributions are expected before normalization. The lower box shows normalised Log Expression signal probe set signal values. Both Box plots display satisfactory signal distributions with no outlier samples identified.

9.13. The relative log expression values before and after normalization

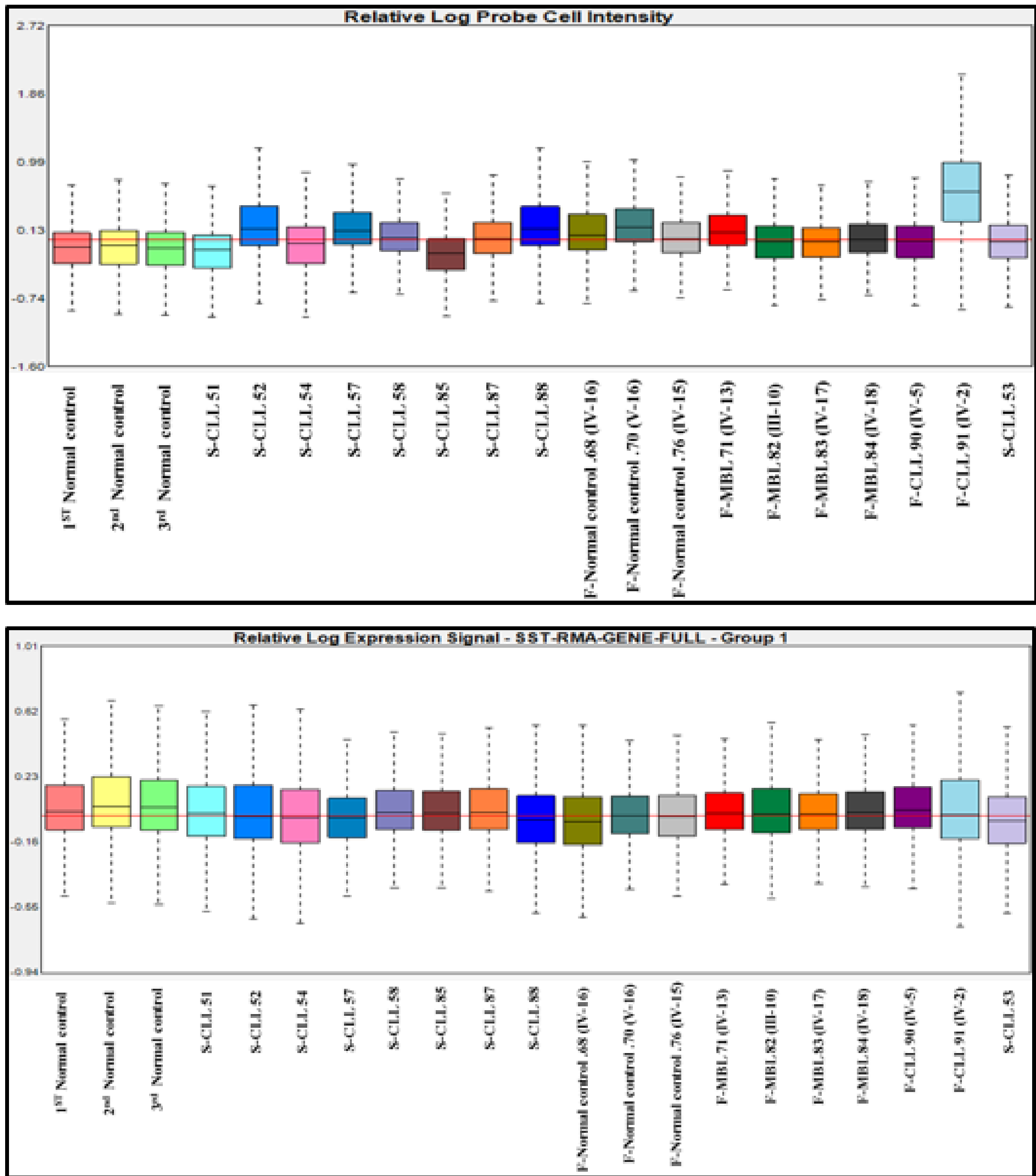

Figure 9-12. Box plots for the relative log expression values before and after normalization

The relative Log Expression (RLE) signal shows the differences of the intensity signals of each probe to the median probe intensity across all arrays in the study. Box plots of RLE before (upper) and after normalisation (lower), for all samples these ratios are ≈ zero and consistently low after normalisation. Therefore, none of these arrays are flagged as outliers.

