Evaluation of Genomic Aberrations to Define Treatment Response Predictors in Chronic Lymphocytic Leukaemia



Dr Ruth Clifford

A thesis submitted for the Degree of Doctor of Philosophy

at Hertford College, University of Oxford

January 2018

For Eoin, Derval and Niamh

Is scíth mo chrob ón scríbainn

My hand is cramped from penwork.

My quill has a tapered point.

It's bird-mouth issues a blue-dark

Beetle-sparkle of ink

Wisdom keeps welling in streams

From my fine-drawn sallow hand:

Riverrun on the vellum

Of ink from green-skinned holly.

My small runny pen keeps going

Through books, through thick and thin

To enrich the scholars' holdings-

Penwork that cramps my hand.

Taken from a translation of Colum Cille Cecinit, by Seamus Heaney

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work, except where specifically indicated in the text. This dissertation contains approximately 43,000 words and 260 pages.

Ruth Clifford January 2018

ABSTRACT

Chronic lymphocytic leukaemia is a clinically and genetically heterogenous disease. Despite advances in therapeutics, markers to predict response to treatments are lacking, particularly in the setting of chemotherapy resistance. During the course of this thesis I sought to discover, investigate and validate novel candidate gene mutations using NGS. By detailing the genomic landscape of pre-treatment CLL, I evaluated whether novel predictive markers of early progression could be found in the context of CLL clinical trials.

In a longitudinal whole genome sequencing study of CLL, a homozygous *SAMHD1* mutation was revealed as a potential founder event. Furthermore, a pathogenic germline *SAMHD1* mutation identified in a case of Aicardi-Goutieres Syndrome who developed CLL at 24 years lent further support to the hypothesis for a founder mutation status of *SAMHD1* in CLL. By testing clinical trial samples, I have shown that somatic *SAMHD1* mutations are associated with chromosome 20q cnLOH, reduced *SAMHD1* expression and occur in 3-11% of CLL patients. Furthermore evidence that SAMHD1 regulates cell proliferation and engages in the DNA damage response is presented. The work suggests that SAMHD1 may have a function in DNA repair lending support to the idea that *SAMHD1* mutations may lead to genomic instability.

In addition to the *SAMHD1* studies, a comprehensive genomics study to investigate response predictors in frontline treatment setting was undertaken using samples from 250 CLL patients in UK FCR-based clinical trials. Key CLL driver mutations and their interactions were assessed using massively parallel sequencing and high-resolution genome-wide arrays. In univariate analyses, the presence of *TP53* mutation ($p<0.001$), *TP53* deletion ($p<0.001$), *SAMHD1* mutation ($p=0.001$), number of CLL drivers ($p=0.04$) and chromothripsis ($p=0.005$) predicted short progression free survival. Combined trisomy 12 and 19 strongly associated with minimal residual disease negativity and non-progression. In multivariate analysis, both *TP53* mutations ($p<0.001$) and *SAMHD1* mutations ($p=<0.001$) had the strongest predictive value in two independent statistical models. Global measures in inter- and intra-tumour heterogeneity did not impact on clinical outcome.

In conclusion these data highlight *SAMHD1* as a novel CLL driver and importantly, indicate that this approach may identify patients who will benefit from frontline targeted treatments.

ACKNOWLEDGEMENTS

I would firstly like to show my great appreciation to both of my supervisors, Prof. Anna Schuh and Dr. Samantha Knight. Anna, thank you for your encouragement and vision over the last six years. Your expertise in CLL and cutting edge views on genomic technology nurtured the nerd in me and opened up the field of molecular haematology, thanks for this fantastic opportunity. Sam, you have become a close friend and I too now appreciate the need for the finer details. The knowledge you kindly shared relating to genomics smoothed my transition from clinical to research, and late nights viewing Nexus trumped Netflix. In addition, I would like to express my gratitude to my collaborators Prof. Jon Strefford, Dr. Stuart Blakemore, Prof. Monsef Benkirane and Dr. Tania Louis with whom I have worked closely.

I would like to thank all the staff, past and present, of the Molecular Haematology Laboratory, John Radcliffe Hospital, Oxford. Merci beaucoup to my very dear friend, Dr Pauline Robbe,. We have exchanged mentoring roles numerous times over the years, and together we have become keen researchers of cancer genomics. Drs Adam Burns and Adele Timbs, who were with me all the way, I am so grateful for all the mentoring and patience you have shown me. I would like the three of you to appreciate that asking Google to help with IT-related queries is not at all rewarding! I would also like to thank Helene Dreau, Maite Cabes, Sara Ramos, and Drs Pavlos Antoniou, Dimitris Vavoulis and Basile Stamatopoulos who continuously helped and encouraged me during

my years in the laboratory. Many thanks also to my clinical colleagues in Oxford and Limerick. In Oxford, I am especially grateful for the support and guidance from Drs Deborah Hay and Joseph Browning. My good friend and colleague Dr Nicola Curry has always managed gems of advice, and made me laugh, so cheers for that. My consultant colleagues in Limerick have shown the utmost patience and I thank you for this, Drs Maeve Leahy, Denis O' Keeffe and Hilary O' Leary.

There are many friends and family, on the Clifford and Flannery sides, to thank as you have supported my 'never ending story' for some time, too numerous to mention individually. A special mention for my late friend, Martin Vereker, who is much loved. I would like to remember him for his japery and ambition. Our great friends, Gobnait and David O' Shaughnessy, you have shown constant support and are always up for a bit of craic. A particular thanks to Vikki Spiers, Fiona Urquhart, Cristina Navarro and Siobhan Pieroni, during a year when everything came to a head, you made it possible to keep some semblance of a work-life balance.

A huge thank you to my in-laws, Mary and John Flannery, for providing warm welcomes and essential childcare. A very warm thank you to my sisters, Sinead and Sarah, who have provided words of wisdom and shoulders to lean on, you have always shown faith in me. To my brother Shane, thank you for instilling the competitiveness of a second born child in me! The greatest relief lies with my parents, John and Mary Clifford, who have always encouraged me and

facilitated many, many years of education and training. I hope that you can be very proud. Thank you both for everything.

My greatest thanks and love are to my husband Eoin and my two daughters, Derval and Niamh. This work has come about in part from your belief in me, but also through your sacrifices. I hope this encourages you, Derval and Niamh, for your future, and reinforces a desire and confidence to follow your dreams.

Finally I would like to thank all of the CLL patients whose contributions to research are invaluable, and my funding bodies the Anya Sturdy Research Fellowship and the Cancer Research UK Clinical Research Fellowship.

INDIVIDUAL CONTRIBUTIONS

This thesis is the result of my work with the additional contributions listed below.

In Chapter 3, the WGS including data alignment and primary data analysis was undertaken by Illumina, Chesterford Research Park, Little Chesterford, UK. The WES and data aligment was performed by the Core Facility, WTCHG, Oxford as outlined in Chapter 2.15.

Maite Cabes and Dr Pauline Robbe contributed to the DNA extraction and quantification of samples used in this thesis as outlined in Chapter 2 Sections 2.6, 2.7 and 2.10.

In Chapter 3 as part of the functional validation of SAMHD1 was undertaken by Dr Tania Louis under the supervision of Prof. Monsef Benkirane. This was inclusive of that described in Chapter 2 Sections 2.22 to 2.28.

In Chapter 4 in addition to my laboratory work, genome-wide array and MPS was undertaken by Drs Adele Timbs and Pauline Robbe on a subset of the whole-genome arrays. Dr Pauline Robbe performed part of the MPS also. Drs Samantha Knight and Pauline Robbe contributed to the primary reporting of whole-genome SNP array data and Dr Pauline Robbe contributed to the data analysis of the MPS panel.

In Chapter 5 the statistical validation of the clinical outcome findings was wholly performed by Dr Dimitris Vavoulis, as outlined in Chapter 2 Section 2.30. The MPS on CLL4 samples in Chapter 5 Section 5.4.7 was performed by Ewa Stec-Dziedzic and initial data analysis by Dr Stuart Blakemore.

PUBLICATIONS

Publications Related to the work in this Thesis

Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns

Schuh A, Becq J, Humphray S, Alexa A, Burns A, **Clifford R,** Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S, McBride D, Murray L, Menju T, Timbs A, Ross M, Taylor J, Bentley D. *Blood. 2012 Nov 15;120(20):4191-6. doi: 10.1182/blood-2012-05-433540.*

SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage

Clifford R, Louis T, Robbe P, Ackroyd S, Burns A, Timbs AT, Wright Colopy G, Dreau H, Sigaux F, Judde JG, Rotger M, Telenti A, Lin YL, Pasero P, Maelfait J, Titsias M, Cohen DR, Henderson SJ, Ross MT, Bentley D, Hillmen P, Pettitt A, Rehwinkel J, Knight SJ, Taylor JC, Crow YJ, Benkirane M, Schuh A. *Blood. 2014 Feb 13;123(7):1021-31. doi: 10.1182/blood-2013-04-490847.*

Presence of multiple recurrent mutations revealed by targeted NGS confers poor trial outcome of relapsed/refractory CLL

Guièze R, Robbe P, **Clifford R**, de Guibert S, Pereira B, Timbs A, Dilhuydy MS, Cabes M, Ysebaert L, Burns A, Nguyen-Khac F, Davi F, Véronèse L, Combes P, Le Garff-Tavernier M, Leblond V, Merle-Béral H, Alsolami R, Hamblin A, Mason J, Pettitt A, Hillmen P, Taylor J, Knight SJ, Tournilhac O, Schuh A *Blood. 2015 Oct 29;126(18):2110-7. doi: 10.1182/blood-2015-05-647578*.

Towards Response Prediction using Integrated Genomics in Chronic Lymphocytic Leukaemia: Results from frontline treated patients in UK Clinical Trials.

R Clifford, P Robbe, D Vavoulis, S Weller, A Timbs, A Burns, B Stamatopoulos, M Titsias, P Antonius, M Cabes , S Blakemore, R Alsolami, A Hamblin, H Dreau, J Mason, S Henderson, D R Howard, A Chalmers, M Oates, A Pettitt, D Catovsky, J Strefford, C Yau, C Holmes, P Hillmen, J Taylor, S JL Knight, Anna Schuh, *Under review with Haematologica*

Publications unrelated to the Thesis published during DPhil

Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia

Knight SJ, Yau C, **Clifford R,** Timbs AT, Sadighi Akha E, Dréau HM, Burns A, Ciria C, Oscier DG, Pettitt AR, Dutton S, Holmes CC, Taylor J, Cazier JB, Schuh A.*Leukemia. 2012 Jul;26(7):1564-75. doi: 10.1038/leu.2012.13.*

State-of-the-Art Management of Patients Suffering from Chronic Lymphocytic Leukemia

Clifford R, Schuh A. *Clin Med Insights Oncol. 2012;6:165-78. doi: 10.4137/CMO.S6201.*

Impact of isolated germline JAK2V617I mutation on human hematopoiesis

Mead AJ, Chowdhury O, Pecquet C, Dusa A, Woll P, Atkinson D, Burns A, Score J, Rugless M, **Clifford R**, Moule S, Bienz N, Vyas P, Cross N, Gale RE, Henderson S, Constantinescu SN, Schuh A, Jacobsen SE.*Blood. 2013 May 16;121(20):4156-65. doi: 10.1182/blood-2012-05-430926.*

Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations.

Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, **Clifford R**, Sutton LA, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, Höglund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier D, Gaidano G, Schuh A, Davi F, Pott C, Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sjöblom T, Rosenquist R.*Blood. 2016 Feb 25;127(8):1007-16. doi: 10.1182/blood-2015-10-674572*

Transposon mutagenesis reveals fludarabine-resistance mechanisms in chronic lymphocytic leukemia.

Pandzic T, Larsson J, He L, Kundu S, Ban K, Akhtar-Ali M, Hellström AR, Schuh A, **Clifford R**, Blakemore SJ, Strefford JC, Bauman TS, Lopez-Guillermo A, Campo E, Ljungström V, Mansouri L, Rosenquist R, Sjöblom T, Hellström M. *Clin Cancer Res. 2016 Mar 8. pii: clincanres.2903.2015.*

Genomic disruption of the histone methyltransferase *SETD2* in chronic lymphocytic leukemia

H Parker, MJ Rose-Zerilli, M Larrayoz, **R Clifford**, J Edelmann, S Blakemore, J Gibson, J Wang, V Ljungström, TK Wojdacz, T Chaplin, A Roghanian, Z Davis, A Parker, E Tausch, S Ntoufa, S Ramos, P Robbe, R Alsolami, AJ Steele, G Packham, AE Rodríguez-Vicente, L Brown, F McNicholl, F Forconi[1], A Pettitt, P Hillmen, M Dyer, MS Cragg, C Chelala, CC Oakes, R Rosenquist, K Stamatopoulos, S Stilgenbauer, S Knight, A Schuh, D G Oscier, JC Strefford.*Leukemia 20 May 2016; doi: 10.1038/leu.2016.134*

NCRI phase II study of CHOP in combination with ofatumumab in induction and maintenance in newly diagnosed Richter syndrome

Eyre TA, **Clifford R**, Bloor A, Boyle L, Roberts C, Cabes M, Collins GP, Devereux S, Follows G, Fox CP, Gribben J, Hillmen P, Hatton CS, Littlewood TJ, McCarthy H, Murray J, Pettitt AR, Soilleux E, Stamatopoulos B, Love SB, Wotherspoon A, Schuh A., *Br J Haematol. 2016 Jul 5. doi: 10.1111/bjh.14177*

Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia

B Stamatopoulos, A Timbs, D Bruce, T Smith, **R Clifford** , P Robbe , A Burns D V. Vvoulis, L Lopez, P Antoniou, J Mason, H Dreau, A Schuh, Leukemia. *2017 Apr;31(4):837-845. doi: 10.1038/leu.2016.307.*

***EGR2* mutations define a new clinically aggressive entity of chronic lymphocytic leukemia patients**

E Young, D Noerenberg, L Mansouri, V Ljungström, M Frick, LA Sutton, S.J. Blakemore, J Galan-Sousa, K Plevova, P Baliakas, D Rossi, R Clifford, D Roos-Weil, V Navrkalova, B Dörken, C A. Schmitt, K.E. Smedby, G Juliusson, C Belessi, P Panagiotidis, N Chiorazzi, F Davi, A.W. Langerak, David Oscier, A Schuh, G Gaidan, P Ghia, L Fan, O A. Bernard, F Nguyen-Khac, K Stamatopoulos, J Li, S Pospisilova, T Zenz, J.C. Strefford, R Rosenquist, F Damm.Leukemia. *2017 Jul;31(7):1547-1554. doi: 10.1038/leu.2016.359. Epub 2016 Nov 28.*

TABLE OF CONTENTS

DECLARATION V
ABSTRACT VII
ACKNOWLEDGEMENTS IX
INDIVIDUAL CONTRIBUTIONS XIII
PUBLICATIONS XV
TABLE OF CONTENTS XIX
LIST OF FIGURES XXV
LIST OF TABLES XXIX
LIST OF ABBREVIATIONS XXX
Chapter 1 : Introduction 1
1.1 Chronic Lymphocytic Leukaemia 1
1.1.1 Diagnosis 2
1.1.2 Spectrum of disease with CD19/5/23 positive immunophenotype 2
1.1.3 Clinical Presentation 3
1.1.4 Autoimmune phenomena 4
1.1.5 Transformation to high grade lymphoma – Richter's syndrome 4
1.2 Pathogenesis of Chronic Lymphocytic Leukaemia 6
1.2.1 Normal B cell development 6
1.2.2 Haematopoiesis and the bone marrow niche 6
1.2.3 B cell lineage commitment 8
1.2.4 The developing B cell 9
1.2.5 B cell receptor signalling in CLL 11
1.2.6 Apoptosis in CLL 12
1.2.7 Cell of origin 14
1.3 Genomics of Chronic Lymphocytic Leukaemia 15
1.3.1 Overview of genomic profile of CLL 15
1.3.2 Structural variation 15
1.3.3 Deletions of chromosome 13q 16
1.3.4 Deletions of chromosome 11q 18
1.3.5 trisomy 12 18
1.3.6 Deletions of chromosome 17p 20
1.3.7 Structural variants of lower frequency in CLL 21

1.3.8 Copy neutral loss of heterozygosity 24
1.3.9 Genomic Complexity 24
1.3.10 Key driver genes and pathways in CLL 26
1.3.11 DNA damage response pathway: TP53 and ATM proteins 27
1.3.12 NOTCH signalling pathway and NOTCH1 mutations 29
1.3.13 RNA processing pathways 29
1.3.14 NF-Kappa B signalling pathway 31
1.3.15 Intra-clonal heterogeneity defined by CLL drivers 32
1.3.16 Clonal evolution 33
1.4 Staging and Prognosis 34
1.4.1 Staging systems 35
1.4.2 Somatic Hypermutation 35
1.4.3 Surrogate markers of IGHV status 37
1.4.4 Serum markers and the lymphocyte doubling time 37
1.4.5 Genetic Risk Stratification and Response Prediction 38
1.5 CLL Treatment and Outcomes 39
1.5.1 Aims of treatment 40
1.5.2 When to treat 42
1.5.3 Measuring treatment response 42
1.5.4 Cheson criteria 42
1.5.5 Minimal residual disease response assessment 43
1.5.6 Fit ‘Go Go’ patients 43
1.5.7 High-risk patients 44
1.5.8 Allogeneic Transplantation 45
1.5.9 Frail or elderly ‘Slow Go’ patients 46
1.5.10 Novel agents for fit and frail patients 47
1.5.11 Unmet clinical need – early relapses after frontline treatment 49
1.6 Next Generation Sequencing Applications in Cancer 50
1.7 Thesis Aims and Objectives 53
Chapter 2 : Materials and methods 55
2.1 Patient samples 55
2.2 Research Ethics 57
2.3 Ficoll Preparation of Peripheral Blood Mononuclear Cells 57
2.4 Thawing of CLL samples stored in -80°C freezer 58
2.5 Cell purification using fluorescence-activated cell sorting 59
2.6 DNA extraction from CLL samples 59
2.6.1 Phenol-Chloroform DNA extraction 59

2.6.2 Manual kit DNA extraction 60
2.7 DNA extraction from Saliva 62
2.8 RNA extraction 63
2.9 Complementary DNA preparation 64
2.10 Nucleic acid quantification 65
2.10.1 Nanodrop spectrophotometry 65
2.10.2 Qubit fluorometric quantitation of DNA 65
2.11 DNA Sequence Analysis 66
2.11.1 Polymerase Chain Reaction 66
2.11.2 Agarose gel electrophoresis 67
2.11.3 Purification of PCR Product 68
2.11.4 Capillary Sequencing 68
2.11.5 Sequencing Amplification 69
2.11.6 Ethanol Purification 70
2.11.7 DNA sequencing Data Analysis 71
2.12 cDNA sequencing in selected SAMHD1 cases 71
2.13 Quantitative real-time PCR 72
2.14 Whole genome sequencing 73
2.14.1 WGS library generation and sequencing 73
2.14.2 Analysis and Somatic Variant Calling 73
2.14.3 Somatic Single Nucleotide Variant Time-series Analysis 74
2.15 Whole Exome Sequencing - workflow 74
2.16 Targeted Next Generation DNA Sequencing 75
2.16.1 TSCA panel design 75
2.16.2 TSCA methodology 77
2.16.3 TSCA analysis pipeline 81
2.17 Sensitivity analysis 82
2.18 Validation and reproducibility of mutations found using TSCA panel 82
2.18.1 Validation of mutations 82
2.18.2 Reproducibility 82
2.19 Definition of CLL Driver lesion 83
2.20 Genome-wide SNP array 83
2.20.1 Genome-wide SNP array hybridisation overview 84
2.20.2 Genome-wide SNP array hybridisation in detail 84
2.20.3 Genome-wide SNP array analysis overview 85
2.21 Subclonal evaluation 86
2.22 Cell lines and expression constructs 87

2.23 Plasmids 88
2.24 Cell extracts preparation and western blot analysis 88
2.25 Immunofluorescence 88
2.26 Antibodies 89
2.27 Drug treatment and cell proliferation 89
2.28 Cell fractionation and glycerol gradient 90
2.29 Statistical analysis overview 90
2.30 Statistical Validation of the Multivariate Model 91
2.31 Hierarchical model 92
Chapter 3 : Candidate gene selection using whole genome technologies in CLL: the *SAMHD1* story 93
3.1 Introduction 93
3.2 Aims and Objectives 94
3.3 Materials and Methods 95
3.4 Results 95
3.4.1 Whole genome sequencing reveals putative founder mutations 95
3.4.2 Genome-wide arrays refine the minimal deleted region involving the SAMHD1 locus 97
3.4.3 Literature Review of SAMHD1 locus guides further studies 98
3.4.4 Review of patients with congenital AGS and germline SAMHD1 mutations 100
3.4.5 Genomic evaluation of tumour and germline samples from the AGS CLL patient 102
3.4.6 SAMHD1 is recurrently mutated in CLL 102
3.4.7 Expression of SAMHD1 in mutated CLL samples 116
3.4.8 SAMHD1 regulation of cell proliferation and survival 120
3.4.9 SAMHD1 involvement in DNA repair 123
3.4.10 SAMHD1 is recruited to site of DNA damage 126
3.5 Discussion 128
Chapter 4 : Defining the genomic landscape of pre-treatment CLL 135
4.1 Introduction 135
4.2 Aims and Objectives 137
4.3 Materials and Methods 137
4.4 Results 138
4.4.1 Patient characteristics 138
4.4.2 Validation of Next Generation Sequencing method 140
4.4.3 Results of TSCA panel tests: Mutation frequency and distribution 146
4.4.4 Genes with a mutation frequency greater than 10% 149

4.4.5 Genes with a mutation frequency less than 10% 154
4.4.6 CNAs identified using high resolution genome-wide array 164
4.4.7 Integrated sequencing and copy number analysis 178
4.4.8 Quantification of subclones to define samples with intra-tumour heterogeneity 185
4.5 Discussion 187
4.5.1 MPS using TSCA panel is a reliable and sensitive platform 187
4.5.2 Gene mutations and associations from MPS 189
4.5.3 Whole genome array reveals common and novel aberrations in the majority of cases 190
4.5.4 Heterogeneity at 13q locus points towards convergent evolution 191
4.5.5 Integrating genomics reveals CLL driver in 95% of cases 192
4.5.6 Pair-wise associations 192
4.5.7 Clonal and subclonal mutations discovered both on array and sequencing panel 193
4.5.8 Conclusion 194
Chapter 5 : Towards Response Prediction in CLL - Integrated Genomics USING Clinical Trial samples 196
5.1 Introduction 196
5.2 Aims and Objectives 199
5.3 Materials and Methods 200
5.3.1 Patient samples 200
5.3.2 Statistical analysis 200
5.4 Results 201
5.4.1 Patient Characteristics 201
5.4.2 Predictive markers of clinical response 203
5.4.3 Predictive markers of minimal residual disease status 204
5.4.4 Survival analysis reveals high-risk treatment response predictors 207
5.4.5 Mutations alone, not CLL driver burden, independently predict for early progression 226
5.4.6 Statistical validation of the model 228
5.4.7 Validation of high-risk genomics in historical clinical trial 234
5.4.8 Hierarchical model 237
5.5 Discussion 239
Chapter 6 Discussion 249
6.1 Introduction 249
6.2 Summary of thesis and key findings 250
6.3 *SAMHD1* as a driver gene in CLL 251
6.4 SAMHD1 is involved in genomic integrity 254

6.5 Clinical implications of integrated genomics 256
6.6 Global heterogeneity measures do not predict resistance to treatment 258
6.7 Future perspective 261
6.8 Conclusion 263
REFERENCES 264
Appendix 1 List of sequencing primer for Sanger Sequencing validation 331
Appendix 2 List of variants reported by targeted sequencing 335
Appendix 3 List of copy number variants reported by SNP array 343

LIST OF FIGURES

Figure 3.1 WGS of a CLL patient at 5 treatment time-points 96

Figure 3.2 Genome-wide Array across the *SAMHD1* locus 97

Figure 3.3 SAMHD1 mechanism of action 99

Figure 3.4 Analysis of CLL sample from Aicardi-Goutieres syndrome patient 101

Figure 3.5 Example of Sanger sequencing and genome-wide array results from *SAMHD1* mutated samples 103

Figure 3.6 *SAMHD1* mutation frequency across cohorts 104

Figure 3.7 Schematic distribution of *SAMHD1* mutations 106

Figure 3.8 Conservation tracks for all *SAMHD1* mutations 107

Figure 3.9 Mutation association chart of *SAMHD1* mutated samples 111

Figure 3.10 Variant allele frequency of *SAMHD1* mutations 112

Figure 3.11 Example of Genome-wide array results for sample CLL007 114

Figure 3.12 Phylogenetic trees for CLL007 116

Figure 3.13 *SAMHD1* expression in various tissues 118

Figure 3.14 SAMHD1 expression in Primary CLL cells 119

Figure 3.15 SAMHD1 expression in transduced HeLa Cells 121

Figure 3.16 SAMHD1 effects on cell proliferation and cell viability in response to DNA damaging agents 122

Figure 3.17 SAMHD1 expresion from nuclear extracts 125

Figure 3.18 SAMHD1 and DNA damage response 127

Figure 4.1 Distribution of read depth per gene. 141

Figure 4.2 Expected versus Observed VAFs for sample ARC604 142

Figure 4.3 Graph showing correlation between expected versus observed VAFs for *eight* mutations in sample ARC604 144
Figure 4.4 TSCA results: Number of mutations and functional consequence 148
Figure 4.5 Schematic of *SF3B1* and *ATM* protein domains and mutations .. 150
Figure 4.6 Variant allele frequencies of samples with more than one *ATM* mutation 153
Figure 4.7 Schematic of *NOTCH1* and *TP53* protein domains and mutations 155
Figure 4.8 Schematic of *SAMHD1* and *BIRC3* protein domains and mutations 157
Figure 4.9 Variant allele frequencies of cases with more than one *BIRC3* mutation 158
Figure 4.10 Schematic of *XPO1*, *MYD88* and *MED12* protein domains and mutations 161
Figure 4.11 *TP53* mutations grouped by variant allele frequency 162
Figure 4.12 Summary of CNA results from genome-wide array on 250 samples 165
Figure 4.13 Frequency of all CNAs 166
Figure 4.14 Genome-wide array of CLL sample depicting chromothripisis of chromosome 2 167
Figure 4.15 SNP array analysis reveals regions of cnLOH 170
Figure 4.16 Heterogenous 13q deletions on 132 cases 172
Figure 4.17 Genome-wide array showing 13q deletions involving *RB1* 172
Figure 4.18 Gemone-wide array reveals 13q heterogeneity 174