9.14. Monitoring hybridization efficiency

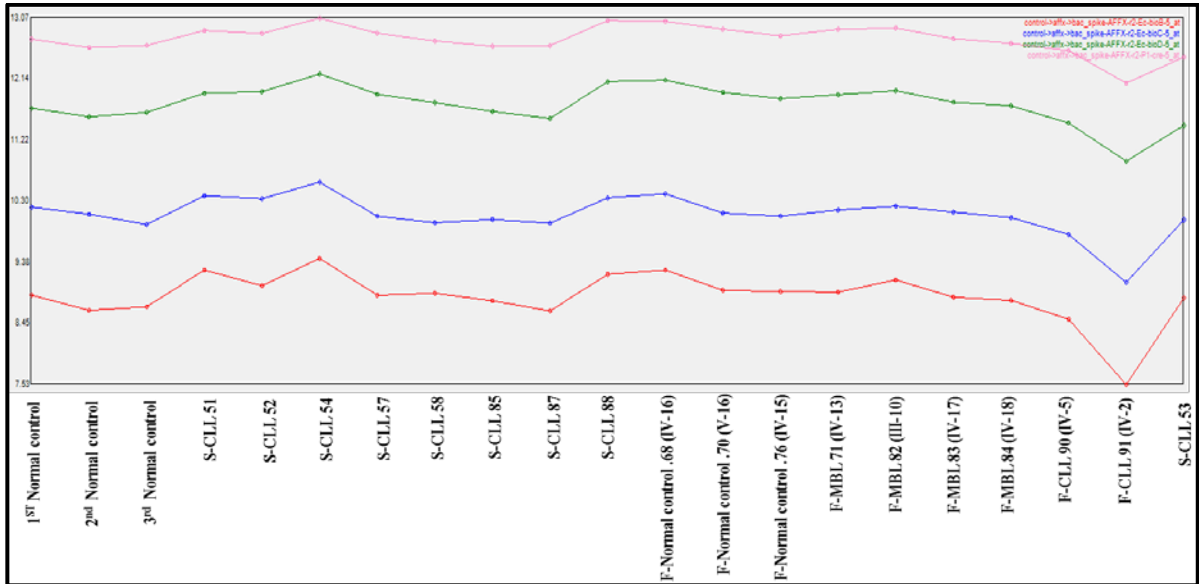

Figure 9-13. Monitoring hybridization efficiency

The efficiency of hybridization steps was determined by the correct rank order of the bacterial spikes positive controls. BioB<BioC<BioD<Cre is observed in the correct order of signal intensities in all samples. These indicate that the hybridization of all samples on gene expression arrays was clearly efficient.

9.15. Monitoring labelling efficiency

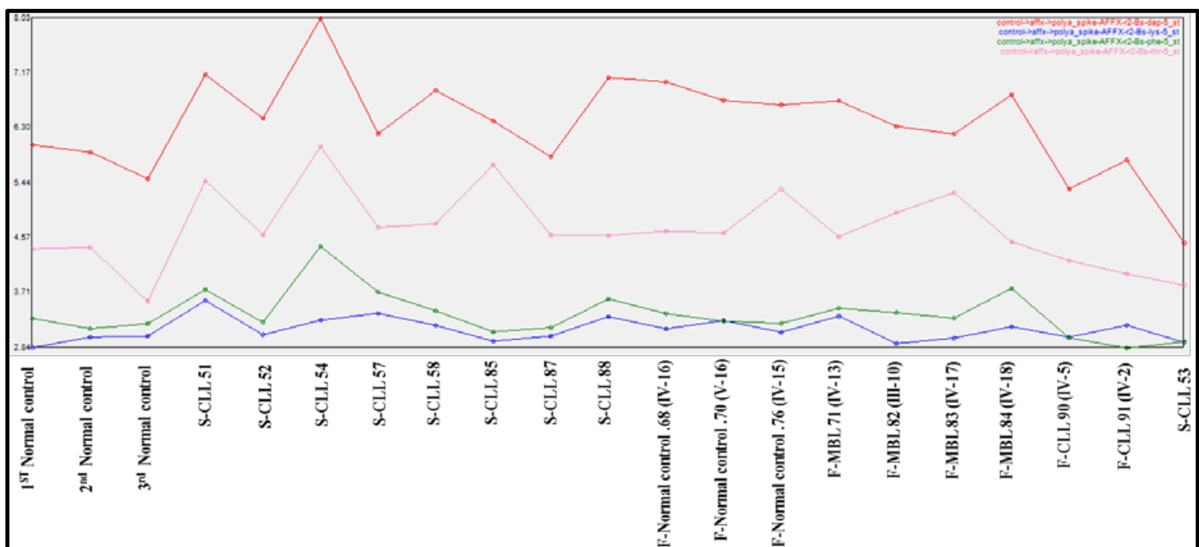

Figure 9-14. Monitoring labelling efficiency

The efficiency of labelling steps was determined by the correct rank order of the Poly-A spike controls Lys < Phe < Thr < Dap. [F-CLL91 (IV-2)] not fall within the expected order for the poly-A spikes (Lys and Phe).