Figure 4.19 Deletions and cnLOH of 13q 175
Figure 4.20 Case of two 13q deletions: ADM391 176
Figure 4.21 Separate deletions across the 13q locus illustrated by whole-genome array 177
Figure 4.22 Number of CLL driver lesions 178
Figure 4.23 Plot integrating sequencing and SNP array data 179
Figure 4.24 Heatmap of genetic lesions outlining significant associations ... 181
Figure 4.25 Total number of CLL drivers and 13q deletion type 182
Figure 4.26 Association chart showing 11q locus findings 184
Figure 4.27 Subclonal analysis 186
Figure 5.1 Kaplan-Meier estimate of survival for IGHV mutation status 208
Figure 5.2 Kaplan-Meier estimates of survival for *TP53* mutations and *SAMHD1* mutations 210
Figure 5.3 Kaplan-Meier estimates of PFS for trisomy 12 212
Figure 5.4 Kaplan-Meier estimates of Survival for Biallelic *ATM* lesions 213
Figure 5.5 Kaplan-Meier estimates of Survival and 13q deletions 215
Figure 5.6 Kaplan-Meier estimate of survival based on total number of CNAs and cnLOH per sample 217
Figure 5.7 Kaplan-Meier estimates of PFS for Chromothripsis 218
Figure 5.8 Total number of CLL driver lesions defined by PFS 219
Figure 5.9 Kaplan Meier estimates of PFS based on low and high numbers of CLL driver lesions 221
Figure 5.10 Kaplan-Meier estimate of survival based on presence of clonal or subclonal mutations 222

Figure 5.11 Kaplan-Meier estimates of PFS on various VAF subgroups of *TP53* mutations 224
Figure 5.12 Kaplan-Meier estimates of PFS based on subclonal or clonal mutation 225
Figure 5.13 Kaplan-Meier estimates of Survival based on high risk genetic variables and CLL drivers 227
Figure 5.14 LASSO regularization path for the Cox proportional hazards model 229
Figure 5.15 Stability and Ranking of Genomic Variables from Cox model ... 232
Figure 5.16 Prediction error of the stable Cox proportional hazards model . 233
Figure 5.17 Kaplan-Meier estimates of PFS for *SAMHD1* mutation in CLL4 trial cohort 234
Figure 5.18 Kaplan-Meier estimates of progression free survival in LRF CLL4 clinical trial 236
Figure 5.19 Kaplan-Meier estimates of Survival ranked by genomic variable 237

LIST OF TABLES

Table 1.1 Staging Systems 35

Table 2.1. Primer sequences for SAMHD1 sequencing 71

Table 2.2 Genes included in TruSeq Custom Amplicon Panel 77

Table 3.1 Pathogenicity scores of *SAMHD1* mutations 109

Table 3.2 CLL007 *SAMHD1* mutation details 115

Table 4.1 Patient characteristics of all ARCTIC and ADMIRE patients 138

Table 4.2 Dilution experiment results for the *SF3B1* mutation 143

Table 4.3 Variants detected by NGS and validated by Sanger sequencing. 145

Table 4.4 Details of somatic and germline *ATM* mutations 152

Table 4.5 Features of *SAMHD1* mutations 156

Table 4.6 Details of *BIRC3* mutations 159

Table 4.7 Low level *TP53* mutations with VAF <0.05 163

Table 4.8 Samples with Chromothripsis 168

Table 5.1 Patient characteristics including clinical outcome 202

Table 5.2 Univariate regression analysis on Clinical Response 204

Table 5.3 Univariate Cox regression analysis on Minimal Residual disease positive status 206

Table 5.4 Multivariate Cox Regression Analysis using Minimal Residual Disease on genomic alterations. 207

Table 5.5 Univariate Cox regression analysis on Progression Free Survival of all variables (n=250) 209

Table 5.6 Multivariate Cox Regression Survival Analysis 226

Table 5.7 Patient characteristics of LRF CLL4 trial 235

LIST OF ABBREVIATIONS

°C	degrees Celsius
53BP1	p53-binding protein 1
aCNA	acquired copy number alteration
AGS	Aicardi Goutieres syndrome
AID	activation-induced cytidine deaminase
AIHA	autoimmune haemolytic anaemia
ATM	Ataxia-Telangiectasia Mutated
BAK	BCL-2 antagonist/killer
BAX	BCL-2 associated X
BCL-2	B cell lymphoma-2
BCL-xL	B cell leukaemia -xL
BCR	B cell receptor
BH3	BCL-2 homology 3
BMT	Bone marrow transplant
BTK	Bruton's tyrosine Kinase
C-terminal	carboxyl-terminal
cDNA	complementary DNA
CIT	chemo-immunotherapy
CLL	chronic lymphocytic leukaemia
CN	copy number

CNA	copy number aberration
cnLOH	copy neutral loss of heterozygosity
CNV	copy number variation
CPT	camptothecin
CT	computerised tomography
CXCR4	C-X-C chemokine receptor 4
ddNTPs	dideoxynucleotides
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSB	double strand breaks
dsDNA	double stranded DNA
EBF1	early B-cell factor1
EDTA	ethylenediamenetriacetic acid
FACS	fluorescence-activated cell sorting
FAT-C	focal adhesion kinase-C domain
FCM-miniR	fludarabine cyclophosphamide mitoxantrone mini-rituximab
FCM-R	fludarabine cyclophosphamide mitoxantrone rituximab
FCR	fludarabine cyclophosphamide rituximab
FISH	fluorescent in-situ hybridisation
HD	histidine-aspartate
HEAT	huntingtin, elongation factor 3, protein phosphatase 2A , yeast kinase TOR1

HMW	high-molecular-weight
HSC	haematopoietic stem cells
IGH	immunoglobulin heavy chain
IGHV	immunoglobulin heavy chain variable
IGHV-M	mutated immunoglobulin heavy chain variable gene
IGHV-UM	unmutated immunoglobulin heavy chain variable gene
IGV	Integrated Genomic Viewer
IL7Rα	interleukin 7 receptor alpha
Indels	insertions or deletions
ITP	immune mediated thrombocytopenia purpura
KHCO3	potassium carbonate
KM	Kaplan and Meier
LMW	low-molecular-weight
LRF	Leukaemia Research Foundation
M	Molar
Mb	megabases
MBL	monoclonal B cell lymphocytosis
MCL-1	myeloid cell leukaemia sequence-1
MDR	minimally deleted region
$MgCl_2$	magnesium chloride
min	minute
miR	microRNA

MPS	massively parallel sequencing
MRD	minimal residual disease
mRNA	messenger RNA
NaCl	Sodium Chloride
NCRI	National Cancer Research Institute
NCRN	National Cancer Research Network
NE	nuclear extract
NGS	next generation sequencing
NH_4Cl	ammonium Chloride
nM	nano Molar
ORF	open reading frame
PAX5	paired box gene factor 5
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
PI3K	phosphoinositide 3-kinase
PMSF	phenylmethylsulfonyl fluoride
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT-qPCR	real time quantitative polymerase chain reaction
s	second
SAMHD1	SAM domain and HD domain-containing protein 1

SBS	Sequencing by Synthesis
SDS	Sodium Dodecyl Sulphate
SLL	small lymphocytic leukaemia
SNP	single nucleotide polymorphism
TDS	targeted deep sequencing
Thr592	threonine 592
TPH	triphosphohydrolase
TRRAP	Transformation/transcription domain-associated protein
UTR	untranslated region
VAF	variant allele frequency
VH	heavy chain variable gene
w/v	weight per volumne
WES	whole exome sequencing
WGS	whole genome sequencing
wt	wild type
µM	microMolar

CHAPTER 1 : INTRODUCTION

1.1 Chronic Lymphocytic Leukaemia

CLL is a malignant proliferation of mature B cells, which presents clinically with a circulating peripheral blood lymphocytosis. It is the most common adult leukaemia in the Western world and affects males more frequently than females (1.7:1) (Watson et al. 2008). CLL affects 4.5 per 100,000 cases and the incidence is rising over the last four decades, most likely due to improved diagnostics (Howlader 2016). While CLL is known as a disease of the elderly with a median age of 72 at presentation, a significant population of younger patients are also affected with 270 new cases diagnosed in the under 55 year old group in the UK during 2014. With an aging population the burden of CLL on healthcare systems will likely increase over the coming decades. The fact that Caucasians have a higher incidence of CLL than other ethnic groups and the known increased incidence of CLL in family members, suggests a genetic basis to this disease (Morton et al. 2006; Neuland et al. 1983; Yuille et al. 2000).

CLL is a heterogenous disease with a wide variability in disease presentation and clinical course. While some patients with CLL will never require treatment and run a benign course, many others require multiple lines of treatment and can die from the disease. As CLL is generally an incurable disease, periods of remission after therapy alternating with disease relapse have facilitated the study of clonal evolution and effects of therapeutic interventions over time. As

CLL cells circulate in the periphery there is the added ease of access to study material, with the caveat that proliferating events may be more frequent in the lymph node sites (Caligaris-Cappio & Ghia 2008; Herishanu et al. 2011).

1.1.1 Diagnosis

Current guidelines recommend making the diagnosis of CLL based on the presence of morphologically abnormal circulating lymphocytes at a level of $>5x10^9$/l in the peripheral blood with the characteristic immunophenotype of CD19 and CD5 positivity (Michael Hallek et al. 2008; Oscier et al. 2012). A CLL score was devised to aid diagnosis, as not every case will stain typically, allocating one point to each of the following: expression of weak surface membrane immunoglobulins, CD5, CD23, and absent or low expression of CD79b and FMC7 (Matutes et al. 1994). Using this system, 92% of CLL cases score 4 or 5. Further investigations including karyotype or histology can be required in low scoring cases.

1.1.2 Spectrum of disease with CD19/5/23 positive immunophenotype

Two further conditions carry the same immunophenotype as CLL; monoclonal B cell lymphocytosis (MBL) and small lymphocytic leukaemia (SLL). These are distinguished based on circulating B lymphoctyes, below the level of $5x10^9$/L without lymphadenopathy for MBL, and B cell count $<5 \times 10^9$/L with lymphadenopathy for SLL (Marti et al. 2005; Rawstron et al. 2007). MBL is considered a pre-leukaemic condition and harbours many of the cytogenetic lesions found in CLL (Rawstron et al. 2008; Shanafelt et al. 2009). MBL can be further classified depending on the level of clonal B cells above (high-count

MBL) or below (low-count MBL) 0.5 $x10^9$/L (Rawstron et al. 2010). Different clinical and genetic features are seen between both groups. Low-count MBL can be found in up to 75% of people over 90 years of age (Nieto et al. 2009). Low-count MBL is also associated with low risk genetics such as isolated 13q deletion or mutated immunoglobulin heavy chain variable (IGHV-M*)* status and very rarely progresses to high-count MBL and CLL (Henriques et al. 2013; Fazi et al. 2011). The risk of progression from high count MBL to CLL is 1 to 2% per year and more patients are found to have an unmutated IGHV (IGHV-UM) locus and stereotyped CDR3 comparable to CLL cases (Rawstron et al. 2008; Molica et al. 2016; Strati & Shanafelt 2015). SLL is considered the tissue equivalent of CLL. Both SLL and CLL are largely managed in the same way and are classified together for many studies. The underlying reasons for tissue propensity rather than peripheral blood are not completely understood but may be related to differential expression of chemokines and integrins between CLL and SLL and a higher incidence of certain genetic lesions such as trisomy 12 found in SLL (Santos & O'Brien 2012).

1.1.3 Clinical Presentation

Patients may present with B symptoms (weight loss, night sweats, >10% weight loss in <6 months), fatigue, recurrent infections, enlarging lymphadenopathy or symptoms of bone marrow suppression. However the majority of patients are detected as an incidental finding on routine blood counts. As the diagnosis can be made on peripheral blood immunophenotyping, further tests such as bone marrow investigation and computerised tomography (CT) scans are typically

only indicated if treatment is being considered to complete staging and to aid future response assessments (Oscier et al. 2012).

1.1.4 Autoimmune phenomena

Autoimmune phenomena are another feature of CLL. Autoimmune haemolytic anaemia (AIHA) and less frequently, immune thrombocytopenia purpura (ITP) are common presentations often in early stage disease. When both conditions occur together it is known as Evan's syndrome. Treatment is generally of the autoimmune condition, rather than of CLL for Stage A disease. However, for resistant cases treatment of the underlying CLL may be required to control the autoimmune complications (Michael Hallek et al. 2008).

1.1.5 Transformation to high grade lymphoma – Richter's syndrome

A feature of low-grade lymphomas and CLL is the continuous risk of transformation to high-grade lymphoma. In CLL this condition is known as Richter's syndrome and accounts for 2-10% of all CLL depending on the population studied (Parikh et al. 2013; Maddocks-Christianson et al. 2007; Parikh et al. 2014). The histology is diffuse large B cell lymphoma in 95% of cases and Hodgkin's lymphoma in the remaining cases (Swerdlow SH, Campo E 2008). Risk factors for transformation are lymph node size, CD38 expression, lack of 13q deletion and IGHV subset 8 usage (Rossi et al. 2008). Number and type of prior treatments is also a risk factor and fludarabine is associated with an increased risk of transformation, possibly in combination with an alkylating agent (Thornton et al. 2005; Parikh et al. 2013). Initial reports of novel Bruton's tyrosine kinase (BTK) inhibitors revealed a number of cases of Richter's

syndrome developing within months of commencing treatment possibly due to undiagnosed Richter's syndrome prior to trial enrolment (Byrd et al. 2013). Single-agent treatment with Ibrutinib had a 6% incidence of transformation in a phase II trial, all occurring within 15 months of commencing treatment (Ahn et al. 2017). However, a recent UK multi-centre Phase II trial on Richter's transformation, revealed that almost 50% of patients had no prior CLL treatment and this was supported in a single-centre US study (Eyre et al. 2016, Parikh et al. 2013). Two distinct groups of DLBCL have been identified in these patients, those clonally related (80%) and clonally unrelated (20%) to the underlying CLL, when tested using IGHV mutation analysis (Rossi & Gaidano 2009). The prognosis of patients with clonally related RS is poor, with a median overall survival of 8 months compared to up to 5 years in the clonally unrelated group which compares to *de novo* DLBCL (Parikh et al. 2014). Genetic alterations distinct from those seen in de novo DLBCL are found in Richter's syndrome. Inactivation of cell cycle regulation via *TP53* or *CDKN2A* disruption accounts for up to 50% of cases, with c-myc activation seen more commonly in the *TP53* inactivated group (Fabbri & Dalla-Favera 2016; Chigrinova et al. 2013). Trisomy 12, typically in combination with *NOTCH1* mutations, is seen in a further 20% of cases (Ling et al. 2016). Treatment options have limited success for clonally related disease. Standard of care remains anthracycline- or platinum- based chemotherapy followed by stem cell transplantation if remission is achieved however novel agents including BTK inhibitors, BCL-2 inhibitors and immune modulators are currently under investigation (Parikh et al. 2014). Clonally unrelated Richter's syndrome has similar outcomes to *de novo* DLBCL.

1.2 Pathogenesis of Chronic Lymphocytic Leukaemia

Over the last 20 years scientific advances including gene expression profiling and more recently genomic and epigenomic research, coupled with studies of the B cell receptor (BCR) and IGHV mutation status have expanded our understanding of CLL pathogenesis and informed on therapeutic strategies.

1.2.1 Normal B cell development

B cells are blood cells that express clonally diverse cell surface immunoglobulins and were discovered in the 1960's while studying patients with immune deficiency syndromes (Good & Zak 1956). T lymphocytes are involved in the innate immune response while B lymphoctyes are broadly involved in antibody production – a key component of the adaptive immune system. This important aspect of immunity protects humans from rapidly evolving and genetically diverse infectious organisms (Hoehn et al. 2016). B cell development begins in the primary lymphoid organs of the fetal liver and adult bone marrow. Cells circulate to the secondary lymphoid organs of the lymph node and spleen for maturation. The functional endpoint is to become a fully differentiated plasma cell capable of antibody production (LeBien & Tedder 2008).

1.2.2 Haematopoiesis and the bone marrow niche

In the bone marrow, self-renewing haematopoietic stem cells are responsible for the development of all downstream mature functioning blood cells (Zhu & Emerson 2002). The hierarchy of haematopoiesis is subject to continuous

research and debate. The identification of cells at different stages of development can be determined using monoclonal antibodies against cell surface markers or based on the metabolic properties of the cell. Isolation of individual compartments is possible to a high level of purity with contemporary fluorescence-activated cell sorting (FACS). The use of human stem cell transplants as curative treatments to reconstitute haematopoiesis in patients with cancers and constitutional genetic diseases confirms the powerful properties of haematopoietic stem cells (HSC). As haematopoiesis is highly conserved through vertebrate species, studying mouse and zebrafish models has greatly added to our knowledge of haematopoiesis. Reconstitution of human blood cells in immunodeficient mouse models is the current gold-standard experimental model to determine the functional properties of groups of haematopoietic cells (Orkin & Zon 2008).

B lymphopoiesis is one of the leading models for studies of haematopoiesis given the ease for in vivo culture studies on these cells. The multipotent progenitor (MPP) is the earliest known progeny of the HSC. The hierarchy suggested by Reya *et al*. (2001) with a clear separation into myeloid and lymphoid committed cells from the HSC, has had further iterations and an intermediate step with a lymphoid-primed multipotent progenitor (LMPP) is suggested by Adolffson *et al*. (2005) (Reya et al. 2001; Adolfsson et al. 2005). The LMPP contains early lymphoid progenitors (ELPs) that give rise to common lymphoid progenitors (CLPs), which in turn develop primarily into B lymphocytes and natural killer (NK) cells. Moving through these stages of development from immature self-renewing HSC precursors with self-renewal

potential to mature, differentiated, functional blood cells requires the interplay of many transcription factors and growth hormone receptors within the bone marrow niche. The bone marrow niche is an *in vivo* microenvironment that provides signals, nourishment and scaffolding to influence stem cell behaviour (Ferraro et al. 2010). The aging process affects not only the stem cell but also the microenvironment leading to inefficient regulatory processes and exhaustion. The stem cell niche can be perturbed in cancer with abnormal vascular networks and supportive cells in the microenvironment such as fibroblasts maintaining the malignant stem cell niche. Immune cells play a dual role in cancer both in immune surveillance removing abnormal cells but also in the setting of chronic inflammation and skewing local molecular signals can lead to tumour growth (Mantovani et al. 2008). In the setting of CLL, signals from mesenchymal stromal cells, monocytes and T cells in the microenvironment, which refers to both the bone marrow and lymphoid organs, may play a role in CLL cell survival, disease progression and possibly resistant disease (Burger 2011). Following stimulation of the B cell receptor, chemokines and adhesion molecules are released and facilitate cross-talk between CLL cells and these supportive cells. Over the last decade this has become an area of intense investigation and some of these molecules have become therapeutic targets such as the BCR and C-X-C chemokine receptor 4 (CXCR4).

1.2.3 B cell lineage commitment

Relative levels of key transcription factors are at the heart of lineage fate decisions (Zhu & Emerson 2002). Transcription factors function to repress signals towards development of an alternative cell fate while promoting lineage specification through induction of lineage-specific gene expression programs.

Pre-pro B cells are the first stage of B cell precursors and are marked by expression of the B cell associated marker B220. For B lymphoid cells, the lineage-commitment factor paired box gene factor 5 (PAX5) takes centre stage and lack of PAX5 leads to a block in differentiation at the pro-B cell stage (Nutt & Kee 2007). Confirmation of B lineage commitment is seen with the expression of CD19, a target of PAX5. In addition, normal B cell development requires expression of other transcription factors such as Purine-rich DNA sequence (PU.1), Ikaros and early B-cell factor 1 (EBF1) at different stages and knock down of these factors leads to severely suppressed or absent B cells in mouse models. In addition, fms-like tyrosine kinase 3 (Flt-3) ligand has a pertinent role at the LMPP developmental stage, and interleukin 7 receptor alpha (IL7Rα) at the CLP stage (Nutt & Kee 2007). Finally, the concept of lineage specification and lineage commitment as events that were synonymous seem to be separate events with EBF1 and E2A1 acting to prime precursor cells and PAX5 as the main factor to commit these progenitor cells to the B lineage (Traver & Akashi 2004). Continuous PAX5 expression until plasma cell differentiation is required with the potential to revert to a multipotent progenitor cell, even at the differentiated post lymph node B cell stage in the absence of PAX5 (Delogu et al. 2006).

1.2.4 The developing B cell

The developing B cell undergoes numerous steps to increase the diversity of antigens that will ultimately bind to the mature BCR. The structure of an antibody is made up of four polypeptide chains, two heavy chains and two light chains. The heavy chain genes (IGH) are located on chromosome 14 while the

light chain genes, either kappa or lambda (IGκ, IGλ), are found on chromosomes 2 and 22 respectively. A given immunoglobulin has either a κ or a λ chain, with no known functional difference between them. An excess of one chain over the other can be a sign of disease. The constant (C) region is composed of the carboxy-terminal domains of the heavy chains and is responsible for the class of the immunoglobulin. The function of the C region is to interact with effector cells and molecules. Five classes of immunoglobulin exist: IgG, IgM, IgD, IgA and IgE (Herrmann et al. 2016).

The variable (V) region at the amino terminus of the four polypeptide chains is made up of three highly diverse regions known as the complementarity determining regions (CDR) (Schroeder et al. 2010). Antibody receptor diversity is produced initially by genetic recombination of segments of the heavy and light chain components of the B cell receptor at the pre-B cell stage. The variable (V), diversity (D), and joining (J) gene regions of the heavy chain, and the V and J gene segments of the light chain, are subject to V(D)J somatic recombination. The majority of the recombination occurs in the variable gene that contains 44 open reading frames out of the 123 - 129 gene segments. This compares to 25 of 27 D segments and 5 of 7 J segments (Lefranc & Lefranc 2001). During this process, further diversification can be introduced by deletions and insertions at the junctions of each segment. The result is fusing of highly variable CDRs to more conserved, structurally supportive regions known as the constant regions.

Each naïve B cell has it's own BCR and the number of possible sequences is predicted to be at least 10^{18} (Elhanati et al. 2015). Class switch recombination (CSR) and somatic hypermutation (SHM) occur at a later stage of B cell development and take place in the germinal centre. CSR affects interactions with effector molecules by altering the constant region of the immunoglobulin rather than the variable region and therefore antigenic specificity is retained. The SHM process is known as affinity maturation and occurs after exposure to antigen. This process of rapid mutation of the order of one mutation per cell division per region is tightly controlled by the enzyme activation-induced cytidine deaminase (AID) (Teng & Papavasiliou 2007). A round of competitive B cell selection follows this mutation burst resulting in high-affinity antigen binding (Peled et al. 2008). Differences in the degree of SHM have been found in CLL and have shaped our understanding of the origins of CLL in addition to acting as prognostic markers based on markedly different outcomes between the two subgroups. This will be discussed further in the following sections and also in Subsection 1.4.2 in relation to prognostic markers.

1.2.5 B cell receptor signalling in CLL

The BCR is a membrane-bound immunoglobulin that binds antigen. Levels of BCR expression in CLL are low compared to other B cell lymphomas (Zomas et al. 1996). Consequently, with the finding that both mutated and unmutated immunoglobulin genes are present in CLL patients, the study of the BCR in CLL became of major importance (Damle et al. 1999; Hamblin et al. 1999). The process of chronic low-level stimulation of the BCR is known as anergy (Cambier et al. 2007). CLL cells have variable levels of BCR stimulation from

anergic to more responsive (Packham & Stevenson 2010). While anergic CLL cells are relatively quiescent they have the potential for re-entry into the cell cycle and a return to proliferation. This is in keeping with more recent experiments using heavy water that revealed approximately 0.1-1% of circulating CLL cells are proliferating each day with the remainder held in a state of anergy (Messmer et al. 2005). Depending on the mutation status of the IGHV region, more cells can be in this proliferative phase (Ian Mockridge et al. 2007; Guarini et al. 2008). Early reports from BCR inhibitor use in CLL revealed a better response in the IGHV-UM group, possibly a reflection of the higher state of BCR stimulation (Byrd et al. 2013). The state of BCR stimulation is relatively stable for IGHV-UM patients, in that there is a continuous level of stimulation, whereas for IGHV-M patients there is a spectrum from anergy to stimulation of the BCR, reflecting a further level of heterogeneity (Ian Mockridge et al. 2007). It is known that IGHV-UM cells have higher levels of genomic alterations and whether this holds true for the more stimulated BCR subgroup of IGHV-M patients is not yet clear (Rossi et al. 2013).

1.2.6 Apoptosis in CLL

Circulating CLL cells are predominately in the G0 phase of cell cycle (Andreeff et al. 1980). This finding is in keeping with the theory that CLL involves disruption of normal B cell apoptosis. Apoptosis is a form of programmed cell death and is the normal cellular response to stress such as from reactive oxygen species, chemotherapy or radiation. Regulators of apoptosis include proteases of the B cell lymphoma-2 (BCL-2) family protein including both pro- and anti-apoptotic molecules. In normal homeostasis a fine balance between

the pro-apoptotic and anti-apoptotic proteins is required. In the setting of cell damage, the normal response is activation of pro-apoptotic BCL-2 associated X (BAX) or BCL-2 antagonist/killer (BAK) leading to mitochondrial damage, release of cytochrome c and subsequently cell death. Overexpression of BCL-2, one of five key anti-apoptotic molecules, has been found in CLL, which prevents the normal cellular response to stress (Robertson et al. 1996). Other means of pathological anti-apoptotic signalling can occur via Myeloid cell leukaemia sequence-1 (MCL-1) overexpression, disruption of tumour suppressor genes, and pro-apoptotic under-expression (Ortíz-Maldonado et al. 2016).

1.2.7 Cell of origin

The malignant cell counterpart to a normal B cell can retain the immunophenotype and gene expression profile of this cell. In lymphoid neoplasms it has been possible to identify the normal cell counterpart where arrested development of the B cell occurs. It is currently unknown at which stage of B cell development leukaemia-initiating events occur in CLL. It was initially thought that the normal counterpart for CLL was a naïve B cell from the follicular mantle zone that had not yet reached the germinal centre, due to the presence of CD5 expression (Nicholson et al. 1995; Caligaris-Cappio 1996). Early work on the somatic mutation status of CLL cells revealed that some samples were IGHV-M indicating a stage in B cell development that required germinal centre experience (Schroeder & Dighiero 1994). This theory was more in keeping with an activated mature B cell as the normal counterpart. Subsequent studies on the gene expression profiles of CLL cells highlighted the memory B cell as the closest counterpart, with a global pattern for all CLL cases irrespective of IGHV mutation status reported (Klein et al. 2001). However, as further studies unraveled the two distinct populations defined by either IGHV-M or IGHV-UM genes outlined above, it was believed that two separate diseases may exist, or at least two different cells of origin for the initiating leukaemogenic event (Hamblin et al. 1999; Stevenson et al. 2001). The poor outcome of patients with IGHV-UM status possibly reflects a diverse origin for the initiating leukaemia cell(s) in both patient groups. Furthermore, the finding of haematopoietic stem cells (HSC) isolated from the bone marrow of patients with CLL which gave rise to clonal lymphopoiesis in immunodeficient

mice indicates that the primary defect in CLL may be within a bone marrow derived progenitor cell (Kikushige et al, XX, Damm et al, 2014).

1.3 Genomics of Chronic Lymphocytic Leukaemia

1.3.1 Overview of genomic profile of CLL

For over 35 years the genetic profile of CLL has been investigated. Initially a controversial point, it has become clear since the turn of the century that genetic changes influence outcome in CLL (Juliusson et al. 1990; Döhner et al. 2000). Genomic imbalances are frequent in CLL however reciprocal translocations were rarely found in conventional metaphase cytogenetics (Haferlach et al. 2007). The impact of genetic changes on patient survival is dynamic and alters as treatments change (Hallek et al. 2010). An individual patient's genetic changes are also dynamic with particular alterations present from disease initiation at a clonal level while other alterations may become more dominant as the disease progresses (Schuh et al. 2012).

1.3.2 Structural variation

Initial B-cell metaphase cytogenetic studies recognised structural abnormalities in 218 of 391 (56%) CLL samples and linked cytogenetic abnormalities with reduced survival using multivariate analysis (Juliusson et al. 1990). This key study first introduced the concepts of both the clonal burden of genetic abnormalities and also of a complex karyotype associating with a poor outcome. In addition, this study highlighted trisomy 12 and deletions of chromosome bands 13q14 and 14q32 as significant alterations in CLL.

The cumbersome application of metaphase cytogenetics in CLL led to the investigation of interphase fluorescent *in situ* hybridisation (FISH) to improve the detection of genetic aberrations. A study by the German group reported both 13q deletions and 11q deletions as the most common alterations in CLL. trisomy 12 was the next most frequently involved region, and also recurrent deletions of 17p containing the *TP53* gene. Alterations of chromosomal regions 6q, 3p and 14q were also noted but at a low frequency (Döhner et al. 1999; Döhner et al. 1995). Subsequently, a ground-breaking publication linked genomic alterations discovered by FISH with clinical outcome (Döhner et al. 2000). To this day the Döhner hierarchy remains relevant for the prognostic evaluation of CLL patients, even in the era of targeted therapies.

1.3.3 Deletions of chromosome 13q

The most common chromosomal aberration in CLL involves deletion of the chromosomal band 13q14 (del(13q)) in approximately 55% of cases (Döhner et al. 1999). This deletion carries an excellent prognosis when found in isolation (Döhner et al. 2000). The minimally deleted region (MDR) has been refined over many studies and includes the genes *DLEU1*, *DLEU2* and the microRNAs 15a and 16-1 (miR15a/16-1) (Kalachikov et al. 1997; Liu et al. 1995; Calin et al. 2002; Corcoran et al. 1998; Bullrich et al. 1996; Ouillette et al. 2008; Edelmann et al. 2013). Mutations in tumour suppressor genes leading to gene inactivation are a key feature in cancer. While no sequence mutations were found in either of the *DLEU1* and *DLEU2* genes, a more recent study revealed a low frequency of germline mutations in the precursor RNA of MicroRNA 16.1 (Migliazza et al. 2001; Calin et al. 2005). Decreased expression or deletion of

the microRNA cluster was found in 68% of cases (Calin et al. 2002). A study of a murine knock out with deletion of this MDR led to development of CLL at low penetrance (21%) further supporting the notion that this region is important in leukaemogenesis (Klein et al. 2010). Dissecting the contribution of both the miRNA locus and *DLeu2* this group showed a more aggressive phenotype with early death when the entire MDR was knocked out versus solely the miRNA cluster suggesting several interacting genes in the tumour suppressive process. This study suggested a potential role for the miRNA cluster in cell cycle regulation. Repression of BCL-2 by miR-15a/miR-16-1 leading to apoptosis in a leukaemia cell model has been reported (Cimmino et al. 2005). A link between TP53 and this microRNA cluster has also been defined in leukemic cell lines and primary CLL cells whereby TP53 stimulated the transcription of miR-15/miR-16-1(M. Fabbri et al. 2011).

Del(13q) can be monoallelic or biallelic (20%) with acquisition of biallelic loss as the CLL clone evolves (Stilgenbauer et al. 2007; Knight et al. 2012; Shanafelt et al. 2006). Varying lengths of del(13q) have been described and linked with clinical outcome (Ouillette et al. 2008; Gunnarsson et al. 2011; Mian et al. 2013; Parker et al. 2010). A prognostic role for longer deletions (Type II del(13q)) compared to shorter Type I deletions may relate to the deletion of the *RB1* gene which is involved in Type II deletions (Parker et al. 2010; Dal Bo et al. 2011; Ouillette et al. 2008). In addition, the percentage of cells with del(13q) can impact on outcome, this relates to increasing numbers of cells with a monoallelic del(13q) rather than the presence of a biallelic loss (Hernández et al. 2009; Van Dyke et al. 2010).

1.3.4 Deletions of chromosome 11q

While del(13q) in isolation related to a better prognosis, the finding of deletion of a region of chromosome 11q (del(11q)), the second most common chromosomal deletion found in approximately 15-20% of cases, has historically been linked with more aggressive disease, younger age at onset, rapid disease progression and bulky lymphadenopathy (Neilson et al. 1997; Döhner et al. 1997). In Döhner's hierarchical model del(11q) retained independent prognostic significance for shorter survival in a multivariate analysis including other cytogenetic lesions and biologcial characteristics. The MDR relating to del(11q) is located at 11q22.3 and contains many genes including the *ATM* gene. This gene plays a central role in the DNA damage repair response and mutations of the *ATM* gene, both germline and somatic, are found in up 30% of CLL patients leading to biallelic inactivation of *ATM* in some patients (Austen et al. 2005; Austen et al. 2007). No reports of bilalleleic deletion of the 11q locus have been described in CLL. Compared to the initial FISH-based cytogenetic model from 2000, patients with del(11q) enrolled in clinical trials showed an improved outcome with chemo- and chemoimmunotherapy (CIT) regimens especially with the addition of Rituximab (Stilgenbauer 2008; Hallek et al. 2010; Tsimberidou et al. 2009).

1.3.5 Trisomy 12

Trisomy 12 defines a heterogenous group and while initially found to be the most frequent aberration when using conventional cytogenetics, the application of interphase FISH reveals that it is the third most common lesion affecting 10-20% of CLL patients and can be the sole lesion identified in the majority of these

cases (Juliusson & Gahrton 1990; Matutes et al. 1996; Döhner et al. 2000; Lehmann et al. 2008). Conflicting reports relating to clinical outcome have been reported with an intermediate or good risk profile associated with trisomy 12 both in population based and frontline treatment settings (Döhner et al. 2000; Hallek et al. 2010; Gunnarsson et al. 2011; Thorsten Zenz, Mertens, et al. 2010). Earlier studies reported that trisomy 12 was associated with more aggressive behaviour (Oscier et al. 1990; Matutes et al. 1996; Athanasiadou et al. 2006). More recently the association with *NOTCH1* mutations reveals further subgroups for patients with trisomy 12 that may explain the heterogeneous outcomes seen (Del Giudice et al. 2012; Balatti et al. 2012). In addition, while infrequent, deletions of 14q32 can be seen with cases of trisomy 12 and have a poorer prognosis which was independent of IGHV mutation status (Reindl et al. 2010). In contrast, cases with trisomy 19 and/or trisomy 18 as additional cytogenetic gain(s) have been reported in patients with trisomy 12 with a IGHV-M status and an improved prognosis (Sellmann et al. 2007; Ibbotson et al. 2012).

As the entire chromosome is duplicated the search for key oncogenes is challenging when a minimally affected region is not available to interrogate. Numerous putative candidate genes have been selected for further investigation from gene expression profiling reports of overexpression of genes on chromosome 12 such as *HIPIR*, *PYF6*, *CDK4* and *MDM,* a regulator of *TP53* (Porpaczy et al. 2009; Kienle et al. 2005; Koníková & Kusenda 2003; Haupt et al. 1997).

1.3.6 Deletions of chromosome 17p

Deletion of chromosome 17p13 is seen at a low frequency at diagnosis and prior to frontline treatment in the order of 3-8% (Döhner et al. 1995; Döhner et al. 2000; Delgado et al. 2012). However, this increases to up to 30% at relapse corresponding with the failure of standard chemotherapy including CIT that is typical in del(17p) patients (Stilgenbauer et al. 2009; Hallek et al. 2010). The *TP53* gene is located at chromosome 17p13 and this gene is within the MDR defined by CLL patients. The TP53 protein responds to diverse cellular stresses and induces cell cycle arrest, DNA repair and apoptosis (Lane 1992; Levine & Oren 2009). *TP53* mutations are reported in a number of cancers including hereditary cancer predisposition syndromes such as Li-Fraumeni (Srivastava et al. 1990; Malkin et al. 1990). The mode of action of conventional chemotherapy such as fludarabine, a nucleoside analogue, is via *TP53*-driven apoptosis (Pettitt 2003). In the setting of del(17p) the *TP53* pathway is affected and resistance to chemotherapy is seen (Stilgenbauer 2008; Hallek et al. 2010). The negative impact of del(17p) is independent of other predictive markers and typically affects patients with IGHV-UM status, complex karyotype and enrichment of this deletion in the relapsed refractory setting (Stilgenbauer et al. 2007; Stilgenbauer et al. 2009; Edelmann et al. 2013). For a subgroup of del(17p) patients a more indolent course was seen for patients with de novo del(17p) compared to the acquisition of this lesion as a second or subsequent genetic hit (Tam et al. 2009). The percentage of 17p deleted nuclei can impact on outcome as reported initially at a level of >3% and subsequently at levels of >10% and >25% by FISH (Döhner et al. 1995; Oscier et al. 2010; Catovsky et al. 2007).

1.3.7 Structural variants of lower frequency in CLL

Recurrent copy number alterations of lower frequency have also been discovered, particularly with the increasing use of genome-wide arrays (Kujawski et al. 2008; Gunnarsson et al. 2011; Lehmann et al. 2008; Pfeifer et al. 2007; J. R. Brown et al. 2012; Edelmann et al. 2013; Knight et al. 2012). These lesions vary from 1-5% at diagnosis and may increase in frequency with progressive CLL.

Deletions of chromosome 6q have been noted at varying frequencies of <1% to 7% of CLL cases (Lehmann et al. 2008; Gunnarsson et al. 2011; Ouillette et al. 2011; Knight et al. 2012; Edelmann et al. 2013). It has been difficult to determine an MDR, even with the advent of genome-wide array technology, as the deletions are very heterogenous. For a subset of cases a region within 6q21 of 2.5Mb was commonly deleted in 80% of cases and this region was also refined as the MDR for three cases involving 6q21 in a previous study (Edelmann et al. 2013; Lehmann et al. 2008). Lesions in this region have not been associated with good or poor risk disease, this may be due in part to the low frequency of lesions and also the lack of clarity on whether there are one or more key regions affected (D.-M. Wang et al. 2011; Döhner et al. 1999; Lehmann et al. 2008).

Copy number gains of chromosome 2p are also described at a frequency of 1.9-28% (Gunnarsson et al. 2010; Chapiro et al. 2010; Rinaldi et al. 2011; Edelmann et al. 2013). The *MYCN and REL* genes are within the commonly

affected region and increased mRNA expression of *MYCN* has been reported (Chapiro et al. 2010). More recently, in the frontline treatment cohort from the German CLL8 clinical trial of CIT, a frequency of 7% for 2p gains was noted, and the MDR contained nine genes including *XPO1*, a novel gene recurrently mutated in CLL (Edelmann et al. 2013; Puente et al. 2011). A link with IGHV-UM genes and also with 17p deletions has been reported (Chapiro et al. 2010; Forconi et al. 2008).

Alterations across chromosome 8 including copy number gain of 8q and copy number loss of 8p (del(8p)) have been described at low frequency of between 1.6 and 7% (Grubor et al. 2009; Gunnarsson et al. 2011; Rinaldi et al. 2011; Edelmann et al. 2013; Knight et al. 2012). The *c-myc* oncogene which has proven leukaemogenic potential across a range of B cell malignancies(Adams et al. 1985), is within the MDR of some cases with 8p gains (Jennifer R. Brown et al. 2012) . A 7% frequency for 8p gains in a subsequent study revealed that the MDR was 500 kilobases centromeric of *c-myc* (Edelmann et al. 2013). Gains of 2p and 8q, and del(8p) when studied against clinical outcome predicted a poor prognosis, and were independent predictors of decreased overall survival in multivariate analyses (Forconi et al. 2008; Rinaldi et al. 2011; Chapiro et al. 2010).

Chromosomal translocations are noted as an infrequent occurrence in CLL (Haferlach et al. 2007). The immunoglobulin heavy chain (IGH) locus has long been recognised as a region involved in balanced and unbalanced translocations across a range of B cell neoplasms including CLL. In addition,

copy number deletions of regions centromeric of the IGH locus on chromosome 14 have been described but at low frequency of 1-2% (Reindl et al. 2010; Haferlach et al. 2007; López et al. 2012). A prognostic impact related to these lesions has been noted and may aid in further classifying larger more heterogenous groups such as trisomy 12 affected cases when both lesions occur together (Reindl et al. 2010).

While copy number losses are the predominant feature when analysing sub-chromosomal lesions in CLL, trisomy 12 is a recurrent finding as outlined above, and in a low number of cases it associates with additional trisomies, most frequently trisomy 19 but also trisomy 18 in 50% of cases in one study (Edelmann et al. 2013). The significance of these findings may help further subgroup the heterogeneous group that is currently defined by trisomy 12. The additional trisomies tend to occur in lower risk groups with better prognosis but the frequency of these findings is approximately 2-3% in the frontline treatment setting therefore it is difficult to confirm their independent prognostic value outside of very large clinical studies.

1.3.8 Copy neutral loss of heterozygosity

Chromosomal changes can be seen without any change in the total copy number of chromosomes as in the case of unusually large regions of homozygosity. This can affect germline cells when both maternal or both paternal alleles are passed to their offspring due to an error occurring during meiosis (uniparental isodisomy) or in cases where the parents are related (identity by descent). In cancer, errors can lead to acquired regions of homozygosity or 'copy neutral loss of homozygosity' (cnLOH), resulting in two copies of a region of a chromosome harbouring an inactivated tumour suppressor gene for example (Mao et al. 2007).

CnLOH can be detected using single nucleotide polymorphism (SNP) array technology by analysing the B allele frequency plot in relation to the $\log_2$ ratio plot. Large areas of cnLOH are a feature of cancer and haematological malignancies (Mao et al. 2007; Gondek et al. 2007). In CLL these regions have previously been reported to be infrequent but significant by finding cnLOH in regions that have been associated previously with copy number loss in CLL such as del(13q) and del(17p) (Knight et al. 2012; Edelmann et al. 2013; Gunnarsson et al. 2011; Ouillette et al. 2008)

1.3.9 Genomic Complexity

Complex chromosomal abnormalities of ≥3 genomic alterations have been found in 20% of patients, referred to as a complex karyotype, and linked with a more aggressive disease course using conventional metaphase cytogenetics (Juliusson & Merup 1998; Haferlach et al. 2007). Building on technological

advances with the more widespread use of high-density SNP arrays several groups have demonstrated that elevated numbers of copy number alterations (CNAs) are associated with progressive disease and shorter survival (Kujawski et al. 2008; Ouillette et al. 2011; Jennifer R. Brown et al. 2012; Knight et al. 2012). While the majority of patients carried 0-2 CNAs, a subset with increased numbers of CNAs were found with ~20% of cases affected by ≥3 CNAs titled genomic complexity (GC). The impact of the overall size of genomic lesions was also investigated and lesions greater than 1Mb and certainly >5Mb were found in patients with shorter overall survival. These findings however accompanied the already defined high-risk groups of del(17p) and del(11q) (Gunnarsson et al. 2010). GC has been identified in patients affected by del(17p) or del(11q) which are lesions affecting the *TP53* and *ATM* genes respectively and while interference with the DNA damage repair response pathway may explain the genomic instability in these cases, up to two thirds of cases with GC are not affected by these lesions (Malek 2012). The underlying mechanism here is yet to be explained. Interestingly cases without genomic complexity harbouring either del(17p), *TP53* mutations or del(11q) did not progress to require treatment compared with GC without additional high-risk features (Ouillette et al. 2011).

Chromothripsis is the finding of a high number (n≥10) of alternating copy number changes in an individual chromosome and has been described as a one-step cellular catastrophic event in CLL (Rausch et al. 2012; Stephens et al. 2011). The finding of alternating copy number alterations of greater than seven or more, referred to as chromothripsis-like events, were associated with

a shorter PFS after frontline FCR treatment in a clinical trial setting (Edelmann et al. 2013).

1.3.10 Key driver genes and pathways in CLL

The concept of recurrent copy number alterations leading to the definition of MDRs in CLL has facilitated the study of key genes that may be functioning as tumour suppressors in CLL as outlined above in the cases of del(13q), del(11q) and del(17p). Furthermore, cnLOH events in key CLL-related regions can highlight regions with mutated tumour suppressor genes or inactivated oncogenes. However, other regions in the genome may play a role in CLL pathogenesis and disease progression as the clinical heterogeneity of CLL and lack of response to treatment has not completely been explained by the events described to date. The advent of next-generation sequencing has enabled the unbiased exploration of genome-wide structural and sequence mutations leading to the discovery of novel gene mutations and signalling pathways not previously recognised in the pathobiology of CLL (L. Wang et al. 2011a; Quesada et al. 2011; G. Fabbri et al. 2011; Puente et al. 2011). Affected genes in CLL can be grouped into a number of cellular pathways.

1.3.11 DNA damage response pathway: TP53 and ATM proteins

Constant assaults on the genetic material of human cells led to the evolution of complex systems of DNA damage detection and repair to ensure cellular genomic integrity is maintained (Jackson & Bartek 2009). Intact cellular responses to DNA damage are necessary to prevent cancer development. Key proteins involved in signalling a cellular response to DNA damage are TP53 and ATM (Lakin & Jackson 1999; Shiloh 2003).

TP53 mutations affect 4-10% of CLL patients at diagnosis or prior to frontline treatment and accompany del(17p) in over 80% of cases (Thorsten Zenz, Eichhorst, et al. 2010; Rossi et al. 2009; Gonzalez et al. 2011). Monoallelic *TP53* mutations impact on prognosis and were the strongest independent predictor of overall survival in the setting of Fludarabine-based treatment in the German CLL4 trial (Zenz et al. 2011; Zenz et al. 2008; Malcikova et al. 2009). This finding was confirmed in the frontline LRF CLL4 trial (Gonzalez et al. 2011). Acquisition of new *TP53* mutations at the time of refractory disease has been reported and the frequency of *TP53* mutations increases as CLL progresses (T Zenz et al. 2010; Rossi et al. 2009). Testing for a *TP53* mutation is recommended based on a lack of response to standard frontline CIT approaches and the need for alternative treatment strategies upfront for this group of patients (Pospisilova et al. 2012; Hallek et al. 2010). This group of patients with *TP53* disruption; accounting for any combination of *TP53* mutation or del(17p) is referred to as ultra high risk CLL as these patients relapse within 2-3 years of standard CIT approaches (Stilgenbauer & Zenz 2010). Novel

treatments including novel monoclonal antibodies and kinase inhibitors targeting the B cell receptor pathway may have efficacy in this difficult to treat patient cohort (Clifford & Schuh 2012; Stilgenbauer & Zenz 2010).

ATM is a protein kinase and is central to the DNA damage repair response that becomes activated in response to double-stranded DNA breaks in order to minimise the risk of DNA damage (Mavrou et al. 2008). Germline mutations in *ATM* are a feature of Ataxia Telangiectasia, from which the name arises (Stankovic et al. 1998). This hereditary neurodegenerative disease is a rare autosomal recessive disorder characterized by chromosome instability, radiosensitivity, immunodeficiency and a predisposition for cancer including lymphoid malignancies (Mavrou et al. 2008). The study to identify the candidate tumour suppressor gene(s) within the del(11q) MDR has revealed convincing evidence that the *ATM* gene plays a role in CLL pathogenesis and disease progression. In addition to AT patients having a predisposition to lymphoid malignancies, this gene is recurrently mutated in CLL patients in 10-16% of all patients and in 36-50% of patients with del(11q) (Austen et al. 2005; Stankovic et al. 1999; Austen et al. 2007; Skowronska et al. 2012). However, several other possible candidate genes lie within the large del(11q) MDR. Studies on patients with biallelic *ATM* inactivation have revealed a poor prognosis comparable to that seen with *TP53* deletion and/or mutation (*TP53* disruption) (Skowronska et al. 2012; Rose-Zerilli et al. 2014). *ATM* mutations can appear over the course of a patient's disease and can lead to chemotherapy resistance (Austen et al. 2007). More recently, a beneficial response for del(11q) patients with the addition of Rituximab to combination chemotherapy has been seen, however

details on the *ATM* mutation status of these patients was not known (Hallek et al. 2010).

1.3.12 NOTCH signalling pathway and NOTCH1 mutations

The NOTCH signalling pathway is involved in cell fate decisions in both embryonic and adult life. NOTCH activity affects the implementation of differentiation, proliferation, and apoptotic programs (Artavanis-Tsakonas et al. 1999). *NOTCH1* encodes for a NOTCH1 trans-membrane protein that acts as a ligand-activated transcription factor regulating several downstream pathways important for the control of cell growth (Landau and Wu, 2013). Mutations in *NOTCH1* in haematological malignancies were first reported in T acute lymphoblastic leukaemia with activating mutations involving the C-terminal PEST domain of NOTCH1 (Weng et al. 2004). A landmark whole genome sequencing study by Fabbri *et al*. (2011a) reported *NOTCH1* mutations in two cases out of five leading to the validation of this mutation in a larger cohort (G. Fabbri et al. 2011). The predominant mutation is a 2-bp frameshift deletion (c.7544_7545delCT) in exon 34 which has been validated in subsequent CLL cohorts and is predicted to disrupt the PEST domain (Puente et al. 2011). A negative prognostic impact for *NOTCH1* mutated patients, an association with trisomy 12, and an increased frequency of Richter's syndrome has been linked with *NOTCH1* mutations (Puente et al. 2011; D. Rossi et al. 2012; Del Giudice et al. 2012).

1.3.13 RNA processing pathways

A somewhat surprising result from a study using whole exome sequencing (WES) implicated the RNA processing pathway in CLL for the first time.

Recurrent mutations in the splicing factor 3B subunit 1 (*SF3B1)* were reported at a frequency of 15% and were associated with a poor prognosis (L. Wang et al. 2011a). This finding was confirmed by another next generation sequencing study (Quesada et al. 2012). *SF3B1* is also recurrently mutated in myelodysplastic syndrome (MDS) and is found most commonly in the subgroup of MDS with ring sideroblasts and a better overall outcome (Yoshida et al. 2011; Papaemmanuil et al. 2011). The majority of mutations are located in the evolutionary conserved C-terminal HEAT-repeat domain and consist of three hotspot mutations in the majority of cases with fewer mutations spread across exons 14-16. SF3B1 is a key component of the spliceosome and mutations in the gene may lead to alternative splice variants. The exact mechanism whereby *SF3B1* mutations effect leukaemogenesis is unknown, while several reports have not found a significant impact on alternative splicing, this remains the subject of intense investigation facilitated by the availability of transcriptomics (Quesada et al. 2011; Visconte et al. 2012). Subsequent studies reported an increase in *SF3B1* mutations over the CLL disease course and highlighted *SF3B1* as a negative prognostic marker contrary to the positive prognostic association in MDS (Quesada et al. 2011; Wang et al. 2011a; Rossi et al. 2011).

The nuclear export protein Exportin 1 (*XPO1*) also referred to as CRM1 (chromosome region maintenance 1) carries over 200 proteins such as TP53 and IkB (inhibitor of nuclear factor of kappa light polypeptide gene enhancer in B-cells (NF-kappaB)) out of the nucleus, in addition to a subset of mRNAs thus mediating the distribution of many subcellular molecules. It is the best-characterized nuclear exporter in humans (Xu et al. 2010). Increased

expression of this protein has been reported in other cancers and recurrent copy number gains across the 2p region which encompasses the *XPO1* locus are an infrequent finding in CLL (Yao et al. 2009; Huang et al. 2009; Noske et al. 2008; Forconi et al. 2008; Edelmann et al. 2013). A recurrent mutation at the region of p.E571, of high variant allele frequency (VAF), was discovered in an initial CLL WGS study (Puente et al. 2011). The precise effect of this mutation is unknown but the hotspot mutation in a highly conserved codon points towards a pathogenic role.

1.3.14 NF-Kappa B signalling pathway

The NF-kappa B signalling pathway is constitutionally activated in cancer and notably in haematological malignancies involving mature B cells. A ground-breaking treatment for multiple myeloma patients, bortezomib, inhibits this pathway and was introduced to clinical practice 15 years ago. This has resulted in a paradigm shift with increased overall survival for these patients (Kumar et al. 2008). The NF-kappaB pathway is a family of transcription factors involved in inflammation, proliferation, differentiation and cell survival. Regulation is via inhibitors of NF-kappaB (IKBs) and through various post-translational modifications (Oeckinghaus & Ghosh 2009). The NF-kappaB pathway can be activated via toll-like receptors (TLRs), Interleukin-1 (IL-1) and lipopolysaccharide (LPS), all known as the canonical pathway activation. Non-canonical pathway signalling depends on the activation of NF-kappaB-inducing kinase (Oeckinghaus & Ghosh 2009).

Baculoviral inhibition of apoptosis repeat containing 3 (BIRC3) is a negative regulator of the non-canonical NF-kappaB pathway and is involved in

maintaining TP53 levels via NF-kappaB mediated modifications of MDM2 expression. BIRC3 knockdown leads to downregulation of TP53 and cancer development (Lau et al. 2012). *BIRC3* is recurrently mutated in CLL at a level of 4% in diagnostic samples increasing to 24% in fludarabine-refractory cases (Rossi et al. 2012). *BIRC3* is located at 11q22 telomeric to the *ATM* gene and is affected by the classical del(11q) leading to a biallelic *BIRC3* disruption (Malek 2012).

Toll-like receptor (TLR) signalling is a key part of the innate immune system. Part of this system includes adaptor proteins that effect the signals from TLRs upon pathogen sensing at the cell surface. Myeloid differentiation primary response 88 (*MYD88*) is one of these adaptor proteins and leads to increased NF-kappaB activation (Zhang et al. 2012). Mutations in *MYD88* have been reported at a low frequency of approximately 4% in CLL and lead to the L265P missense mutation. *MYD88* is recurrently mutated in Waldenstrom's Macroglobulinaemia where it is of diagnostic use as it is present in 90% of cases and is a potential therapeutic target (Treon et al. 2012). It is also commonly mutated in the activated B cell subtype of diffuse large B cell lymphoma (Ngo et al. 2011). In the setting of CLL, *MYD88* mutations align with a IGHV-M locus and a less aggressive clinical course (Puente et al. 2011).

1.3.15 Intra-clonal heterogeneity defined by CLL drivers

It is clear that a single high-frequency CLL driver lesion does not exist such as that seen in CML with the BCR:ABL translocation. In this context a driver lesion refers to one found recurrently in CLL and predicted to have a pathogenic

consequence. Outlined in section 1.3 are CLL driver lesions including larger copy number alterations defined by FISH and genome-wide array and somatic single nucleotide alterations and indels. In CLL higher frequency mutations of 10-30% exist along with those of lower frequency in the range of 1-10%. CLL driver lesions may occur as a leukaemogenic disease-causing event as has been suggested for trisomy 12 or heterozygous del(13q) based on their presence early in the disease course and the stable frequency over time (Hjalmar et al. 2001; Chevallier et al. 2002). Other CLL driver lesions may cause the disease to transform to a more aggressive proliferative phase as has been suggested for del(11q) and del(17p) as well as *ATM* and *TP53* mutations, based on the theory that these lesions have a higher frequency later in the disease course when treatment is required or in the relapsed/refractory setting (Chevallier et al. 2002; Stilgenbauer et al. 2007). Defining the clonal composition of the CLL bulk can be used to infer the stage of mutation acquisition and from this the role this lesion may play in CLL pathogenesis and disease progression.

1.3.16 Clonal evolution

Early studies of sequential samples in CLL using conventional cytogenetics revealed that the acquisition of new lesions was uncommon (Juliusson et al. 1990) but this was challenged in further studies and clonal evolution in 43% of patients most frequently involving deletions of the long arm of chromosome 6 were reported (Finn et al. 1998). To evaluate clonal evolution using FISH, a study over an eight year time period revealed that clonal evolution in the first two years was seen in only 1.7% of patients in a cohort of 159 patients but this

increased to 27% after five years or more (Shanafelt et al. 2006). The development of clonal evolution was linked with increasing numbers of prior treatments and IGHV-UM status in a study on 64 patients where 17% developed clonal evolution (Stilgenbauer et al. 2007).

With the application of array technology a more accurate genome-wide view is facilitated. A study of 22 patients with sequential samples revealed that 27% exhibited clonal evolution and complex subclonal patterns were seen with both linear and branching evolution of subclones (Braggio et al. 2012). Further studies on larger numbers of paired samples confirmed complex patterns of evolution and revealed that clonal evolution increased in the post treatment setting. While clonal evolution occurred in samples with IGHV-M status this was only in the setting of prior treatment which points towards treatment resistant clones bearing a selective advantage and driving disease progression (Gunnarsson et al. 2011). Certain genomic aberrations were seen more frequently, not unexpectedly, such as del(17p) and del(11q), in addition to developing biallelic 13q deletions from monoallelic deletions (Knight *et al*, 2012; Gunnarsson *et al*, 2011b).

1.4 Staging and Prognosis

Many factors influence clinical outcome, both in relation to disease progression and response to treatment in CLL. Prognostic markers are of value in broadly defining risk groups and predicting the natural course of the disease irrespective of treatment. In CLL, prognostic scoring systems have been continuously refined since the first prognostic staging systems in the 1970s.

1.4.1 Staging systems

Rai and Binet classified CLL based on peripheral blood count values and lymphadenopathy/organomegaly findings (**Table 1.1**). Both systems remain in use to this day, in routine clinical practice and for clinical trial enrolment.

Table 1.1 Staging Systems

BINET Stage		Features
A		<3 Lymphoid areas*
B		≥3 Lymphoid areas
C		Haemoglobin < 100 g/l or platelet count < 100 x 109/l
RAI Stage	**Risk group**	
0	Low	Lymphocytosis only
I		Lymphadenopathy
II	Intermediate	Hepatomegaly or splenomegaly + lymphocytosis
III/IV	High	Haemoglobin < 110 g/l or platelet count < 100 x 109/l

*The five lymphoid areas comprise: uni or bilateral cervical, axillary and inguinal lymphoid, hepatomegaly and splenomegaly

The value of the Rai and Binet staging systems is found mostly in the intermediate and high-risk groups, helping to identify patients requiring treatment. For stage A asymptomatic patients not requiring treatment, there are highly variable outcomes thus limiting it's application in this setting. Further differentiation with biomarker development aiming to improve prognosis in this broadly defined group are outlined below.

1.4.2 Somatic Hypermutation

Initially it was thought that CLL B cells were naïve B cells as the IGHV gene reflected the germline status and CD5+ B cells are seen early in B cell development. Cases of CD5+ CLL with IGHV-M were subsequently noticed

meaning that these cells had passed through the germinal centre and undergone SHM. It became clear that either IGHV-M *or* -UM genes can be found in CLL (Schroeder & Dighiero 1994; Fais et al. 1998). Restricted use of certain immunoglobulin heavy chain variable genes (VH) has also been reported in CLL with six VH genes accounting for over 50% of all gene usage (Kipps et al. 1989; Fais et al. 1998; Hamblin et al. 1999). As several CLL subsets, defined by usage of certain VH genes, have now been described in CLL this points to the potential of common antigens stimulating the BCR (Tobin et al. 2004; Messmer et al. 2004). The finding of an unfavourable prognosis in patients with a IGHV-UM was first reported in tandem by two groups in 1999 (Hamblin et al. 1999; Damle et al. 1999). This prognostic classification is robust and retains significance in the era of CIT (Fink et al. 2013). Initial reports of differential SHM patterns in subsets of CLL defined by cytogenetic lesions such as trisomy 12 in IGHV-UM cases and sole 13q deletion seen with IGHV-M cases reveal a different clonal origin perhaps for these CLL subtypes (Oscier et al. 1997). However an interesting area of research is how to define the hierarchy of prognostic indicators when discordant findings such as that of high risk del(17p) cases with IGHV-M arise and are shown to have significantly better prognosis than is typical of 17p(del) cases (Gladstone et al. 2012). Preferential use of the VH3-21gene is seen in CLL and two thirds of cases are found in the IGHV-M group (Tobin et al. 2003). This finding correlated with a decreased overall survival and for the purposes of prognostication is included with the IGHV-UM group (Thorselius et al. 2006).

1.4.3 Surrogate markers of IGHV status

Determining the IGHV mutation status is a laborious and expensive test to perform routinely in clinical laboratories, therefore surrogate markers of IGHV mutational status such as zeta associated protein 70 (ZAP70) and CD38 expression have been evaluated. CD38 is a trans-membrane glycoprotein that can be easily measured by flow cytometry. Initial studies revealed a correlation between IGHV mutation status (IGHV-UM) and a level of >30% CD38 positivity which reflected a poor prognosis (Damle et al. 1999; Ghia et al. 2003). Similarly for ZAP70, an intracellular protein tyrosine kinase expressed in some cases of CLL, a link between higher levels of ZAP70 on flow cytometry associated with an IGHV-UM status and it was proposed that this was a more reliable surrogate marker than CD38 positivity (Crespo et al. 2003; Orchard et al. 2004; Rassenti et al. 2004). As studies have progressed CD38 positivity alone has not proven to be a reliable proxy for IGHV mutation status (Thunberg 2001; Hamblin et al. 2000). Both markers continue to hold prognostic value when tested together and can classify CLL patients in to 2 risk groups with a double negative result equating to an excellent prognosis and double positive a poor prognosis (Schroers et al. 2005). The lack of inter-laboratory standardisation for ZAP70 measurement has precluded the widespread uptake of this prognostic test routinely in the clinic (Marti et al. 2006).

1.4.4 Serum markers and the lymphocyte doubling time

Numerous markers such as CD23, thymidine kinase and elevated β2microglobulin (β2-M) have predicted survival or progression free survival (Cramer & Hallek 2011; Knauf et al. 1997). While all three measurements have proven independent value in early stage disease, β2-M continues to have

relevance even in the era of CIT (Hallek et al. 1996). β2-M is an extracellular protein that interacts with the class I major histocompatibility complex. In CLL, elevated levels have been shown to correlate with poor prognosis and this has been shown at different disease stages (Molica et al. 1999). The independent prognostic significance was shown in Phase II trials with fludarabine-based treatment and more recently in the setting of CIT where a level of 3.5mg/L or greater correlated with a shorter PFS (Wierda et al. 2009; Hallek et al. 2010). However, the impact of β2-M when important genetic drivers of disease are taken into account has not been fully answered.

The lymphocyte doubling time (LDT) can inform on disease prognosis and is a simple parameter to follow. The LDT is defined as the length of time, in months, for the lymphocyte count to double in number. A more indolent disease course is seen when the LDT is greater than 12 months and while studies have proven an independent relationship between the LDT and prognosis, this has not been consistently proven (Schroers et al. 2005; Heintel et al. 2001).

1.4.5 Genetic Risk Stratification and Response Prediction

Genomic risk stratification has supplemented the clinical staging systems with the aim of further understanding the heterogeneity of CLL. As outlined in more detail below in subsection 1.4.6, treatment of patients with CLL has evolved in recent years and many more treatment options are now available. Advances in whole genome array and sequencing technologies are transforming response prediction by facilitating the identification of novel genetic markers that may direct treatment choice. Response prediction refers to the identification of a

biomarker that will inform on the outcome of a specific treatment (Zenz et al. 2010). In CLL, as with many cancers, a patient's response to treatment is a prognostic marker of overall survival. The development of pre-treatment response predictors however are of far greater value to the patient and are at the centre of personalised medicine (Alsolami et al. 2013). The genetic features of CLL and the prognostic implications of various findings have been outlined previously in section 1.3. The resounding finding across numerous clinical trials confirm the negative outcome of patients with T*P53* mutations or del(17p) when treated with chemotherapy or CIT regimens and this marker is truly a response predictor in this setting. While the finding of IGHV-UM confirms an aggressive disease course it does not currently influence the choice of treatment. Other genetic features do not inform on clinical decision making at present but with the advent of more sensitive next-generation technologies with the application of genome wide analyses this may change.

1.5 CLL Treatment and Outcomes

Advances in treatments for CLL over the past 15 years have been ground-breaking. Previously treatment, when indicated, was alkylator-based and had no impact on survival. Fludarabine treatment and subsequently the addition of monoclonal antibody treatment to the chemotherapy backbone have led to improved PFS and for the first time an impact on overall survival (Hallek et al. 2010). More recently second generation anti-CD20 antibodies and novel small molecule therapies have further expanded the treatment options available to CLL patients and appear to be active in high-risk groups.

Survival for the general population in the Western world has improved in the last number of decades (Mackenbach et al. 2013). For CLL patients the greatest improvement has been seen in Binet stages B/C and in patients <70 years of age (Brenner et al. 2008).

The advent of many new agents, namely BCR antagonists and anti-CD20 antibodies, has increased patients' choice of treatments and improved clinical outcomes in CLL. In particular, the addition of rituximab to the chemotherapy backbone has changed the natural history of CLL by improving overall survival for the first time (Hallek et al. 2010). However, many issues remain unresolved and the increasing use of more toxic therapeutic regimens demands better risk stratification. The question of early treatment versus active surveillance has re-emerged as an area of research interest and also whether achieving eradication of minimal residual disease (MRD) should become a treatment goal for patients. The treatment of patients with high-risk and purine-analogue refractory CLL remains challenging in clinical practice and optimal strategies for older patients aimed at quality of life rather than overall survival are under investigation.

1.5.1 Aims of treatment

Only 15% of patients diagnosed with CLL are under the age of 65. On the other hand, 66% of patients recruited into the German CLL trials were younger than 64 with only 3% of patients in the older than 75 year age group (Widhalm 2004). Western societies are confronted with a rise in the elderly population and an improvement in life expectancy. There is therefore a clinical need to design

specific treatments for older patients who often have multiple co-morbidities (Dores et al. 2007; Yancik 1997). It is critical to define the goals of therapeutic intervention from the outset together with patients and their families. Scoring systems such as the Cumulative Illness Rating Scale (CIRS) to establish and quantify co-morbidities have been validated in the elderly and are being used as part of clinical trials (Linn et al. 1968). On the other hand, in young patients without co-morbidities, curative options should at least be considered.

1.5.2 When to treat

Although treatment indications have been clearly defined including advanced Binet or Rai stages or patients with symptomatic disease, the exact time when treatment should be initiated can be subjective and dependent more on severity of symptoms then objective criteria. Early treatment is of no proven benefit at initial diagnosis and therefore the natural history of the disease can be observed (Dighiero et al. 1998). However such conclusions were based on alkylator agents with significant short- and long-term side effects including myelosuppression, genotoxicity and increased risk of second malignancies. In the era of novel, less toxic agents including kinase inhibitors, this question will be challenged. As a general principle enrolment in clinical trials where available should be the goal.

1.5.3 Measuring treatment response

1.5.4 Cheson criteria

Guidelines on assessment of response post treatment were developed in 1988 and are particularly important when reporting within clinical trials to enable a uniform assessment (Cheson *et al*, 1996). These criteria define parameters of complete and partial response as well as progressive disease based on constitutional symptoms, physical examination, full blood count parameters and bone marrow evaluation and continue to be in use today (Michael Hallek et al. 2008). Formal publication of updated guidance to account for the increasing lymphocytosis when undergoing treatment with novel agents is awaited (Hallek 2017b).

1.5.5 Minimal residual disease response assessment

The molecular eradication of CLL is possible with modern CLL treatment (Ghia 2012). Minimal residual disease (MRD) negativity defined by guidelines as a level of detection below 10^{-4} cells correlates with prolonged progression free and overall survival (Rawstron et al. 2001; Lamanna et al. 2009). This was confirmed in the German CLL8 trial when MRD negativity at three months after treatment associated with longer progression free survival (PFS) and overall survival (OS) independent of IGHV status, 17p deletion or treatment arm (Böttcher et al. 2012). MRD is currently applied to samples within the clinical trial setting to guide risk-adapted therapy in CLL.

1.5.6 Fit 'Go Go' patients

1.5.6.1 Frontline treatment

Over the last decade, considerable progress has been made in the treatment of physically fit patients with CLL. Purine-analogue combinations have improved treatment outcomes (Rai et al. 2000; Catovsky et al. 2007; Robak et al. 2010; Hallek et al. 2010). Importantly, we have witnessed a paradigm shift in the management of CLL changing for the first time the natural history of the disease. The German CLL8 study compared FC versus FCR and demonstrated that therapeutic intervention in CLL led to an improved overall survival in patients with CIRS scores of less than six (Hallek et al. 2010). On the other hand, neither fludarabine and cyclophosphamide combination (FC) nor FC with Rituximab (FCR) were effective at treating patients with del(17p). FCR is considered the new standard of care for fit patients with CLL in first line treatment. Fludarabine is not without side effects and prolonged

myelosuppression, neutropenic sepsis and an increased risk of second malignancies are concerns (Tam et al. 2008).

1.5.6.2 Treatment of relapsed disease

FCR combination treatment is also effective in the relapse setting for those relapsing 2-3 years after FCR bearing in mind the toxicities outlined above (Ghia & Hallek 2014). As novel agents of proven efficacy are receiving rapid regulatory approval outwith the clinical trial setting, these treatments are becoming the new standard of care in the relapsed setting (outlined in subsection 1.1.7.6).

1.5.7 High-risk patients

Patients with deletions of chromosome 17p or *TP53* mutations or purine analogue refractory disease have a poor prognosis and obtain only limited response to second line chemotherapy (Badoux et al. 2009). High risk CLL was defined by the European Bone Marrow Transplantation (EBMT) CLL transplant consensus (Dreger et al. 2007) as:

1. Non-response or early relapse (within 12 months) after purine analogue-containing therapy
2. Relapse (within 24 months) after purine analogue combination therapy or treatment of similar efficacy (ie, autologous stem cell transplantation)
3. Del(17p)/*TP53* mutation requiring treatment

Alternative treatments have been investigated including the second generation anti-CD20 antibody Ofatumumab and the anti-CD52 monoclonal antibody Alemtuzumab. Ofatumumab showed promise with efficacy in high-risk CLL as

a single agent (Wierda et al. 2011) and is under investigation with novel kinase inhibitors. Alemtuzumab treatment has proven efficacy with response rates ranging between 22% and 34% however the median overall survival times are short, between 10 and 19 months and with significant toxicity, namely infectious complications (Stilgenbauer et al. 2009; Osterborg et al. 1997; Cortelezzi et al. 2009). Despite the absence of randomised studies, this was the standard of care for patients with *TP53* disrupted or purine analogue refractory disease until targeted kinase inhibitors became available within clinical trials (see subsection 1.1.7.6).

1.5.8 Allogeneic Transplantation

In the era of novel small molecule inhibitors the role for allogeneic transplant is rapidly changing and is undertaken less frequently (see section 1.1.7.6). For younger patients without co-morbidities and high-risk CLL, bone marrow transplantation (BMT) to consolidate remission should still be considered(Schetelig et al. 2008). Numerous retrospective studies inclusive of high-risk patients reveal that long-term remissions are achieved for 30-40% of cases indicating that del(17p) loses its adverse prognostic significance in this therapeutic context (Sorror et al. 2008; Schetelig et al. 2008). Autologous PBSCT are not performed in CLL due to improved upfront treatments namely chemo-immunotherapy, the lack of overall survival benefit despite improved PFS and EFS, and the increased risk of development of myelodysplasia and acute myeloid leukaemia (Brion et al. 2012; Sutton et al. 2011).

1.5.9 Frail or elderly 'Slow Go' patients

1.5.9.1 Frontline Treatment

While FCR is the standard of care for fit, younger patients the significant toxicity profile makes purine-analogue based treatment unsuitable in the elderly or patients with comorbidities. Chlorambucil has been in use for the past 50 years (Galton et al. 1961). Approximately 70% of frontline patients are expected to obtain a response to chlorambucil, however complete remissions are low and the mean PFS is 18 months (Catovsky et al. 2007). There is no overall survival benefit for elderly patients treated with fludarabine versus chlorambucil (Eichhorst et al. 2009). Chlorambucil combined with Rituximab induces more responses with an overall response rate of>80% and longer PFS (Hillmen 2010), findings reciprocated in another phase II clinical trial. The addition of novel monoclonal antibodies including Ofatumumab and Obinutuzumab are under investigation in combination with Chlorambucil. Bendamustine, a purine analogue- alkylator hybrid compared favourably to chlorambucil in a frontline study for elderly patients (Knauf 2009). The study of combined Bendamustine with Rituximab in patients up to 81 years of age is being investigated in a Phase III trial compared to FCR (Eichorst 2009).

1.5.9.2 Treatment of relapsed/refractory disease

Second generation monoclonal anti-CD20 antibodies represent an attractive alternative for this group of patients. Patients with a PFS of over two years can be re-treated with the frontline single agent chemotherapy. Bendamustine, in combination with Rituximab (BR), also showed significant activity in relapsed/refractory patients where 37% of patients were over 70 years of age with a reported median PFS of 15 months (Fischer et al. 2011). However, the majority of patients experienced at least one grade 3 or 4 adverse event during the course of treatment. Patients with del(17p) and fludarabine refractory patients had the least benefit from BR treatment. Furthermore, the use of novel agents (discussed below) is becoming the standard of care.

1.5.10 Novel agents for fit and frail patients

Mechanism-based treatment of CLL with small molecules is improving outcomes in CLL patients. These have developed from a better understanding of CLL pathophysiology and well-conducted clinical trials. As outlined in Subsections 1.2.5 and 1.2.6 both the BCR signalling and apoptotic pathways are involved in CLL pathogeneisis. Bruton's tyrosine kinase (BTK) is a key enzyme in the B cell receptor signalling pathway. Mutations in BTK were discovered in the inherited condition of B cell dysfunction and recurrent infections known as X-linked agammaglobulinaemia (Tsukada et al. 1993; Vetrie et al. 1993). Ibrutinib, an orally bioavailable, non-selective, irreversible BTK inhibitor has shown activity in CLL and low grade lymphomas (Herman et al. 2011; Ponader et al. 2012). It has been approved for use in CLL, both upfront

for high risk patients and in the relapsed and refractory setting (O'Brien et al. 2014; Byrd et al. 2013; Burger et al. 2014). Clinical activity has been shown in all risk groups including *TP53* disrupted cases and is well tolerated in all age groups making it an option for both fit and frail patients (O'Brien et al, 2014). An increasing lymphocytosis from baseline is seen for the first few of months while rapid resolution of lymphadenopathy is typical as lymphocytes mobilise from tissues to the peripheral circulation (Hillmen 2011). Non-specific BTK inhibition may explain the side effect profile such as the bleeding tendency and diarrhoea, therefore second-generation BTK inhibitors under investigation may have better tolerability. Ibrutinib is under investigation in combination with chemotherapy regimens in addition to other novel agents aiming to achieve deeper responses and possible discontinuation of treatment at a point in the future analogous to the treatment approaches employed for CML in patients with sustained deep molecular responses to tyrosine kinase inhibitors (Saußele et al. 2016).

Inhibition of the phosphoinositide 3-kinase (PI3K) delta isoform, a downstream signalling kinase of the BCR pathway, has proven efficacious in CLL with the orally bioavailable small molecule Idelalisib (Herman *et al*, 2010). It has a dual mechanism of action and decreases cell survival signals via the BCR while down-regulating chemokines that facilitate key interactions between CLL cells and the microenvironment (Hoellenriegel et al. 2011). This agent combined with Rituximab has led to improved PFS compared with single agent Rituximab in a Phase III trial in the relapsed/refractory setting (Furman et al. 2014) leading to rapid regulatory approval. Idelalisib is also active in the frontline setting, is

suitable for elderly patients, and shows activity in high-risk patients (O'Brien et al. 2015; Coutre 2013).

BCL-2 is overexpressed in CLL and plays a critical role in the pathogenesis of the disease (Hanada et al. 1993) It is a pro-survival protein that inhibits apoptosis via the intrinsic pathway as detailed in Subsection 1.2.6. Numerous Phase I/II trials using small organic molecules of high affinity against the BCL-2 family, such as Oblimersen and Obatoclax, have been studied but have not been granted regulatory approval due to significant toxicity or lack of sufficient clinical impact (Abou-Nassar & Brown 2010; O'Brien et al. 2009; O'Brien et al. 2008). The second-generation BCL-2 inhibitor, Navitoclax (ABT-263), while active against CLL led to concomitant on-target thrombocytopenia caused by B cell leukaemia –xL (BCL-xL) inhibition and limited the use of this molecule (Tse et al. 2008; Roberts 2009). This molecule was re-engineered to create Venetoclax (ABT-199), a selective inhibitor of BCL-2, which shows promise in CLL without causing thrombocytopenia (Souers et al. 2013; Anderson 2013; Seymour 2013). Phase III trials combining this agent with other novel agents and monoclonal antibodies are underway.

1.5.11 Unmet clinical need – early relapses after frontline treatment

CLL treatment has changed dramatically in the past decade and thanks to chemo-immunotherapy remission durations of several years with improvement in overall survival have been achieved. However, not all patients benefit from current treatment strategies. Efforts from biological studies and well run clinical trials benefits from the plethora of new anti-cancer agents for both high risk and

elderly patients or patients with co-morbidities are being realised. To genetically define all patients and inform on risk stratification and response prediction that directs targeted therapies to the right patient is the goal of precision medicine. As outlined above a number of patients with high-risk disease can be defined by *TP53* disruption in the era of CIT however up to 20% of early relapses cannot be defined by this aberration(Hallek et al. 2010). Furthermore, certain groups of patients may be more sensitive to CIT and aiming to define this good risk group prior to receiving treatment could ensure durable remissions, even a cure, for select groups of patients with alternative non-CIT regimens for high-risk patients not solely those defined by a *TP53* disruption.

1.6 Next Generation Sequencing Applications in Cancer

For more than two decades sequencing technologies have advanced our knowledge of the cancer genome (Stratton et al. 2009). With the implementation of massively parallel sequencing (MPS) large amounts of DNA can be sequenced in parallel at a lower cost than the gold standard Sanger sequencing. While Sanger sequencing is a robust and highly reliable method the major limitations lie in the lack of throughput and cost (Rizzo & Buck, 2012).

Initial reports of whole genome sequencing studies across cancer and CLL revealed novel gene mutations and signaling pathways previously not linked to cancer (Ley et al. 2008; Shah et al. 2012; Desmedt et al. 2012; Puente et al. 2011; G. Fabbri et al. 2011; L. Wang et al. 2011a; Schuh et al. 2012). The greater coverage afforded by NGS allows the description of intraclonal dynamics within an individual tumour. When spatially or temporally sampled the

evolutionary process of individual cancers can be described and an understanding of the timing of mutation acquisition based on comparing samples from different compartments in one individual or across various time-points is possible (Anderson et al. 2011; Gerlinger et al. 2012; Schuh et al. 2012; Ding et al. 2012). Whole genome sequencing interrogates all of the genome, usually at a depth of 30-100X. For cancer samples both the tumour and germline sample can be sequenced with removal of the germline variants computationally from the data to allow investigation of somatic variants only. Whole exome sequencing investigates the coding genome, which represent only 1.2% of the overall genome and can be sequenced using capture arrays to allow a more targeted, sensitive, efficient and cost-effective investigation than WGS (Rizzo & Buck 2012; Sims et al. 2014).

Targeted NGS panels are also available whereby custom probes are designed to target genes/regions of interest within the genome with much greater sensitivity. This targeted approach facilitates a greater depth of sequencing as a much smaller percentage of the genome is interrogated (Fernandez-Mercado et al. 2013; Metzker 2010). This technique has led to improved knowledge on gene mutations and signalling pathways by screening large cohorts (Guièze & Wu 2015; Baliakas et al. 2015). For targeted NGS, if solely one time-point is available, intraclonal dynamics can be studied using the variant allele frequency (VAF) to infer the timing of mutations (Papaemmanuil *et al*, 2013).

The general principles of targeted NGS panels are based on MPS of an amplified DNA template. As with other applications the DNA template is a

double-stranded DNA construct consisting of short fragments of DNA flanked by adapters of known sequence. Illumina sequencing is the most broadly used NGS technique and consists of a solid phase amplification of the DNA template on a flow cell with subsequent sequencing-by-synthesis (Metzker 2010).

As laboratory techniques have evolved to produce high quality sequencing data on a massive scale, the challenge to accurately and efficiently analyse the vast quantities of genomic data remains. An explosion has taken place in the availability of bioinformatics tools and cloud based data storage to address current challenges accompanying the genomic revolution. Quality metrics and annotation programs have been developed to improve the reliability of results. International collaborations continue to compile cancer mutation databases, and guidelines are available to support the analysis and reporting of pathogenic mutations in clinical practice (Forbes et al. 2009; McVean et al. 2012; Richards et al. 2015).

1.7 Thesis Aims and Objectives

CLL is clinically and biologically heterogenous and it is realistic to hypothesise that further candidate mutations in CLL remain to be discovered. At the outset of my studies, an unmet clinical need in CLL related to the group of high-risk patients without *TP53* disruption who are resistant to CIT. The overall aim of this thesis was to address this need by developing novel response predictors in CLL for these patients and to avoid unnecessary toxicity of standard treatments that may be ineffective. This was facilitated by the availability of next generation sequencing and array technologies that enable the efficient and sensitive analysis of large numbers of CLL samples.

In Chapter, 3 the main aim of the work was to investigate whether one or more candidate genes discovered in CLL by applying WGS to longitudinal CLL samples in CLL might be involved in mechanisms of chemo-resistance that could not be explained by *TP53* disruption. WGS would be applied to longitudinal CLL samples to investigate this. Validation of any findings using NGS technology on larger numbers of CLL cohorts at different disease time-points would be undertaken in addition to *in vivo* functional analyses. In Chapter 4, the aim of defining the pre-treatment mutational landscape of CLL on a clinical trial cohort of 250 CLL patients and evaluating relevant interactions using integrated results from MPS and genome-wide array technologies was undertaken. Finally, in Chapter 5, the overall aim was to define genetic response predictors in the setting of CIT by survival analyses applied to the

findings of the CLL genomic landscape from Chapter 4. The methods, results and discussion of this work are presented in this thesis.

CHAPTER 2 : MATERIALS AND METHODS

2.1 Patient samples

For patients included in this thesis the diagnosis of CLL was made using the International Working Group on CLL defined criteria including the use of immunophenotyping (M. Hallek et al. 2008).

For Chapter 3 CLL samples were chosen from our institution initially (Oxford University Hospitals). For the WGS project three patients who exhibited a poor response to chemo-immunotherapy despite the absence of *TP53* abnormalities were selected. All three patients had unmutated IGVH (IGHV-UM) status, but were CD38 negative at diagnosis. All three patients had relapsed less than one year from purine-analogue therapy and underwent successive rounds of additional treatment. Five peripheral blood samples from diagnosis onwards and one buccal swab sample were submitted for WGS.

To validate *SAMHD1* mutations in a CLL cohort, samples were initially chosen from our institution (Oxford University Hospitals Trust, Cohort 1= 100 samples). Overall, the patient cohort was that of a tertiary referral centre with a high proportion of treatment refractory cases and of IGHV-UM gene cases.

In order to evaluate the frequency of *SAMHD1* mutations in a pre-treatment clinical trial context, 200 samples from two UK NCRN clinical trials were sequenced (Cohort 2: AdMIRe Trial=108, ARCTIC trial= 92). Sample selection was based on the first samples available to our laboratory from the UK National

CLL Biobank in Liverpool. AdMIRe is a randomized Phase II Trial of fludarabine, cyclophosphamide and rituximab (FCR) with or without mitoxantrone (M) in previously untreated physically fit patients with CLL (Munir et al. 2017). ARCTIC is a randomized phase IIB trial in previously untreated physically fit patients with CLL to compare fludarabine, cyclophosphamide and rituximab (FCR) with FC, mitoxantrone and low dose rituximab (FCM-miniR) (Howard et al. 2017).

A further 118 samples from relapsed/refractory clinical trial patients were evaluated for *SAMHD1* mutations. These included two UK NCRN relapsed and refractory trials (Cohort 3 n=63: CLL 201 n=37, CLL 202 n=26) and one French CLL Intergroup ICLL01 phase II clinical trial (Cohort 4 n=55). CLL201 is a randomised phase II multicentre trial to assess the efficacy and safety of fludarabine, cyclophosphamide and mitoxantrone in combination with rituximab (FCM-R), and fludarabine, cyclophosphamide and mitoxantrone alone (FCM) in previously treated but not fludarabine-refractory patients with CLL (Hillmen et al. 2007). CLL202 is a phase II non-randomised study of subcutaneous alemtuzumab plus fludarabine in fludarabine-refractory patients. Refractoriness was defined as no response or relapse within 6 months from fludarabine treatment. The ICLL01-BOMP Phase II clinical trial is salvage treatment with bendamustine, ofatumumab and high-dose methylprednisolone for fit relapsed CLL patients after 1-3 prior lines of therapy (de Guibert et al. 2014).

In Chapters 4 and 5 we studied 250 of 415 patients from two UK NCRN Phase II randomised controlled trials (AdMIRe and ARCTIC) receiving frontline FCR-based treatment as outlined above. The validation cohort discussed in Chapter

5 consisted of 247 of 615 CLL patients from the UK LRF CLL4 trial. This was a phase III randomised controlled trial for CLL patients requiring treatment who were randomly assigned to fludarabine (n=194) or fludarabine plus cyclophosphamide (196) for six courses, or chlorambucil (387) for 12 courses (Catovsky et al. 2007). A subset of samples was tested based on the availability of stored extracted DNA.

2.2 Research Ethics

Informed consent from CLL patients was obtained in line with the Declaration of Helsinki. Research ethics approval for use of samples obtained from the CLL biobank in Oxford is from Oxfordshire C REC, 09/H0606/5. Additional samples studied from the UK NCRI CLL clinical trials were obtained from the National CLL Biobank in Liverpool (Principal Investigator: Andrew Pettitt, Materials Transfer Agreement number 11/03/2011).

2.3 Ficoll Preparation of Peripheral Blood Mononuclear Cells

Whole blood or bone marrow collected in Ethylene Diamine Triacetic Acid (EDTA) tubes was transferred into 50ml falcon tubes ((Thermo Fisher Scientific, Waltham, MA, USA). Phosphate buffered saline (PBS) (10010031, Life Technologies, Carlsbad, California, USA) with 1% fetal bovine serum (FBS) (10270-106, Life Technologies, Carlsbad, CA) were added at 2.5x volumes. Four millilitres of Ficoll (17-5442-02, GE Healthcare, Little Chalfont, Buckinghamshire, UK) were added to 15ml centrifuge tubes. Ten millilitres of blood/PBS mix were layered on top of the Ficoll. The tubes were centrifuged for 20mins at 400g at 4°C. The upper layer was removed, whilst the white cell

fraction was transferred to a new 15ml tube. PBS was added to make a total volume of 14ml, then the tubes were centrifuged for 5mins at 280g. The supernatant was removed and 500µl of red blood cell (RBC) lysis solution (158904, Qiagen, Hilden, Germany) added. The tubes were incubated at room temperature for 10mins then made up to a volume of 10ml with Dulbecco's Modified Eagle Medium (DMEM) (11995-065, Life Technologies, Carlsbad, CA). The tubes were centrifuged for 5mins at 280g and the supernatant removed. The cell pellet was resuspended in freezing medium containing 5ml complete media (CM) (Lonza , Basel, Switzerland) (Media RPMI 1640 + 10%FBS+ 1%Pen/Strep+ 1%L-Glutamine), 4.5ml FBS (Gibco, Thermo Scientific, Waltham, MA, USA) and 0.5ml dimethyl sulfoxide (DMSO) (Radnor, PA, USA) to give a final concentration of 1-5x10^7 cells/ml. Cells were stored in 2ml cryogenic vials containing 1ml of cell preparation. Vials were stored initially inside an isopropanol box in a -80ºC freezer. 24 hours later the vials were transferred to a liquid nitrogen freezer.

2.4 Thawing of CLL samples stored in -80°C freezer

Frozen peripheral blood mononuclear cells were thawed prior to any further downstream procedure. Vials containing 1ml of cells frozen in freezing medium as described in section 2.3 above were thawed by gently agitating the vials in a 37° waterbath until all ice crystals had thawed. Cells were swiftly transferred drop by drop into 10ml of Dulbecco's Modified Eagle Medium (DMEM) (11995-065, Life Technologies, Carlsbad, CA) at room temperature. The sample is centrifuged for 5mins at 180g. The supernatant was removed carefully not

disturbing the pellet and then resuspended in Phosphate Buffered Saline (PBS) (10010031, Life Technologies, Carlsbad, California, USA).

2.5 Cell purification using fluorescence-activated cell sorting

PBMCs were stained and subsequently FACS sorted to improve the purity of the cells for particular downstream analysis including real-time PCR (Methods section 2.12) and whole exome sequencing of test CLL DNA from sample ID NY239 (section 2.14). Briefly the cells were thawed as outlined in section 2.4. A final volume of 1mL of PBS was added to resuspend the pellet and the sample was divided for cell staining and a further centrifuge step at 300g for 4min was undertaken. The cells were resuspended in 25-50µl of PBS and 10µl of in house dilutions of fluorophores anti-CD19PE and anti-CD5 APC (Becton Dickinson, Franklin Lakes, New Jersey, U.S.) were added to the sample. Samples were incubated at room temperature in the dark for 15min. Cells were washed twice in 3ml of PBS followed by centrifugation at 300g for 3 min, discarding the supernatant and then final resuspension of the sample in 500µl of PBS prior to cell sorting using the FACS Aria cytometer (Becton Dickinson, Franklin Lakes, New Jersey, U.S.). Post-sort purity was >95% checked by flow cytometry on 500 cells of the sample sorted.

2.6 DNA extraction from CLL samples

2.6.1 Phenol-Chloroform DNA extraction

Genomic DNA was manually extracted from isolated Buffy coat samples using phenol-chloroform. One millilitre of 2x lysis solution comprising 0.77M NH_4Cl (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 0.046M $KHCO_3$ (Sigma-

Aldrich Corporation, St. Louis, MO, USA) was added to the buffy coat sample tube. The sample was then inverted and rotated for 10 minutes on a rotary mixer to lyse the red blood cells. The lysed sample was centrifuged at 28,341g for 5mins. After discarding the supernatant, the cell pellet was re-suspended in 200µl lysing solution comprising 100mM NaCl (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 25mM EDTA pH 8.0 (Sigma-Aldrich Corporation, St. Louis, MO, USA) along with 30µl of 10% SDS (Promega Corporation, Madison, WI, USA) and 200µl of 10mg/ml proteinase K (Qiagen N.V., Hilden, Germany). The sample was mixed by pipetting and incubated overnight at 37^{o}C in a hotblock. The addition of 700µl of phenol (Sigma-Aldrich Corporation, St. Louis, MO, USA) and centrifugation at 28,341g for 5 min was performed to remove proteins. The aqueous upper layer was shifted to a new eppendorf tube (Sigma-Aldrich Corporation, St. Louis, MO, USA) containing 700µl of phenol and the centrifugation was repeated. The two wash steps were repeated, with 700µl of chloroform (Sigma-Aldrich Corporation, St. Louis, MO, USA) instead of phenol. The remaining upper aqueous layer was transferred to a clean 0.5ml eppendorf tube containing 800µl of cold (4°) 100% ethanol (VWR International, Radnor, PA, USA) and 1µl 7.5M ammonium acetate (Sigma-Aldrich Corporation, St. Louis, MO, USA). The mixture was centrifuged to collect the DNA pellet, washed in 500µl 70% ethanol, air dried and finally re-suspended in 20µl nuclease free water (ThermoFisher Scientific, Waltham, MA, USA).

2.6.2 Manual kit DNA extraction

The QIAamp DNA mini kit (51306, Qiagen, Hilden, Germany) isolates DNA from cells using silica membrane technology. This kit is ideal for processing multiple samples and involves minimal handling of sample. The mini kit is suitable for

processing 200µl of whole blood, buffy coat or up to 5 x 10^6 lymphocytes. Fresh or frozen sample can be used. The sample is lysed using buffers and enzymes that stabilise the DNA and increase binding to the silica-gel membrane present in the spin column. Wash buffers are used to remove protein and other contaminants over four steps, and the DNA is eluted from the column using a low-salt buffer (Buffer AE).

Cells were made up to 200µl with PBS in a 1.5ml microcentrifuge tube after the above outlined thawing process. 200µl of buffer AL and 20µl Qiagen protease were added to the microcentrifuge tube and the tube pulse-vortexed for 15s. The tubes were incubated at 56°C for 10mins. 200µl of 100% ethanol was added and the tubes pulse-vortexed for 15s followed by a brief centrifuge to remove drops from inside the microcentrifuge column. The mix was transferred to a QIAamp Mini spin column and centrifuged at 6000g for 1min. The spin column was transferred to a clean 2ml collection tube and 500µl Buffer AW1 added. The column was centrifuged at 6000g for 1min and then transferred to a clean collection tube. After carefully opening the spin column, 500µl Buffer AW2 was added and the column centrifuged at 20,000g for 3min. The column was transferred to a clean microcentrifuge tube and centrifuged at 20,000g for another 1min. The spin column was transferred to a 1.5ml microcentrifuge tube and 200µl Buffer AE added to the membrane. The Buffer AE was brought to room temperature prior to use. The column was incubated at room temperature for at least 1min, up to 5 minutes to increase yield, then centrifuged at 6000g for 1min and stored at 2-8°C.

2.7 DNA extraction from Saliva

Germline DNA was extracted from saliva samples provided by the patients using the Oragene DNA kits (DNA Genotek Inc, Ottawa, ON, Canada). This procedure describes the process for extraction of DNA from the entire sample collected. For sufficient mixing of reagent with each saliva sample the collection tube was inverted 10 times and then incubated at 50°C in a hotblock for a minimum of 1 hour or overnight. This step increases DNA yield and ensures that nucleases are permanently inactivated. The sample was transferred to a 15ml Falcon tube (Thermo Fisher Scientific, Waltham, MA, USA) containing a 1/25th volume of PT-L2P solution (DNA Genotek Inc, Ottawa, ON, Canada), mixed by vortexing for 10 seconds and incubated on ice for 10 minutes to precipitate any impurities. The sample becomes turbid as impurities and inhibitors precipitate out. The sample was then centrifuged at a minimum of 3,500xg for 10 minutes. The majority of the supernatant was transferred to a clean 15ml Falcon tube leaving a small volume of the supernatant behind to ensure that the pellet was not disturbed. The pellet was then discarded. To precipitate the DNA, a 1.2x volume of room temperature 100% ethanol (VWR International, Radnor, PA, USA) was added to the supernatant, followed by inverting 10 times to mix. The sample was incubated at room temperature for 10 minutes to allow the DNA to fully precipitate, before centrifuging at a minimum of 3,500xg for another 10 minutes. The supernatant was then carefully completely removed and discarded without disturbing the pellet. The remaining pellet was washed with the addition of 1ml of 70% ethanol (VWR International, Radnor, PA, USA) over the pellet and gently swirling. The latter was incubated at room temperature for 1 minute. Centrifugation at 3,500xg for

1 minute was used to facilitate complete removal of the supernatant. The DNA was resuspended with 500µl of TE buffer (10mM Tris-HCL, 1mM disodium EDTA, pH 8.0) (Sigma-Aldrich, St. Louis, MO, USA) and vortexed for 30 seconds. The DNA was incubated at room temperature overnight to ensure maximal resuspension of DNA, followed by a 30 second vortex and transfer to a labelled 1.5ml Eppendorf tube and stored at 2-8°C.

2.8 RNA extraction

Total RNA was isolated from sorted cell populations using RNeasy kits (Qiagen N.V. Hilden, Germany) that purify up to 100ug of total RNA. These kits purify RNA molecules longer than 200 nucleotides, which provides enrichment for mRNA. During the cell sorting process, described in Section 2.5, the sample was collected into a tube containing PBS stored on ice. The cell preparation was then transferred to 1.5ml microcentrifuge tubes, aliquoting 1ml per tube. The tubes were spun at 18,000g for 2mins then the supernatant discarded. The cells were resuspended in 600µl RLT buffer (Qiagen N.V., Hilden, Germany) and 10µl β-Mercaptoethanol/ml (Sigma-Aldrich Corporation, St. Louis, MO, USA). The sample was homogenised manually using a BD Microlance 21G x 1½ needle (Becton Dickinson, Franklin Lakes, New Jersey, USA). The RLT buffer contains guanidine-thiocyanate that protects RNA by destabilising RNases and contains a high concentration of salt to ensure optimal binding to the silica spin column. For optimal RNA yield the maximum threshold of cells per ml was $1x10^7$. Next, 600µl of 70% ethanol was added, and then two lots of 700µl of sample were placed in an RNeasy spin column and centrifuged at 13,226g for 15 seconds. The sample was then washed with 700µl RW1 buffer

(Qiagen N.V., Hilden, Germany) and centrifuged at 13,226g for 15 seconds. Two further washes were performed using 500µl RPE buffer (Qiagen N.V., Hilden, Germany) with centrifugation at 13,226g for 15 seconds after each step. Once the membrane had air-dried, 30µl of nuclease free water (ThermoFisher Scientific, Waltham, MA, USA) was added to elute the RNA with a final spin through the column, collecting the RNA into a new labelled collection tube, stored on ice, then at -80°C.

2.9 Complementary DNA preparation

Complementary DNA (cDNA) was generated from the RNA template directly after the RNA extraction process outlined above in section 2.8. A volume of 0.6µl of RNA in solution was transferred to a 0.5ml microcentrifuge tube and 2µl 5x VILO, 1µl 10X Superscript (both ThermoFisher Scientific, Waltham, MA, USA) and 1µl nuclease free water were added. The VILO reaction mix contains random primers, dNTPs and $MgCl_2$. The 10X Superscript in addition to containing Superscript III reverse transcriptase also contains a ribonuclease inhibitor and a helper protein.

Next, the tube was incubated at 25°C for 10mins, then 42°C for 60mins and finally the reaction terminated at 85°C for 5mins, within a T3 Biometra thermocycler (Biometra GmbH, Göttingen, Germany). The cDNA was made up to a final volume of 50µl with nuclease free water and stored at 2-8°C.

2.10 Nucleic acid quantification

2.10.1 Nanodrop spectrophotometry

The concentration and quality of DNA/cDNA/RNA samples was determined using a Nanodrop 2000 UV-vis spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). All nucleic acids absorb ultraviolet light at a wavelength of 260 nanometres (nm). Proteins absorb ultraviolet light also but at a wavelength of 280 nm. A ratio of absorbance at 260 and 280 (A260/A280) is used to determine a sample's contamination with proteins. With the measurement arm raised, 1.5µl of sample is pipetted onto the pedestal. By closing the measurement arm the sample is between the light source and the nucleic acid detector. Prior to commencing any sample measurements, 1µl of the solution used to re-suspend the test sample was tested. The Nanodrop 2000 software package (v1.6.198 ThermoFisher Scientific, Waltham, Massachusetts, USA) determines the quantity and quality of the DNA and records the results. The sample pedestal is cleaned with distilled water between each test sample and after use. An acceptable A260/A280 ratio was considered to be between 1.8 and 2.

2.10.2 Qubit fluorometric quantitation of DNA

The Qubit fluorometric quantitation method utilises fluorescent dyes that are specific to the target of interest, in this case DNA. This approach is more sensitive than spectrophotometric-based methods, as it is able to distinguish between DNA, RNA and other contaminates and can also quantify low levels of DNA. A Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA) was used to determine the concentration of extracted DNA samples. A working

dilution of Qubit reagent was prepared by diluting the Qubit dsDNA reagent by a factor of 1:200 in Qubit dsDNA Buffer. For each of the two standards, 190µl of Qubit solution was required, along with 199µl for each test sample. Ten microliters of each standard was added to separate tubes and mixed with 190µl of the Qubit working solution. For each test sample, 199µl of Qubit mix was added to 1µl of DNA. All samples were mixed by vortexing and allowed to stand at room temperature for 2 minutes. Measurements of the standards were taken first to establish a standard curve, followed by each sample in turn.

2.11 DNA Sequence Analysis

2.11.1 Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) produces millions of copies of a DNA region from a single copy. The reaction uses forward and reverse primers that anneal to the template DNA at complementary sites. There is an exponential increase in the number of DNA products with each subsequent product becoming a new DNA template. This reaction requires changes in temperature facilitated by a thermocycle to firstly denature double stranded DNA at a high temperature, next to anneal the primers and finally the correct temperature for Taq polymerase to extend the new DNA copy incorporating dNTPs from the reaction mix. All primers were designed, and melting temperatures were calculated, using Primer3 software (http://primer3.ut.ee). All primers were supplied by Sigma-Aldrich (St. Louis, MO, USA). The reaction mixes for PCRs referred to in this thesis were based on the following: each 25µl reaction included 50-250ng of genomic DNA, 12.5µl of 2x Qiagen Multiplex Mastermix (Qiagen, Hilden, Germany), 10pmol each of forward and reverse primers, with

the final volume made up to 25µl with nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA). Where the target sequence was GC-rich, 5x Q solution (Qiagen N.V., Hilden, Germany) was also included. PCR reactions were performed under mineral oil in either 0.5ml PCR tubes (Thermo Fisher Scientific, Waltham, MA, USA) or 0.2ml 96-well PCR plates (Corning Inc, Corning, NY, USA) covered with adhesive foil on a T3 Biometra thermocycler (Biometra GmbH, Göttingen, Germany). The standard cycling conditions were as shown below:

- Initial denaturation step at 96°C for 15mins
- 30 cycles of 1) denaturation at 92°C for 30secs, 2) primer annealing at 50°C (primer dependent) for 30secs, 3) primer extension at 72°C for 20secs
- Final extension at 72°C for 10mins

2.11.2 Agarose gel electrophoresis

PCR products were separated according to nucleotide length using agarose gel electrophoresis, which uses an electrical field to move the negatively charged DNA towards a positive electrode through an agarose gel matrix. Shorter DNA fragments are able to move through the matrix more efficiently than larger fragments, and so will migrate further. The length of the DNA fragment can be visualised by running molecular weight DNA ladders of known length beside the test sample. Agarose gels of 3% (w/v) were made by dissolving 6g of agarose (Sigma-Aldrich, St. Louis, MO, USA) in 200ml of 1x Tris-Borate-EDTA

(TBE) (Geneflow Ltd, Lichfield, UK) with 0.5µg/ml Ethiduim Bromide (Sigma-Aldrich, St. Louis, MO, USA) in a conical flask. The PCR product was mixed with 1/5 volume of 5x loading dye comprising 2.5% Ficoll (Sigma-Aldrich Corporation, St. Louis, MO, USA), 0.01% bromophenol blue (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 0.01% xylene cyanole (Sigma-Aldrich Corporation, St. Louis, MO, USA) and loaded into a gel well. A DNA marker appropriate for the fragment sizes in a 1x TBE running buffer (Geneflow Ltd, Lichfield, UK) was loaded in a separate well. A voltage of 120V was applied across the gel for 45 to 60 minutes. The gels were placed on a UV transilluminator (Bio-Rad, Hercules, CA, USA) to visualise the DNA under ultra-violet light (302nm wavelength) and the images printed using a Mitsubishi P93D thermal printer (Mitsubishi Electric Corporation, Tokyo, Japan).

2.11.3 Purification of PCR Product

Following PCR amplification, to remove reaction components and other small low molecular weight DNA fragments prior to downstream processes, a clean up step is necessary. 10µl of PCR product was mixed with 10µl of MicroClean reagent (Microzone Ltd, Haywards Heath, UK) and mixed by pipetting. The sample was then centrifuged at 13,000rpm for 40 minutes to collect the pellet, then spun inverted for 10 seconds at a very low speed (100rpm) on a stack of paper towels to gently remove the supernatant. The pellet was re-suspended in 12µl of nuclease free water and allowed to stand for 10 minutes to re-hydrate before being processed as per Section 2.11.4.

2.11.4 Capillary Sequencing

DNA sequencing was based on the selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) known as the chain termination method or Sanger sequencing (Sanger et al. 1977). The sequencing reactions were

performed using an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). PCR reactions were performed as described in section 2.11.1.

2.11.5 Sequencing Amplification

As distinct from PCR this amplification uses only one primer and the DNA product is produced in the same direction as the primer. The amplification is linear rather than exponential. In addition ddNTPs are incorporated randomly into the DNA product terminating the extension whereas in a PCR reaction the dNTPs continue to be incorporated along the DNA template. The ddNTPs are tagged with fluorescent dyes that will provide the signal when separated downstream in the capillary analyser. Each 10µl Sanger sequencing PCR reaction contained 1µl of purified PCR product (as described in sections 2.11.2 and 2.11.3) and 9µl of reaction mix. The reaction mix contained 2µl 2.5x ready reaction premix (Applied Biosystems, Foster City, CA, USA), 1µl 5x BigDye sequencing buffer (Applied Biosystems, Foster City, CA, USA), 1µl of 3.2 µM sequencing primer and 5µl of nuclease free water (ThermoFisher Scientific, Waltham, MA, USA). Each sequencing primer had an M13 tail with the following sequence: forward: 5'-GTAAAACGACGGCCAGT-3' and reverse: 5'-CAGGAAACAGCTATGAC-3'. Cycle sequencing reactions were performed in 0.2ml 96-well plates, under mineral oil, using a T3 Biometra thermocycler (Biometra GmbH, Göttingen, Germany). The cycling conditions were as follows:

- Initial denaturation step at 96°C for 1min

-25 cycles of 1) denaturation at 96°C for 1 min, 2) primer annealing at - 50°C for 5secs, 3) primer extension at 60°C for 4mins

- Hold thermocycle at 4°C

2.11.6 Ethanol Purification

Samples were purified to remove surplus PCR reagents using this technique. 5µl of 125mM EDTA pH 6.0 (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 60µl of 100% ethanol stored in the freezer (VWR International, Radnor, PA, USA) were added to each cycle sequencing reaction. The sample was mixed by inversion 10 times and incubated at room temperature for 15 minutes. The plate was centrifuged at 2296g for 30 minutes at 4°C. The plate was then inverted onto paper towel and spun at 100g for 1 min to remove the supernatant and washed again in 60µl of ice-cold 70% ethanol. The samples were centrifuged at 2054g for 15 minutes, the supernatant removed and the samples allowed to air-dry. The dried samples were re-suspended in 8.5µl of Hi-Di formamide (Applied Biosystems, Foster City, CA,USA) prior to sequencing as outlined in Section 2.11.4.3.

2.11.7 DNA sequencing Data Analysis

Sequencing analysis software (v5.4, Applied Biosystems, Foster City, CA, USA) was used to analyse the raw sequencing data. This software displays, edits, saves and prints the data generated by the genetic analysers. The software performs base calling and generates quality values for each base and sample file. The output files from this process include text files of the DNA sequence and analysed ABI files, either of which can be used for downstream analysis. For the identification of variants in the DNA sample, Mutation Surveyor software (SoftGenetics, State College, PA, USA) was used. This software takes analysed ABI files and performs a direct base-to-base comparison against a FASTA file of the wild-type sequence. Visual inspection of sequences was performed in addition to automated analysis.

2.12 cDNA sequencing in selected SAMHD1 cases

Primers were designed to amplify the complementary DNA of particular exons of *SAMHD1* (**Table 2.1**). Purified amplicons were sequenced as detailed in the Sanger sequencing method above Section 2.11, to confirm the presence of the mutations found in the corresponding DNA samples.

Table 2.1. Primer sequences for SAMHD1 sequencing

Gene_Exon	Forward Primer	Reverse Primer
SAMHD1_exon12	AGCAATTGACGACATGGAAG	CTGGCAACCTCTTTTGGAAG
SAMHD1_exon15	TAAGACTGCCCCCAACAGAG	TGAGTGGGGCTATAACATCG

2.13 Quantitative real-time PCR

Absolute quantification of cDNA was performed in duplicate using real-time PCR and a transcript quantification TaqMan Gene Expression assay kit for *SAMHD1* (Hs00210019_m1, Applied Biosystems, Foster City, CA, USA) and *ABL1* (Hs01104728_m1, Applied Biosystems, Foster City, CA, USA) as a normal control. cDNAs from sorted CLL samples and B lymphocytes from healthy donors (Section 2.6) were used in these analyses. Two microliters of cDNA were included in 10µl reaction mixes using Taqman MasterMix (4304437, ThermoFisher Scientific, Waltham, MA, USA) and the corresponding Taqman Gene Expression probe. Detection was performed using the Rotorgene 6000 system (Qiagen, Hilden, Germany).

PCR cycling conditions were as follows:

- Initial 2min incubation at 50°C
- Denaturation of 95°C for 10mins
- 50 cycles of 95°C for 15sec and 60°C for 1min

Relative quantification of the transcripts was performed following normalisation using the control gene, *ABL1,* using the Comparative Quantification method of the Rotor Gene 6.0 software (Qiagen, Hilden, Germany). In this method, the Ct is defined as the point at which the exponential phase of amplification begins and the reaction efficiency is calculated between this point and the maximum rate of exponential amplification. From this the level of individual transcripts relative to the reference sample can be calculated using the formula:

$$\text{Relative Quantity} = \text{Efficiency}^{\text{ReferenceCt-SampleCt}}$$

From this the relative expression ratios of *SAMHD1* to *ABL* for each sample could be determined (McCurdy et al. 2008).

2.14 Whole genome sequencing

2.14.1 WGS library generation and sequencing

Staff at Illumina technologies, Chesterford, Cambridge, UK performed WGS of 18 samples from three patients as detailed here and based on methods published initially by Bentley (Bentley et al. 2009). Cells purified by Ficoll centrifugation (section 2.3) were used for tumour DNA extraction (section 2.7.2) and DNA was extracted from saliva samples (section 2.8) for germline material. Genomic DNA libraries were generated starting from 1µg of sample resulting in 330bp insert fragments ligated to adaptor sequences. Clusters were formed with the TruSeq V3 Amplification protocol (Illumina, San Diego, CA, USA). SBS sequencing was performed using the TruSeq V3 SBS kit (Illumina, San Diego, CA, USA) with 100bp paired end reads using the HiSeq 2000 (Illumina, San Diego, CA, USA). The HiSeq 2000 is designed to perform multiple cycles of sequencing chemistry and imaging to collect sequence data automatically from each cluster on the surfaces of each lane of two eight-lane flow cells. The base-calling takes place in close to real time on instrument using RTA software version 1.10.28. (Illumina, San Diego, CA, USA).

2.14.2 Analysis and Somatic Variant Calling

The standard Illumina CASAVA v1.8 pipeline (Illumina, San Diego, CA, USA)

was used for quality control, alignment to human GRCh37.1 reference and variant calling for single genome analysis (McLaren et al. 2010). All germline mutations were subtracted from the matched tumour, and the results were exported as both variant call files (VCF) and binary alignment map files (BAM). VCF files contain a list of all mutations detected in a particular sample, while BAM files contain information about the sequence alignment, which can be visualised using software such as the Integrated Genomics Viewer (Broad Institute, Cambridge, MA, USA). Annotation of tumour-specific mutations was performed using the Ensembl database (Flicek et al. 2011).

2.14.3 Somatic Single Nucleotide Variant Time-series Analysis

Variant allele frequencies (VAF) were calculated at each tumour stage for tumour-specific point mutations detected in at least one time point. Positions with less than 25x coverage in at least one tumour stage were discarded to remove background noise. Finally, the somatic SNVs were clustered into groups showing similar VAF profiles with a k-means algorithm (Lloyd 1982).

2.15 Whole Exome Sequencing - workflow

For patient NY239 WES of purified CD19+, CD5+ B cells (section 2.6.) was performed by the Core Genomics Facility at the Wellcome Centre for Human Genetics, Oxford, UK. In brief, the sample was quantified using the Qubit system (section 2.10.2) and sample integrity was assessed using 1% E-Gel® EX (ThermoFisher Scientific, Waltham, MA, USA). Library preparation was performed using the Illumina TruSeq DNA HT Sample Preparation kit (Illumina,

Inc., San Diego, California, USA) and enriched for the exome using the Illumina TruSeq Exome Enrichment kit (Illumina, Inc., San Diego, California, USA). Sequencing was performed using a HiSeq2000 (Illumina, Inc., San Diego, California, USA) as 100bp paired end reads. The sequence reads were mapped to the human genome assembly hg19/NCBI37 using Stampy (Lunter & Goodson 2010) and variants were detected using Platypus (Rimmer et al. 2014). High quality reads scoring 200, with sequence variations found within coding exons and a total coverage of at least 50 reads were analysed further. The database for normal variants (dbSNP) and the 1000 Genomes database (1000 Genomes Project Consortium et al. 2012) were used to exclude common variants. We searched for any published CLL-associated mutations (Quesada et al. 2011; Landau et al. 2013; Schuh et al. 2012; Puente et al. 2011; Skowronska et al. 2012; T Zenz et al. 2010) present in the leukaemia cells and excluded their presence in the germline by Sanger sequencing.

2.16 Targeted Next Generation DNA Sequencing

The Illumina TruSeq Custom Amplicon assay (TSCA) (Illumina, San Diego, CA, USA) is an amplicon-based next generation sequencing method to amplify and sequence a large number of targets in multiple patients simultaneously. This method was used to obtain the mutational frequencies of key CLL driver genes during my thesis as described in the following section 2.15.1.

2.16.1 TSCA panel design

This panel included nine genes involved in CLL pathogenesis (**Table 2.2**). The genes for the TSCA panel were selected based on the evidence available at the time of design. Both *TP53* and *ATM* genes were included, in addition to

NOTCH1, *SF3B1, BIRC3* and *MYD88*, genes more recently found to be recurrently mutated in CLL (G. Fabbri et al. 2011; L. Wang et al. 2011b; Puente et al. 2011; Quesada et al. 2011; Rossi et al. 2012). A pilot whole genome sequencing project had highlighted also a further three genes with predicted pathogenic mutations of potential relevance in CLL: *SAMHD*1, *MED12* and *XPO1* which were also included in the sequencing panel (Schuh et al. 2012). The custom oligonucleotide pool was designed using the online DesignStudio tool (http://designstudio.illumina.com). Certain genes with well defined mutational hotspots had only a selection of exons sequenced as outlined in **Table 2.2** while others had all exons included as these were either less well studied at the time of panel design or had no known hot spots. The panel was designed to avoid regions of known polymorphisms as part of the primer sequences.

Table 2.2 Genes included in TruSeq Custom Amplicon Panel

Gene	Chr	Position (bp; hg19)	Targets	# Amplicons
SF3B1	2	198256648 - 198299821	Exons 14, 15, 16 & 18	8
XPO1	2	61704984 - 61765761	Exon 15	2
MYD88	3	38182608 - 38182792	Exon 5	1
NOTCH1	9	139390508 - 139392025	Exon 34	9
ATM	11	108093509 - 108239876	All Exons	113
BIRC3	11	102188144 - 102208515	All Exons	38
TP53	17	7571670 - 7590913	All Exons	21
SAMHD1	20	35520177 - 35580296	All Exons	26
MED12	X	70338356 - 70362354	All Exons	78

2.16.2 TSCA methodology

Dual-barcoded libraries were generated from 250ng of input DNA using the TSCA kit (Illumina, San Diego, CA, USA) before undergoing 2x150bp paired-end sequencing using the MiSeq platform (Illumina, San Diego, CA, USA) as per the manufacturer's protocol, overviewed briefly as follows.

2.16.2.1 Overview

The oligo pool was hybridised to the genomic DNA template starting at a temperature of 96° decreasing to 40° for an 80 minute incubation step to ensure binding of all oligos to upstream and downstream targets. Removal of unbound oligos used a size-selection filter and three wash steps. The extension and ligation step that follows joined the upstream and downstream oligos. Next the

library was amplified to simultaneously add the indexes (i7,i5) for downstream identification of samples and also to add sequences required for cluster formation. A clean-up step using beads allowed the purification of PCR product from other reagents/contaminants. Next the normalisation step allowed equal representation of samples in the pooled library (pooling of libraries is the process of mixing all samples in equal amounts into one collection tube prior to sequencing). A DNA denaturing step was required prior to loading of pooled samples onto the MiSeq instrument.

2.16.2.2 Detailed protocol using TSCA kit

Following the Illumina online protocol found at https://support.illumina.com the DNA samples were quantified by Qubit (section 2.1.4) and normalised to 25 ng/µl with a final volume of 10µl. The oligos were hybridised to the template DNA by adding 5µl of Custom Amplicon Oligo mix to 10µl of DNA sample and 35µl of OHS2 buffer in a microplate on a heat block set to 95°C and cooling down to 40°C over 80 minutes. The plate was centrifuged for 1 minute at 1000g and samples were transferred on a filter plate mounted on a midi collection plate. The plate unit was centrifuged at 2400g for 2 minutes and washed twice with 45µl buffer SW1 and once with buffer UB1 followed by 2 minutes centrifugation at 2400g. The extension and ligation steps followed next by adding 45ul of ELM4 to each sample and incubating at 37°C for 45mins. After incubation the plate was centrifuged for 2mins at 2400g. On the filter plate 25µl of freshly prepared 50 mM NaOH (72068, Sigma-Aldrich, St Louis, MO, USA) were added to each reaction and incubated for 5mins at room temperature. The index amplification plate (IAP) was prepared during this time with 4µl i5 tag, 4µl

i7 tag, 21.6µl PMM2 and 0.4µl TDP1 into each well. Next 20µl of each sample was transferred to the IAP plate and centrifuged at 1000 g for 1 minute. The amplification conditions consisted of the following:

- Initial denaturation step of 95°C for 3 minutes

- 25 cycles of 1) denaturation at 95°C for 30 seconds, 2) primer annealing at 66°C for 30 seconds and 3) primer extension at 72°C for 60s

- 5 minute final extension at 72°C.

5 µl of the PCR product was run on an electrophoresis gel as described in section 2.11.3. The expected PCR product size for an amplicon size of 250bp was ~350bp. The remainder of the DNA sample was transferred to a new midi plate with 45µl of AMPure XP beads. After shaking the plate at 1800 rpm for 2mins it was incubated at room temperature for 10mins. Then, the plate was placed on a magnetic stand and incubated until the liquid became clear. The supernatant was removed and discarded. Next 200µl of freshly prepared 80% ethanol (VWR International, Radnor, PA, USA) was added, the plate was incubated on a magnetic stand for 30secs and the supernatant was discarded. This step was repeated once. The remaining ethanol was removed using a 20µl pipette. The plate was air dried for 10 minutes. Next 30µl of EBT were added to each well, vortexed at 1800 rpm for 2mins ensuring all beads were resuspended and incubated at room temperature for 2 minutes. The plate was placed on a magnetic stand (Illumina, San Diego, CA, USA) until the liquid was clear and 20µl of the supernatant from each well was transferred to the corresponding well of a new microplate and centrifuged at 1000 g for 1 minute. The libraries were normalized to have equal representation of each library in the pool by incubating a pre-mixed solution of 38µl LNA1 + 7µl LNB1 per

sample. The plate was vortexed for 30mins at 1800 rpm and placed on a magnetic stand until the liquid cleared. The supernatant was discarded and samples were washed with 45 μl of LNW1 buffer, shaken at 1800 rpm for 5mins, placed on a magnetic stand and the supernatant was discarded. This step was repeated once. A volume of 30μl freshly prepared 0.1 NaOH (72068, Sigma-Aldrich, St Louis, MO, USA) was added to each sample and the plate was shaken at 1800 rpm for 5mins before being placed on the magnetic stand until the liquid cleared. Next 30μl of the supernatant from each well were pipetted into a new plate containing 30μl of LNS2. Combining 5μl of each normalized sample into a single tube resulted in the final pooled amplicon library (PAL).

2.16.2.3 Targeted paired-end sequencing on MiSeq

Using the MiSeq Reagent Kit v2 300-cycle kit (MS-102-2002, Illumina, San Diego, CA, USA) the final pooled library, PAL, from the above Section 2.15.2.2 was sequenced. A diluted amplicon library made up of 6μl of PAL and 594μl of buffer HT1 was loaded into the reagents cartridge, allowing cluster generation to proceed. During this process, fragments of DNA from the library bind to the flow cell by oligos complementary to the library adapters, bridge amplification then takes place and clonal clusters are formed across the flow cell. Sequencing by synthesis is then performed in the MiSeq instrument (SY-410-1003, Illumina, San Diego, CA, USA). Paired end sequencing using a reversible terminator-based method detects bases as they are incorporated into the DNA template strands. This proceeds for 150 cycles for the DNA insert, eight cycles for index i7, eight cycles for index i5, and finally 150 cycles for the DNA insert.

2.16.3 TSCA analysis pipeline

Initial alignment and variant calling analysis was performed according to the TSCA Workflow that used the MiSeq Reporter v.2.4 (Illumina, San Diego, CA, USA). The sequence reads were demultiplexed and aligned to the reference sequences using a Smith-Walterman algorithm. A second alignment followed by variant calling using the packages Stampy (v. 1.0.22) (Lunter & Goodson 2010) and Platypus (v.0.5.1) (Rimmer et al. 2014) respectively to detect additional small insertions and deletions was used. High confidence functional sequence variants were identified after merging VCFs from both alignment and variant calling methods and were annotated using Annovar (Wang et al. 2010). The following databases were used: hg19 refGene, dnSNP v.132 and v.137, 1000 genomes project v. April 2012 (1KGP), COSMIC v.67 and prediction scores SIFT and Polyphen (from ljb2). The variants were filtering on quality, predicted functional consequence, predicted pathogenicity by *in silico* algorithms SIFT and Polyphen, previously published in the CLL literature, or in the COSMIC database, or absent from 1000 genomes project. At least 50 reads at the variant position and a variant allele frequency of 0.05 or greater were required in order to progress a variant into the next stage of analysis.

For *ATM* mutations, those reported to be causative in ataxia telangiectasia patients even if present in germline samples were included, and similarly for *SAMHD1* mutations if present in Aicardi-Goutieres patients (Skowronska, Austen, et al. 2012; Stankovic et al. 1998; Rice et al. 2009). All candidate variants were confirmed by visual inspection using the Integrated Genomics Viewer (IGV)(https://www.broadinstitute.org/igv/ n.d.).

2.17 Sensitivity analysis

A dilution experiment as a means of testing the sensitivity of sequencing on the MiSeq instrument was performed. Sample ARC604 with eight mutations was selected. The test sample was spiked with another DNA sample wild-type for these mutations. The undiluted samples and nine spiked samples were prepared according to the following dilution series: 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256 and sequenced using our targeted NGS panel. A correlation between expected VAF (based on sample dilutions) and the observed VAF (number of altered reads / total number of reads) using Pearson's correlation was undertaken to analyse the results.

2.18 Validation and reproducibility of mutations found using TSCA panel

2.18.1 Validation of mutations

For all CLL samples with apparently novel mutations for whom matched germline samples were available, primers were designed to validate the novel mutations (Appendix Table 1). The Sanger sequencing methodology used for validation is outlined above in section 2.11.4.

2.18.2 Reproducibility

Seven CLL samples were sequenced both before and after the nine MiSeq runs to assess reproducibility and examine for the presence of carry-over between runs. The libraries were prepared by two different operators using two different

batches of pool of oligos, and sequenced using two different Miseq instruments. Employing our bioinformatics pipeline we found 100% concordance between both reports: the eight variants in the initial filtered VCF files were present in the second report, with no additional variants found

2.19 Definition of CLL Driver lesion

A CLL driver lesion refers to a recurrent mutation or CNA that has been published previously in a clinical context as being associated with CLL and may not necessarily be a proven biological driver lesion (Schuh, Becq, Humphray, et al. 2012; Papaemmanuil et al. 2013). The number of CLL drivers represents the number of mutations in any of the nine genes of the sequencing panel added to the number of chromosomal CNAs or cnLOH in the following regions: 13q14, 11q22, chr12, 2p, 17p13, 14q32, 6q21, 8q24, trisomy 12, 18 or 19.

2.20 Genome-wide SNP array

Hybridisation to Illumina Infinium arrays (HumanOmni1-Quad, HumanOmniS-8 platforms and HumanOmni2.5-8) was performed according to manufacturer's protocols found on registration at: http://www.illumina.com (Illumina, San Diego, CA, USA). These platforms allow the reliable detection of copy number losses, gains and copy neutral loss of heterozygosity. Unless otherwise stated, the reagents used for this protocol were provided with the genome-wide SNP arrays.

2.20.1 Genome-wide SNP array hybridisation overview

The input DNA was denatured and isothermally amplified in an overnight step using whole-genome amplification. Following this the product was fragmented and then precipitated. The sample was hybridised to the BeadChip in an overnight step allowing the individual DNA fragments to undergo complementary annealing to one of over one million beads. A wash step was followed by staining and elongation of hybridised sample prior to imaging using the Iscan (Illumina, San Diego, CA, USA) which deploys a laser to excite the fluorophore of the single-base extension product on the beads and records also a high-resolution image of the light emitted.

2.20.2 Genome-wide SNP array hybridisation in detail

Template DNA was quantified using the Qubit as outlined in section 2.1.4 and denatured by adding 0.1M NaOH (72068, Sigma-Aldrich, St Louis, MO, USA) and incubated at room temperature for 10mins. The denatured DNA underwent whole-genome amplification by adding 68µl MA2 and 75µl MSM to each sample, and incubating overnight for 20-24hrs at 37°C in a hybridisation oven. The amplified samples were fragmented by adding 50µl FMS and incubating at 37°C for 1hr. Next, the sample was precipitated by adding 100µl PM1 and incubating at 37°C for 5mins. 300µl of isopropanol (I9516 Sigma-Aldrich Corporation, St. Louis, MO, USA) was mixed well and then the sample incubated at 4°C for 30mins followed by centrifugation at 3,000g for 20mins. The supernatant was removed quickly by smacking down on an absorbent pad and then firmly tapping to ensure that the wells became devoid of any liquid. The pellets were air-dried for 1hr and resuspended in 46µl of RA1 followed by incubation at 48°C in a hybridisation oven for 1hr. Samples were next

hybridized to the BeadChips by rocking in a hybridisation oven at 48˚C for 16-24hrs prior to placing the BeadChips into a holder and washing the BeadChips twice in a glass tub containing PB1 reagent. The hybridised primers were extended with labelled nucleotides and then stained in a multi-layer staining process on the Chamber Rack. Finally the BeadChips were washed in PB1 and coated with XC4, vacuum- dried for 1 hour and inserted into the iScan instrument (Illumina, San Diego, CA, USA). Scanned images were downloaded and analysed as per Section 2.19.3 below.

2.20.3 Genome-wide SNP array analysis overview

The iScan data were processed using GenomeStudioV2009.2 (Illumina, Inc., San Diego, California, USA) and analyzed using Nexus 6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, California, USA). Analysis was guided by Nexus Discovery Edition v6.1 calls and only alterations that were visible on manual inspection of the data plots were reported. All losses and gains that had been noted previously in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home n.d.) or in unrelated 'in house' datasets were excluded from the analysis. A further exclusion step of CN events <100kb in size was applied unless the region of change had been noted previously in CLL or if a mosaic alteration was present confirming an acquired status. For reporting copy neutral loss of heterozygosity (cnLOH) a threshold of ≥10Mb was used based on 1) the absence of any region of homozygosity ≥10Mb in 1200 aged matched controls 2) the observation that mosaic cnLOH events in

the CLL cohort were always >10Mb and 3) use of similar criteria in the literature (Edelmann et al. 2013).

2.21 Subclonal evaluation

A mosaic or subclonal region of cnLOH is defined as the occurrence of a region of allelic imbalance in the absence of a copy number change. Allelic imbalance was detected using the B allele frequency (BAF) information from the whole genome SNP array visualised using Nexus v6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, California, USA). This was defined as present when the heterozygous (AB) allele frequency deviated from the expected 0.45-0.55 threshold range but did not reach full AA or BB homozygosity (close to 0 or 1 BAF respectively). For subclonal analysis of samples in Chapters 4 and 5, samples were excluded from the analysis based on potential lack of purity of sample rather than correcting for a sample impurity. Samples with lymphocyte purity (lymphocyte count/total white cell count) of <90% were excluded from the subclonal analysis (n=33). As many samples contained both clonal and subclonal lesions on array, in addition to solely clonal lesions, the purity of these samples was confirmed and these samples were included in the subclonal data analysis.

A subclonal lesion was defined if it was present at 90% or less on whole genome array by OncoSNP quantification (Yau et al. 2010) and/or visual inspection using Nexus v6.1 Discovery Edition (BioDiscovery, Inc., El Segundo, California, USA). To analyse subclonal mutations from the targeted deep sequencing (TDS) data the variant allele frequencies (VAF) of mutations were firstly corrected for local copy number change, integrating the copy number

array data with the sequencing data. A lesion was defined as subclonal if it was present at a VAF of <0.45 from the TDS panel. Using this method to correct for copy number and exclude potentially impure samples, 217 of the 250 samples presented in Chapters 4 and 5 were analysed for subclonal lesions from both whole-genome array and TDS data.

2.22 Cell lines and expression constructs

This method and the following Sections from 2.22 to 2.28 inclusive were undertaken by my colleague Dr Tania Louis, Laboratory of Molecular Virology, Institute of Human Genetics, Montpelier, France.

Adherent and suspension cells were cultured in DMEM (11995-065, Life Technologies, Carlsbad, CA) or RPMI-1640 (Lonza, Basel, Switzerland) respectively, supplemented with 10% FBS (10270-106, Life Technologies, Carlsbad, CA), ultraglutamine (Lonza, Basel, Switzerland) and antibiotics (Lonza , Basel, Switzerland). HeLa cells stably expressing Flag-HA-tagged SAMHD1 have been described previously (Laguette et al. 2011; Nakatani & Ogryzko 2003). HeLa cells stably expressing inducible SAMHD1 wild-type (wt) and mutants were generated by transducing HeLa cells with lentiviral pTRIPZ vector (Thermo Scientific, Waltham, MA, USA) in which the *SAMHD1* open reading frame (ORF) was inserted. Transduced cells were selected with 2µg/mL puromycin (Lonza, Basel, Switzerland). *SAMHD1* expression was induced with 2µg/mL Doxycycline (Clontech, Saint-Germain-en-Laye, France).

2.23 Plasmids

The *SAMHD1* wild-type plasmid used to produce HeLa cell lines has been described previously (Laguette et al. 2011). The SAMHD1 HD/AA mutant and F575 ΔCter mutant has also been described previously (Laguette et al. 2012).

2.24 Cell extracts preparation and western blot analysis

For western blotting, cells were lysed for 30 minutes in a buffer containing 0,5% Triton X-100, 150 mM NaCl, 10 mM KCl, 1,5 mM $MgCl_2$, 0,5 mM EDTA, 10 mM β-mercaptoethanol and 0,5 mM phenylmethylsulfonyl fluoride (PMSF) (Thermo Scientific, Waltham, MA, USA). Equal amount of proteins were resolved on SDS-PAGE (Thermo Scientific, Waltham, MA, USA). The proteins were transferred to nitrocellulose membranes (GE Healthcare, Little Chalfont, Buckinghamshire, UK). When specified, protein load was assessed by Ponceau S staining (Thermo Scientific, Waltham, MA, USA). Membranes were blocked with 5% milk diluted in PBS/0.1% Tween (Roche), incubated with primary antibody (see below Section 2.26) for one hour at room temperature and with HRP-conjugated secondary antibody for one hour at room temperature. Proteins were visualized with SuperSignal West pico or femto (Thermo Scientific, Waltham, MA, USA).

2.25 Immunofluorescence

HeLa cells constitutively expressing Flag-HA-tagged SAMHD1 were seeded on cover glasses and treated with or without 25 nM of camptothecin (CPT) (Sigma-Aldrich Chemical, St Louis, MO, USA) for 18 hrs. Cells were then fixed with 2% paraformaldehyde and permeabilized using 0.1% Triton X-100 (Thermo

Scientific, Waltham, MA, USA). Immunodetection of 53BP1 and HA-tagged SAMHD1 was performed overnight at 4°C using specific antibodies. Images were acquired using the Zeiss confocal microscope system, LSM780 (Oberkochen, Germany).

2.26 Antibodies

Anti-SAMHD1 antibodies used were purchased from Abcam, Cambridge, UK (ab96768 and ab67820) and Bethyl/Euromedex, Schiltigheim, France (A303-690A). Anti-phosphoSAMHD1Thr592 antibody has been described previously (Cribier et al. 2013). Anti-CyclinA2 antibody was purchased from Cell signalling (Leiden, The Netherlands), anti-DCAF1 antibody from Bethyl/Euromedex (Strasbourg, France), anti-HA from Covance/Eurogentec (Seraing, Belgium) and anti-53BP1 antibody was from Millipore (Billerica, MA, USA). Tubulin and actin antibodies were from Sigma-Aldrich Chemical (St Louis, MO, USA). Mouse (NA931) and rabbit (NA934) secondary antibodies were purchased from GE Healthcare (Little Chalfont, UK).

2.27 Drug treatment and cell proliferation

HeLa wild-type or expressing SAMHD1 were seeded in 96 wells plates at a density of 1000 cells per well. Etoposide, camptothecin, mitomycin C and methyl methanesulfonate (Sigma-Aldrich Chemical, St Louis, MO, USA) were applied to the cells six hours after seeding in a total volume of 100µl. Three days after treatment, the Cell Titer-Glo kit (Promega, Madison, WI, USA) was used to measure cell survival according to manufacturer's instructions. Results are shown as percentage of survival compared to mock-treated cells.

For the cell proliferation experiment, the cell count was performed at the indicated time using MACSQuant Analyzer (Miltenyi Biotec, Bergisch Gladbach, Germany).

2.28 Cell fractionation and glycerol gradient

HeLa stably expressing SAMHD1 were seeded in 15cm dishes and treated with 5μM etoposide (Sigma-Aldrich Chemical, St Louis, MO, USA) for 4 hours and 18 hours or mock-treated. Cells were then harvested, washed with PBS and lysed successively in five volumes of cytoskeleton (CSK) buffer (10mM PIPES pH7, 100mM NaCl, 300mM sucrose, 3mM MgCl2, 10mM beta glycerophosphate, 0,2mM sodium vanadate and 5mM Na Fluoride) containing 0,5% and 0,1% of Triton to purify nuclear extracts (Thermo Scientific, Waltham, MA, USA). For density gradient sedimentation, 400μg of nuclear extracts were loaded onto a glycerol gradient (15% to 35% in buffer containing 20mM Tris pH7,5, 0,15M KCl, 2,5mM $MgCl_2$, 0,05% NP40, 0,05% tween, 1mM DTT and protease inhibitors (Roche, Mannheim, Germany) and centrifuged for 4 hrs at 55K. 200μL fractions were collected from the top of the gradient.

2.29 Statistical analysis overview

Univariate and multivariate analyses were performed using IBM SPSS v22 (IBM Corporation, Armonk, NY, USA). Univariate analyses were performed using Fisher's Exact test for binary genetic variables and the Wilcoxon signed rank test for continuous variables. Cox regression with backward selection was selected for model development. Kaplan and Meier (KM) survival analysis with log-rank test was applied for survival data using GraphPad Prism V5

(GraphPad Software Inc, La Jolla, CA, USA). Significance was defined as <0.05. Correction for multiple testing used the False Discovery Rate (FDR) (Benjamini & Hochberg 1995). Exact Binomial Tests were used to test for associations between genetic lesions.

2.30 Statistical Validation of the Multivariate Model

To validate the genomic findings relating to clinical outcome described in Chapter 5, a statistical approach was employed using penalised Cox proportional hazards (PH). An optimal bootstrapping methodology to confirm the improved predictive value of the genetic model was also used.

In detail the well-established L_1-norm penalty function was adapted which gives rise to the *Least Absolute Shrinkage and Selection Operator* (LASSO) methodology (Tibshirani 1997). The question was to estimate the association of a number of identified genomic features with clinical phenotypes (shorter progression free survival in this case), the stability selection procedure by Meinshausen and Bulhmann (Meinshausen & Bühlmann 2010) was applied. This methodology consists of repeatedly subsampling the data with replacement, followed by feature selection through application of the LASSO on each sub-sampled dataset. By counting the number of times each genomic feature is included in the fitted penalized model, an estimate of its inclusion probability is calculated. Features more strongly associated with shorter PFS have higher inclusion probability and they are more likely to appear in the fitted model, i.e. they are more stable. Having identified a set of stable genomic features, the predictive value of a Cox proportional hazards model including these features only and using the time-dependent Brier score was assessed.

In addition, the prediction error using a cross-validation approach was estimated. All confirmatory statistical analyses were performed using the R statistical computing environment and the R packages *glmnet*, *c060* and *peperr* (Simon *et al*, 2011; Porzelius *et al*, 2010; https://www.r-project.org; Sill *et al*, 2014).

2.31 Hierarchical model

The hierarchical model described in Chapter 5 is based on genomic findings and clinical outcome. The variables in the hierarchy were ordered using stepwise backward conditional Cox regression modelling from SPSS version 22 (IBM Corporation, Armonk, NY, USA) with the variables remaining in the multivariable model with highest hazard ratios for early progression being selected first. For example as *TP53* remained in the final multivariate model with the highest hazard ratio, this group of patients were first selected from the 250 sample as depicted in the final KM curve; of the remaining patients the next genetic lesion remaining in the multivariate model with the highest hazard ratio was *SAMHD1*, then trisomy 12 non-trisomy 19 and so on. Each sample was counted one time only, with no overlap.

CHAPTER 3 : CANDIDATE GENE SELECTION USING WHOLE GENOME TECHNOLOGIES IN CLL: THE *SAMHD1* STORY

3.1 Introduction

Relapsed disease in CLL is a leading cause of death (Tam et al. 2014). As outlined in Chapter 1, chemotherapy resistance can be predicted in a low number of CLL cases harbouring *TP53* disruption, yet for the remaining patients no genomic cause has been identified (Gonzalez et al. 2011; Thorsten Zenz, Eichhorst, et al. 2010; Stephan Stilgenbauer et al. 2014). While the percentage mutation status at the IGHV locus separates two risk groups of patients, this measure alone does not currently help with treatment selection (Hamblin et al. 1999). As new therapies such as kinase inhibitors show activity and high progression free survival rates in CLL, safe and effective options are now available for patients predicted by genomics to be chemotherapy resistant (Byrd et al. 2014; O'Brien et al. 2015; Furman et al. 2014; Freise et al. 2016).

Whole genome sequencing of a patient's tumour and germline samples has the power to reveal a comprehensive catalogue of all somatic variants at a set moment in time (Chapman et al. 2011; Ding et al. 2012; Puente et al. 2011; Knight et al. 2012). However, cancers evolve over time and the introduction of treatment can alter the balance between different populations of cells comprised of different mutation profiles (Schuh et al. 2012; Ding et al. 2012; Greaves & Maley 2012; Gerlinger et al. 2012). The study of CLL tumour samples using whole genome sequencing and evaluation of intra-tumour

mutation dynamics has the potential to reveal founder, driver and passenger mutations.

Recurrent copy number aberrations are also important in CLL as outlined in Chapter 1 and the prognostic impact of certain chromosomal changes are well established and continue to be relevant with modernising treatments (Zenz, et al. 2010; Stephan Stilgenbauer et al. 2014; Döhner et al. 2000). Therefore, in addition to sequencing approaches, genome-wide arrays can be used to interrogate novel regions as well as define minimally deleted regions (MDRs) (Kujawski et al. 2008; Ouillette et al. 2011; Edelmann et al. 2013; Knight et al. 2012). Within MDRs one can identify candidate genes that are of potential relevance in CLL pathogenesis or disease progression (Jou et al. 2004).

3.2 Aims and Objectives

The overall aim of the work described in this Chapter was to use WGS to investigate whether one or more candidate genes in CLL might be identified as being involved in mechanisms of chemo-resistance that could not be explained by *TP53* disruption. The specific objectives were (i) to use WGS of CLL samples at different treatment timepoints to study mutation dynamics in the tumour over time. We hypothesised that the subclonal hierarchy of these mutations would facilitate the identification of potential founder or driver mutations in CLL (ii) to validate any newly identified driver genes by testing larger cohorts of CLL patients for recurrent mutations using next generation sequencing and genome-wide SNP arrays. Different cohorts of CLL patients both those enrolled in clinical trials and more heterogenous groups, at both pre-treatment and relapsed/refractory stages, were available for validation (iii) to carry out

functional validation by examining expression of downstream protein products of mutated genes and carrying out *in vitro* assays testing abnormal protein function.

3.3 Materials and Methods

Patient samples and all materials and methodology are outlined in Chapter 2.

3.4 Results

3.4.1 Whole genome sequencing reveals putative founder mutations

WGS was performed on longitudinal studies of samples from three CLL patients undergoing repeated cycles of treatment. The three patients selected did not respond well to chemotherapy despite having wild-type *TP53* status. Variants of likely pathogenic relevance were selected analysing exonic mutations only, removing synonymous mutations and including mutations predicted to be pathogenic by *in silico* algorithms as detailed in Chapter 2. Grouping mutations together based on their variant allele frequency (VAF) over time allowed different patterns of mutations to be investigated, such as persistent mutations despite treatment, or emerging mutations later in the disease course. **Figure 3.1** illustrates these profiles in a single patient, CLL077. Mutations present at all five time-points were the focus of further evaluation (**Figure 3.1**A).

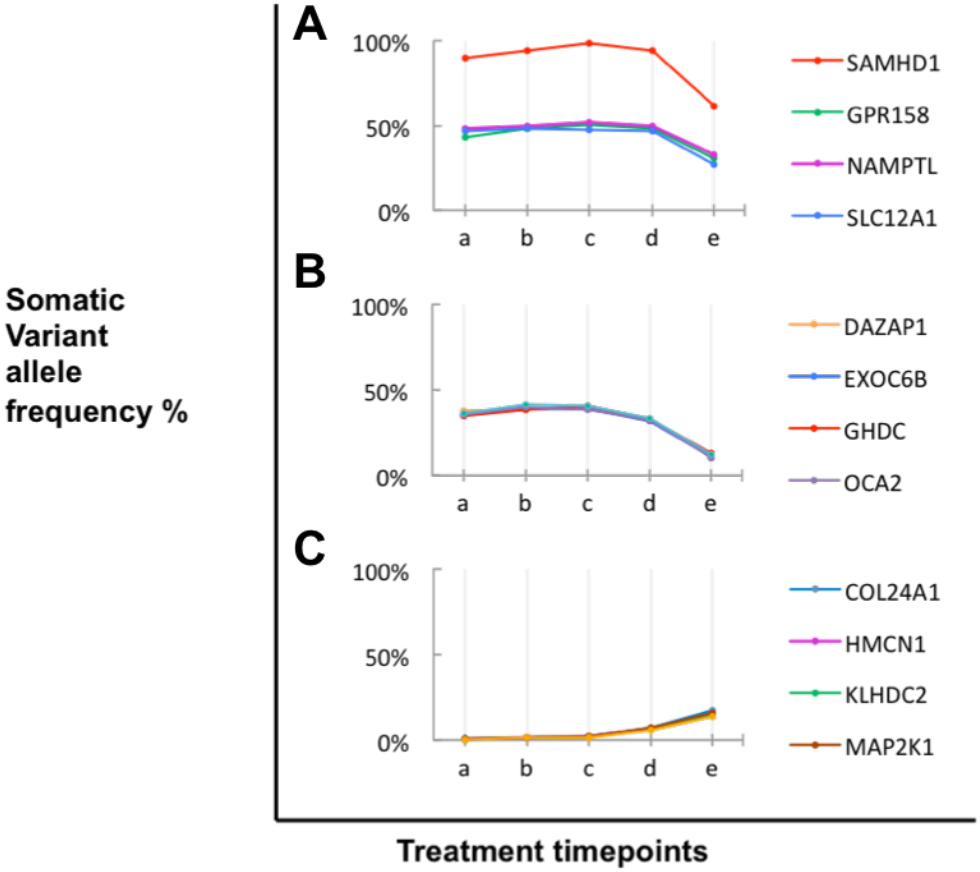


Figure 3.1 WGS of a CLL patient at 5 treatment time-points

Treatment time-points (a-e) shown on the x-axis.

Time-points were as follows: a Pre-Chlorambucil, b Pre-Fludarabine, Cyclophosphamide, c Post-Fludarabine, Cyclophosphamide, d Pre-Ofatumumab, e Relapse Post-Ofatumumab. All mutations identified were clustered according to their somatic variant allele frequency (%) shown on the y-axis. **A** VAF ≥50% for majority of time-points except at time-point E **B** VAF <50% and stable (not increasing or decreasing) for majority of time-points except at time-point E **C** Emerging mutations developing after initial time-point

Focus was placed on mutations causing amino acid changes and those present at all time-points with at least 50% VAF, as these can be present in all cells and potentially present at the beginning of leukaemia development. A purity correction was applied based on total lymphocyte count and may have interfered with VAF calculation most notably for time-point E as total

lymphocyte count was low. *SAMHD1* was identified as a particularly interesting gene as it carried a homozygous mutation, confirmed by Sanger sequencing.

3.4.2 Genome-wide arrays refine the minimal deleted region involving the SAMHD1 locus

The identification of minimally deleted or overlapping regions (MDR/MORs) using array-based technology to highlight potential tumour suppressor genes has proven fruitful for many cancer gene discovery projects (Mitelman et al. 1997; Bullrich et al. 1996; Parker et al. 2011; Jou et al. 2004). I reviewed genome-wide array data on 93 Oxford CLL patients for recurrent abnormalities. This revealed three patients with deletions and three patients with large areas of cnLOH involving the *SAMHD1* locus (**Figure 3.2**). The minimally overlapping region harboured solely the *SAMHD1* gene. This provided further evidence of a potential role for *SAMHD1* in CLL.

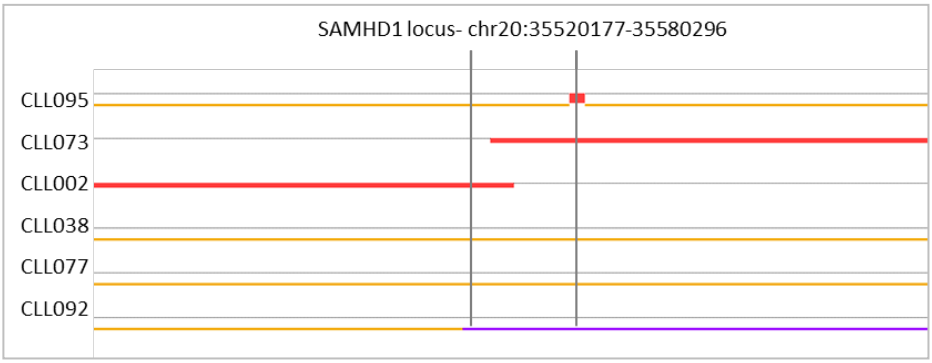


Figure 3.2 Genome-wide Array across the *SAMHD1* locus
This figure depicts whole genome array results spanning the *SAMHD1* locus for six patient's samples, CLL095 – CLL092. The x-axis represents the genomic sequence of the q arm of chromosome 20. Grey vertical lines represent the boundaries of *SAMHD1* gene; yellow lines represent copy neutral loss of heterozygosity; red lines depict heterozygous deletions; purple line represents a mosaic loss.

3.4.3 Literature Review of SAMHD1 locus guides further studies

SAMHD1 is a highly conserved protein and has been identified as an HIV-1 restriction factor operating in non-dividing cells including dendritic cells, macrophages and resting CD4+ T cells (Laguette et al. 2011; Baldauf et al. 2012; Hrecka et al. 2011). SAMHD1 is a deoxyribonucleotide triphosphate (dNTP) triphosphohydrolase (TPH) and a nuclease (Goldstone et al. 2011; Beloglazova et al. 2013), that regulates the intracellular dNTP pool (Powell et al. 2011; Lahouassa et al. 2012). In the cartoon below (**Figure 3.3**) a normal response to viral infection in a cell from exogenous DNA or RNA is depicted. Nucleic acid sensors trigger an immune response secondary to these exogenous nucleic acids. With a defective SAMHD1 protein leading to dNTP imbalance, activation of the immune system in the absence of an exogenous stimulus could possibly take place.

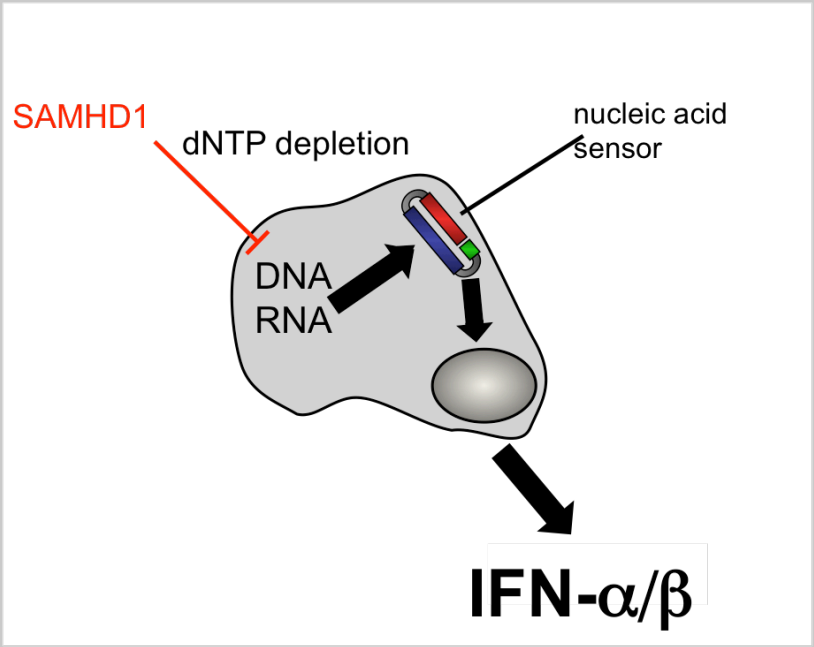


Figure 3.3 SAMHD1 mechanism of action

This figure depicts a cell with DNA or RNA in the cytoplasm stimulating a nucleic acid sensor. This in turn leads to up-regulation of interferon-regulatory genes in the nucleus leading to interferon release. The action of SAMHD1 as a regulator of dNTPs, leads to regulation of DNA/RNA synthesis, potentially leading to inhibition of this inflammatory process. Conversely mutations in *SAMHD1* as seen in Aicardi-Goutieres syndrome may lead to a dNTP imbalance and uncontrolled interferon production.

Adaptation of illustration provided by Prof. Jan Rehwinkel, Weatherall Institute of Molecular Medicine, University of Oxford, UK

Germline mutations in *SAMHD1* have been described in 17% of patients diagnosed with a congenital autoimmune disease called Aicardi-Goutieres syndrome (AGS) (Rice et al. 2009; Crow & Rehwinkel 2009). AGS is an autosomal recessive disorder that is characterized by an autoimmune encephalopathy and chronic activation of the immune system with raised inflammatory markers such as IL-6 in the peripheral blood and cerebrospinal fluid of patients (Rice et al. 2007). It occurs mainly in patients of

consanguineous background. Most patients with AGS succumb to the disease in early childhood (Rice et al. 2007). There is increasing evidence that chronic immune stimulation via the BCR has a role in the pathogenesis of CLL (Chiorazzi et al. 2005; Stevenson & Caligaris-Cappio 2004). Given the link with immune dysregulation and *SAMHD1* mutated AGS, further evaluation of *SAMHD1* as a candidate gene with relevance in CLL was undertaken.

3.4.4 Review of patients with congenital AGS and germline SAMHD1 mutations

Clinical notes were obtained and reviewed on the clinical course of 32 patients with congenital AGS and germline *SAMHD1* mutations. One of these patients was a male born to consanguineous parents who presented with features typical of AGS (Rice et al. 2009), but who had developed also CLL.

This patient had a homozygous germline canonical splice-acceptor site (1609-1G>C) mutation in intron 14 leading to an alternative splice variant (Rice et al. 2009). At the age of 24 years, he presented with cervical and axillary lymphadenopathy and a lymphocytosis. Flow cytometry demonstrated the typical hallmark of CD19+ CD5+ mature B lymphocytes (**Figure 3.4** A) confirming a diagnosis of CLL. High-resolution genome-wide array analysis of DNA from his CLL cells revealed a large region of homozygosity including the *SAMHD1* locus on chromosome 20 (**Figure 3.4** B). This was not associated with a copy number change, but was consistent with a germline inheritance pattern arising from consanguineous parents carrying shared haplotypes across this region. This was confirmed by Professor Crowe's laboratory (Division of Evolution and Genomic Sciences, Manchester University, UK) showing identical STR patterns across this region of chromosome 20 in both

maternal and paternal samples (personal communication). No other chromosomal aberrations associated with CLL were identified, nor were any new structural abnormalities detected.

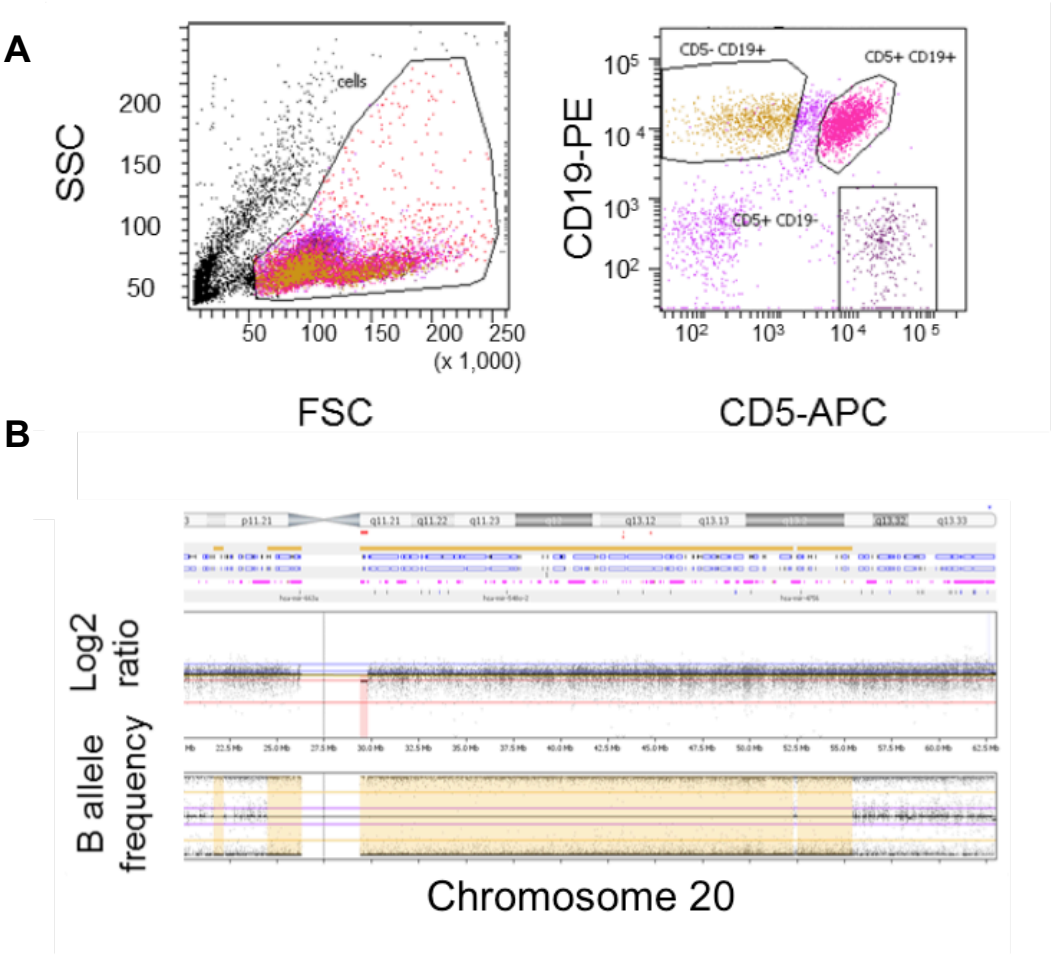


Figure 3.4 Analysis of CLL sample from Aicardi-Goutieres syndrome patient
A Flow cytometry plot from whole peripheral blood of AGS patient showing forward-sideward scatter (left) and CD19 PE/CD5 APC scatter gated on lymphoid cells (right).
B Whole genome array of chromosome 20 from leukaemia cells of AGS patient, visualised in Nexus v6.1 (BioDiscovery, Inc., El Segundo, CA, USA). Log2 ratio indicates copy number changes and B allele frequency plot indicates a region of homozygosity

3.4.5 Genomic evaluation of tumour and germline samples from the AGS CLL patient

To establish whether acquired mutations were present, we tested DNA from the AGS patient's CLL cells using a targeted approach to sequence genes reported as mutated recurrently in CLL (Chapter 2 Table 2.3). None of these genes were mutated in the patient's sample. Next, we performed whole exome sequencing (WES) to identify other acquired mutations that might contribute to CLL pathogenesis including recently described pathogenic mutations (Landau et al. 2013) (see methods). Again, acquired mutations seen previously in CLL or other haematological malignancies were not observed. Whilst these results do not rule out a contribution of other currently unknown genes in CLL pathogenesis, they support an involvement of *SAMHD1*.

3.4.6 SAMHD1 is recurrently mutated in CLL

To establish whether acquired *SAMHD1* mutations were recurrent in CLL patients, targeted sequencing of a single-institution cohort of 100 consecutive CLL patients (Knight et al. 2012) was carried out. Five patients of interest, carrying between them six mutations in *SAMHD1* (**Figure 3.5** A; **Table 3.1**; cohort 1), were identified. Germline analysis using sorted CD5+ CD19- cells confirmed that of the testable mutations all were somatic (**Figure 3.5** A). Furthermore, four of the five patients carried a mutation associated with either a mono-allelic deletion or cnLOH as revealed by genome-wide array data (**Figure 3.5** A). One additional patient had a small bi-allelic deletion of the 5'UTR of *SAMHD1* inclusive of exon 1 of *SAMHD1 s*een by SNP array (**Figure 3.5** B).

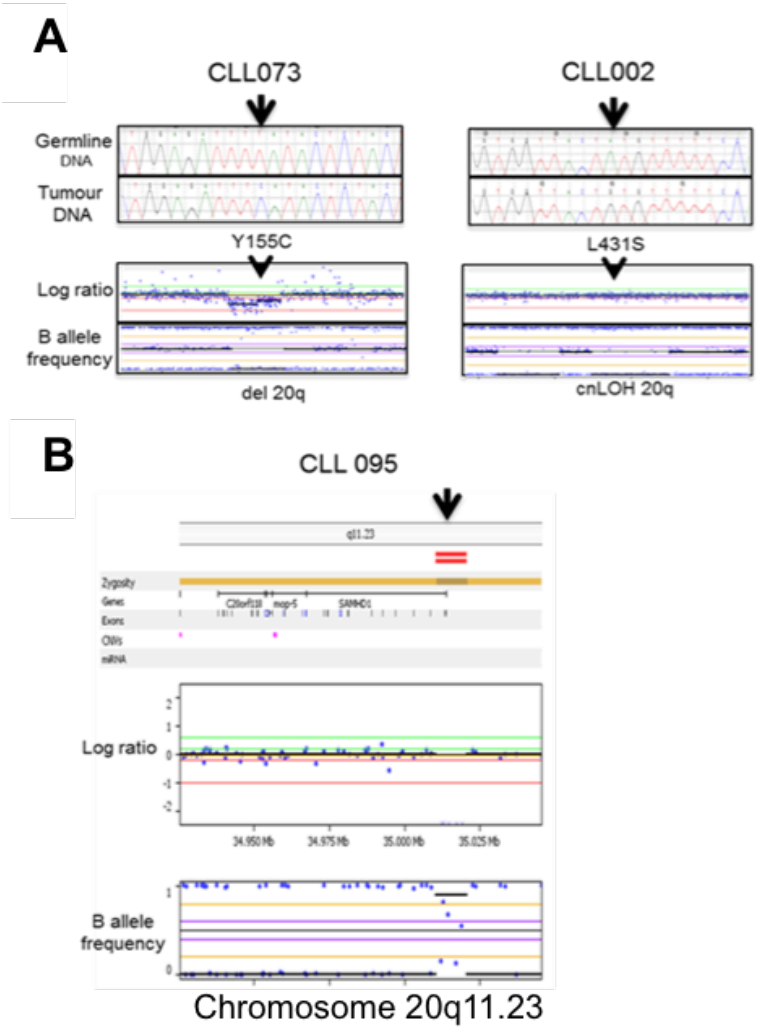


Figure 3.5 Example of Sanger sequencing and genome-wide array results from *SAMHD1* mutated samples

A Electropherogram of *SAMHD1* mutations using paired germ-line and tumour DNA. Representative results from 2 patients are depicted (upper panel). Genome-wide array data viewed in Nexus v6.1 showing a heterozygous deletion (left) and a region of cnLOH (right) involving *SAMHD1* (lower panel). Log2 ratio indicates copy number changes and B allele frequency plot indicates a region of homozygosity. B Genome-wide array viewed in Nexus v6.1 for patient CLL095 showing a small homozygous deletion (arrow) at the 5' end of the *SAMHD1* locus.

3.4.6.1 Establishing the frequency of SAMHD1 mutations across clinical trial studies

Analyses of a further 317 clinical trials samples from Phase II/III clinical trials were performed to establish the frequency of *SAMHD1* mutations across pre-treatment and relapsed/refractory phases of CLL. Of these, 263 patients were from four UK clinical trials; two pre-treatment trials (cohort 2, n=200) and two relapsed/refractory trials (cohort 3, n=63). A further 53 samples from patients with relapsed/refractory CLL treated in a Phase III French clinical trial were also tested (cohort 4). Further details of clinical trials are found in Chapter 2 Section 2.1. Mutation frequencies of 3% in the pre-treatment group and of 9% and 11% in the relapsed and refractory groups were found (Figure 3.6). This increase in mutation frequency at disease relapse could reflect a resistance mechanism to chemo-immunotherapy. Whether the relapsed cohort harboured these mutations at the first-line treatment stage is unknown as first-line treatment samples were not available for analysis.

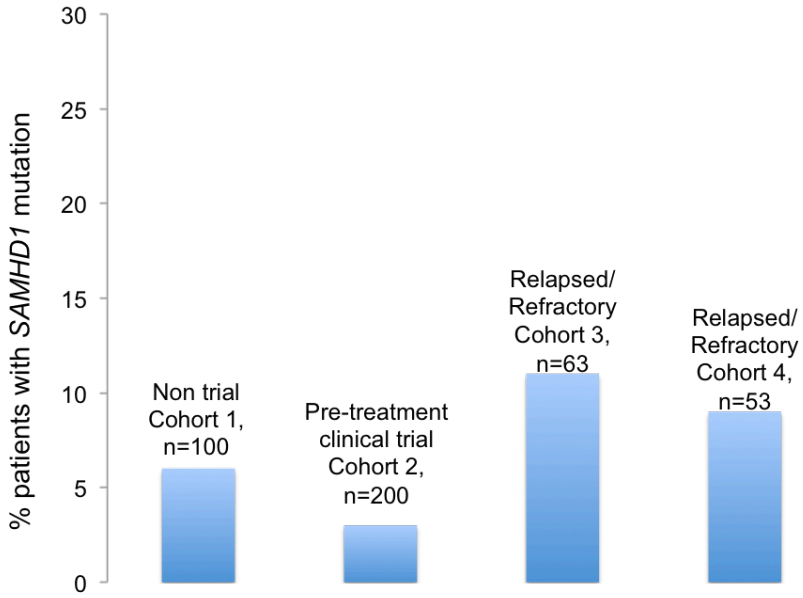


Figure 3.6 *SAMHD1* mutation frequency across cohort

3.4.6.2 Functional consequence and distribution of SAMHD1 mutations

The majority of *SAMHD1* mutations were single nucleotide nonsense or missense variants with predicted pathogenic consequences. Similar to the findings in AGS, mutations were distributed across all exons (Rice et al. 2009) (Figure 3.7) Four mutations (R145X, R145Q, I201N and c.1609-2A>T) were common to both CLL and AGS patients. Furthermore many mutations in both diseases, despite being outside the protein domains, occurred in the same exons (Figure 3.7). Mutations with potential pathogenicity were selected using the inclusion criteria of those conserved from at least mouse to man (Figure 3.8) and those predicted to be pathogenic using *in silico* SIFT and Polyphen algorithms (**Table 3.1**) (Shihab et al. 2013; Adzhubei et al. 2010).

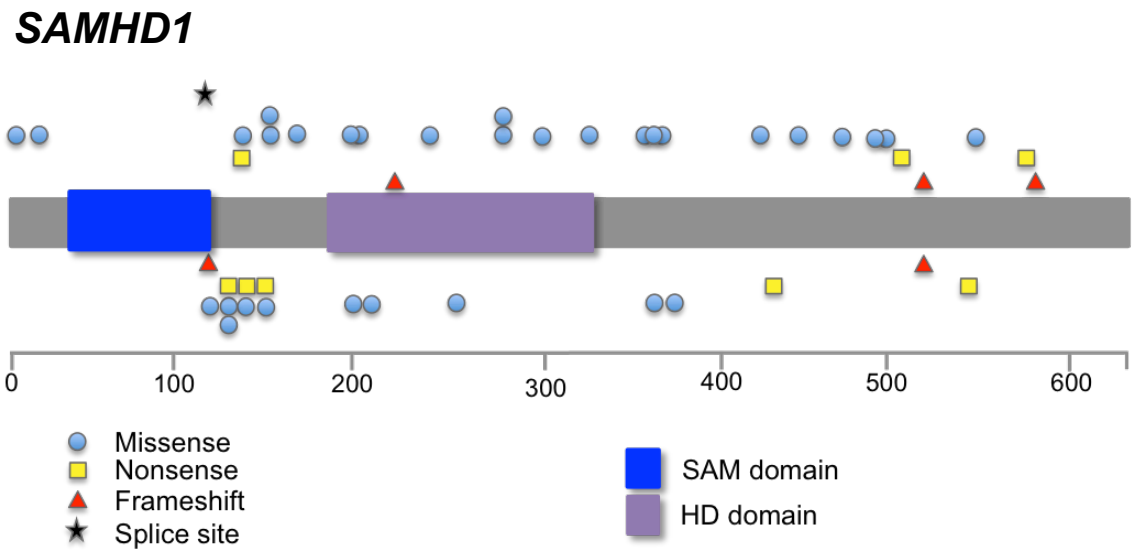


Figure 3.7 Schematic distribution of *SAMHD1* mutations

The horizontal scale represents amino acid number within the protein.Mutations in CLL found in this study are shown above the gene track and mutations found in AGS patients are shown below (Rice *et al.,* 2009). SAM= sterile alpha motif, HD= histidine–aspartic acid

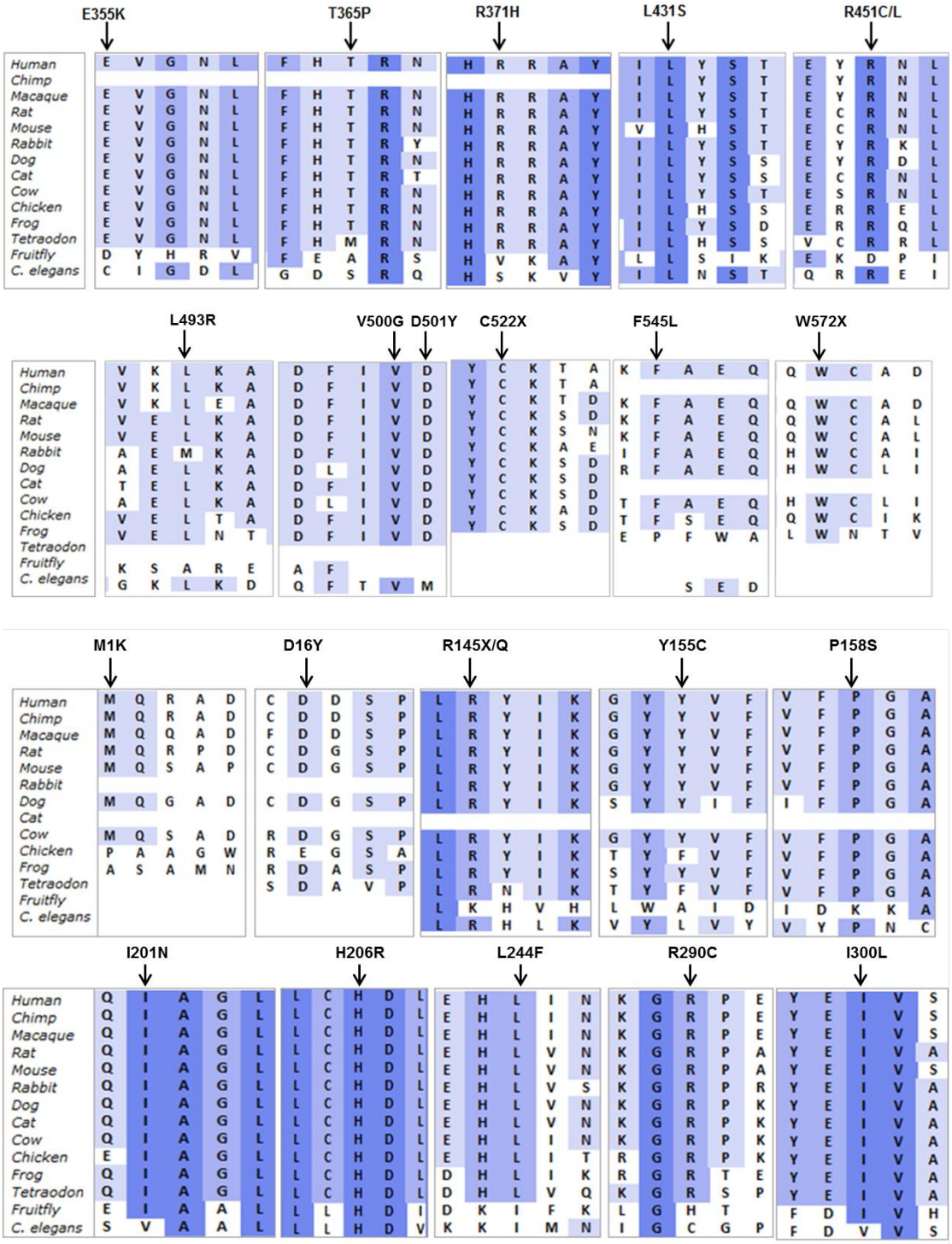


Figure 3.8 Conservation tracks for all *SAMHD1* mutations

Multiple Sequence Alignment of *SAMHD1* around the mutated residue (arrow) of each patient. Degree of conservation is reflected by background colour code (dark blue: highly conserved to white: not conserved) using Alamut database version 2.3 (Interactive Biosoftware, Rouen, France)

Two patients, outlined below, harboured mutations that were 'possibly damaging' rather than deleterious using *in silico* analyses, despite being highly conserved (Figure 3.8)**.** Mutation D501Y (CLL238), was confirmed to be somatic, and reached a SIFT score of 0.07. Similarly, mutation E355K (CLL2537) scored a high pathogenicity score with both *in silico* algorithms (SIFT score 0.2, Polyphen 2 score 0.871), in addition to harbouring a region of copy number loss over chromosome 20 encompassing the *SAMHD1* gene. Both samples were included as likely pathogenic variants based on the above findings.

Table 3.1 Pathogenicity scores of *SAMHD1* mutations

Patient no	Cohort	Sample ID	Amino acid change	SIFT prediction	SIFT score	Polyphen-2 Prediction	Polyphen Score
1	1	CLL077	F545L	T	0.54	D	0.978
2	1	CLL 092	M1K	U	-	P	0.915
	1	CLL 092	C522X	U	-	U	-
3	1	CLL073	Y155C	D	0.05	D	0.999
4	1	CLL 002	L431S	D	0	D	0.992
5	1	CLL 238	D501Y	D	0	P	0.655
6	2	CLL 242	R371H	T	0.09	D	0.999
7	2	CLL 069	R145Q	D	0.02	D	0.999
8	2	CLL 228	W572X	U	-	U	-
9	2	CLL354	L493R	D	0.01	B	0.446
10	2	CLL 007	H206R	D	0.01	D	0.999
	2	CLL 007	I201N	D	0	D	1
	2	CLL 007	A389T	D	0.06	D	0.989
11	3	CLL 2537	E355K	T	0.2	P	0.871
12	3	CLL 2353	R145X	U	-	U	-
13	3	CLL 2247	T365P	D	0.02	D	0.999
14	3	CLL 2198	I300L	D	0.03	D	0.995
	3	CLL 2198	P158S	D	0.02	D	1
	3	CLL 2198	D16Y	D	0	D	1
15	3	CLL 2219	L244F	D	0.03	D	0.993
	3	CLL 2219	R290C	T	0.12	D	1
16	3	CLL 2250	R451C	D	0.01	D	1
	3	CLL 2250	R451L	D	0.02	D	0.999
17	3	CLL 2462	V500G	D	0	D	0.999
18	4	CLL CJ	Y155C	D	0	D	0.999
19	4	CLL DJ	p.576_576del	U	-	U	-
20	4	CLL FB	R366C	D	0	D	1
	4	CLL FB	p.216_216del	U	-	U	-
21	4	CLL LJ	H210R	D	0.02	D	1
	4	CLL LJ	p.117_118del	U	-	U	-
22	4	CLL QP	c.1609-1G>C	U	-	U	-
	4	CLL QP	R290H	D	0.02	D	1

Amino acid changes according to transcript ID NM_015474. A low SIFT score and a high Polyphen score describe deleterious mutations. Scores were not calculated for indels. For SIFT: T= tolerated (benign), D=Deleterious, U=Unknown, For Polyphen: B=benign, P=Possibly damaging, D=Damaging

3.4.6.3 SAMHD1 mutations and associated genetic events

When assessed using genome-wide high-resolution arrays, *SAMHD1* mutations showed a clear association with chromosome 20 abnormalities involving the *SAMHD1* locus (seen in 12 out of 18 patients with data available, Figure 3.9). This finding is reminiscent of genetic lesions affecting tumour suppressor genes such as *TP53*. For example, *TP53* disruption can manifest with an isolated heterozygous mutation or deletion, or a deletion or cnLOH with a mutation leading to a double genetic hit within the *TP53* gene (Thorsten Zenz, Eichhorst, et al. 2010; Kruiswijk et al. 2015; Jasek et al. 2010). The IGHV mutation status was also associated statistically with *SAMHD1* mutated samples. Of the 19 samples with IGHV mutation status available, 16 had an IGHV-UM status (actual 84%, versus 60% expected for this cohort, Chi-squared test p=0.044). No other association was revealed between *SAMHD1* mutations and the genetic lesions tested. Importantly, as an association with *TP53* mutation and/or deletion was not found, this rules out a confounding reason for the increased number of samples with *SAMHD1* mutations in the relapsed and refractory cohorts.

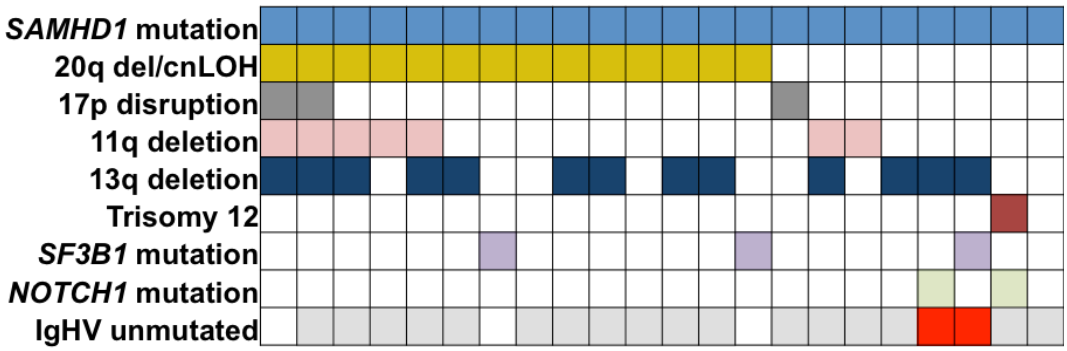


Figure 3.9 Mutation association chart of *SAMHD1* mutated samples

Each box represents one patient's sample. Light blue: *SAMHD1* mutation, yellow: 20q del/cnLOH, dark grey: 17p disruption, pink: 11q deletion, dark blue: 13q deletion, dark red: trisomy 12, purple: *SF3B1* mutation, green: *NOTCH1* mutation, light grey: IGHV-UM, bright red: no data available, Del: deletion, cnLOH: copy neutral loss of heterozygosity

3.4.6.4 Single and multiple SAMHD1 mutations with differing VAFs

Intriguingly, a number of patients had multiple *SAMHD1* mutations with differing VAFs ranging from 8-99% (**Figure 3.10**). Patients with single rather than multiple *SAMHD1* mutations were more likely to have a high VAF and these in turn were associated with cnLOH across the chromosome 20 locus. The presence of SAMHD1 mutations with low VAFs provided evidence against these being founder mutations.

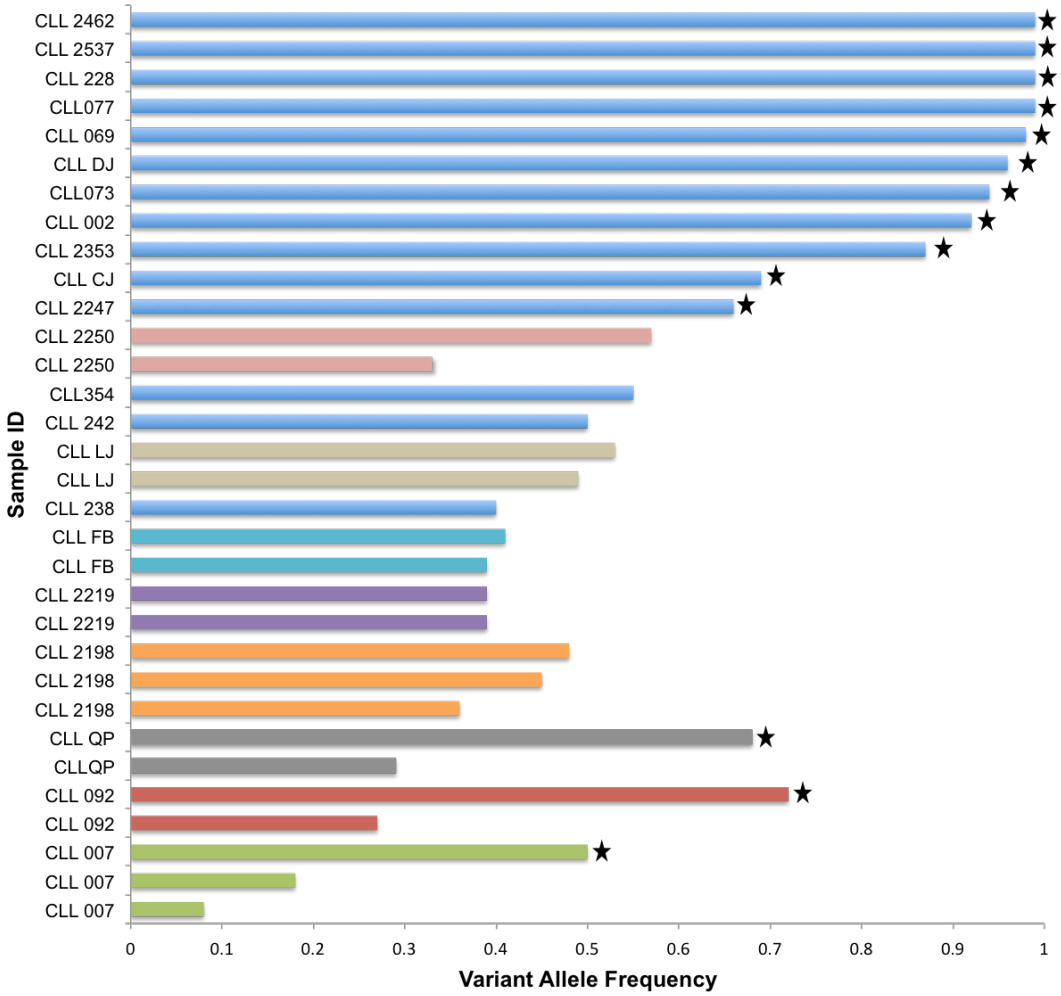


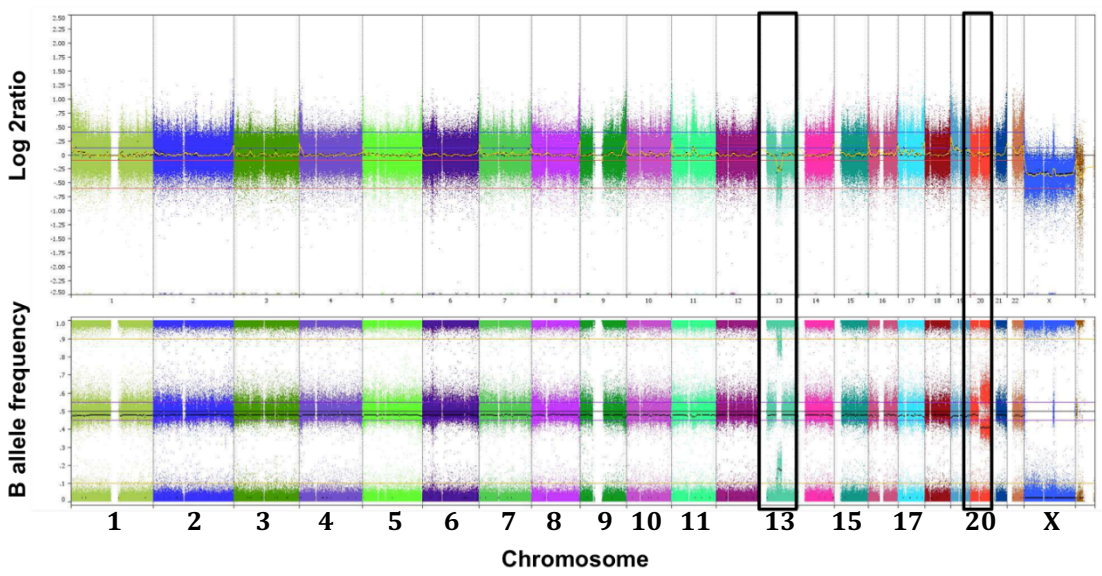
Figure 3.10 Variant allele frequency of *SAMHD1* mutations

Variant allele frequency of *SAMHD1* mutations (x-axis) detected by NGS, on all individual *SAMHD1* mutations (y-axis). Multiple mutations in the same patient are depicted by the same colour, except blue shading which reveals samples with only one *SAMHD1* mutation, seen mostly at the top end of the graph with a higher VAF. The asterisk highlights samples with either cnLOH or deletion on chromosome 20 encompassing the *SAMHD1* locus, seen by SNP array.

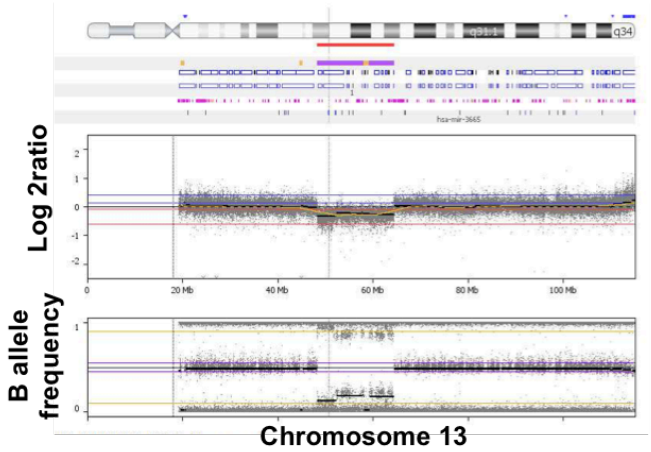
3.4.6.5 Multiple SAMHD1 mutations reveals branching clonal evolution

One patient, CLL007, had six different amino acid changing *SAMHD1* mutations with a predicted pathogenic effect. This 59 year old patient did not have any poor risk cytogenetic features, but had an IGHV-UM status. FISH studies revealed an isolated 13q deletion, and targeted deep sequencing revealed wild type *TP53*, *NOTCH1*, *ATM*, *BIRC3*, and *SF3B1*. He required treatment three years after his CLL diagnosis. His disease progressed 15 months after FCMR CIT. On review of his whole genome SNP array, in addition to a 13q deletion, cnLOH of chromosome 20q was found (**Error! Reference source not found.**A). Two levels of 13q deletion were apparent reflecting intra-tumour heterogeneity. One 13q deletion was larger in a sub-population of approximately 80% of cells (12.9 Mb), compared to the 13q deletion encompassing the MDR at MiR-15a/16-1 that affected >90% of all CLL cells and was 3.2 Mb in size (**Error! Reference source not found.** B).

A



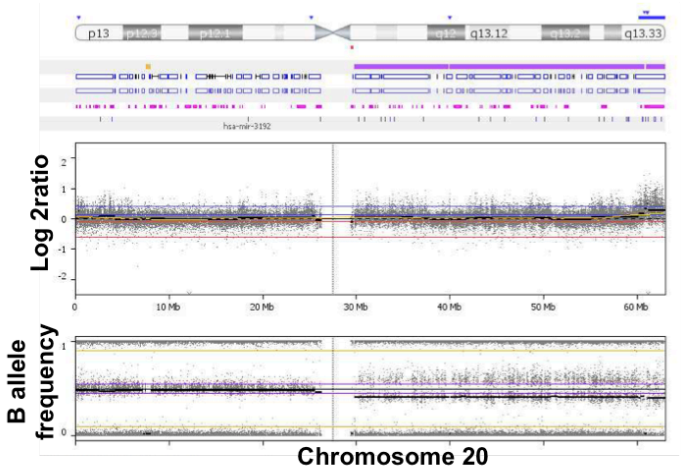
B



80% of cells: larger 13q deletion

90-95% of cells: smaller 13q deletion with MDR

C



20% of cells: cnLOH of 20q

(Example of Genome-wide array results for sample CLL007

A Whole genome view visualised in Nexus v6.1. Log2 ratio indicates copy number changes and B allele frequency plot indicates loss of heterozygosity.

B Chromosome 13 view from genome-wide array. Two levels of deletions are detectable, from the log 2 ratio plot and B allele frequency plot, reflecting different cell populations with different size of the deletion C Chromosome 20 view from genome-wide array. A subset of cells (20%) data points harbour copy neutral loss of heterozygosity encompassing *SAMHD1.*

A large region of copy neutral loss of heterozygosity was seen along chromosome 20q of 33.2 Mb length that included the *SAMHD1* locus (**Error! Reference source not found.** C). This affected approximately 20% of cells tested. The six *SAMHD1* mutations are outlined in **Table 3.2**. One of these, I201N, was confirmed to be a germline mutation and therefore present in every CLL cell. Whether this potentially damaging mutation was involved in CLL pathogenesis is unclear.

Table 3.2 CLL007 *SAMHD1* mutation details

Amino Acid	SIFT	PolyPhen2	Depth	VAF	Germline
p.I201N	Deleterious	Damaging	259	0.5	Yes
p.H206R	Deleterious	Damaging	261	0.17	No
p.A389T	Deleterious	Damaging	801	0.08	No
p.R226H	Deleterious	Damaging	2649	0.04	Unknown
p.D120A	Deleterious	Damaging	3006	0.02	Unknown
p.L369S	Deleterious	Possibly Damaging	4426	0.01	Unknown

By integrating array data with the SNV data, we can postulate a number of phylogenetic trees based on inference of percentage of cells affected by a copy number change or SNV (**Figure 3.11**). If a mutation is present in a subset of cells and increases over time, it may be involved in driving this clone to expand. For this patient, the analysis of a snapshot from one sample taken prior to first treatment revealed that clonal diversity was already present, with at least two clones defined by copy number changes and *SAMHD1* mutations. The clone with a mutated *SAMHD1* was either within the larger 13q deletion that has expanded, (**Figure 3.11** B), or was in opposition to this clone (**Figure 3.11** C). Interestingly, this patient had three additional low-level *SAMHD1* mutations

found on analysis when the filter for VAF of ≥5% was removed. This potentially reveals clonal competition and convergence on a mutated *SAMHD1* as a possible mechanism for clonal evolution in the absence of any known CLL driver mutations other than the 13q deletion. A number of additional trees can be derived, which are not illustrated here, and single-cell sequencing or growth rate studies could confirm the true clonal architecture.

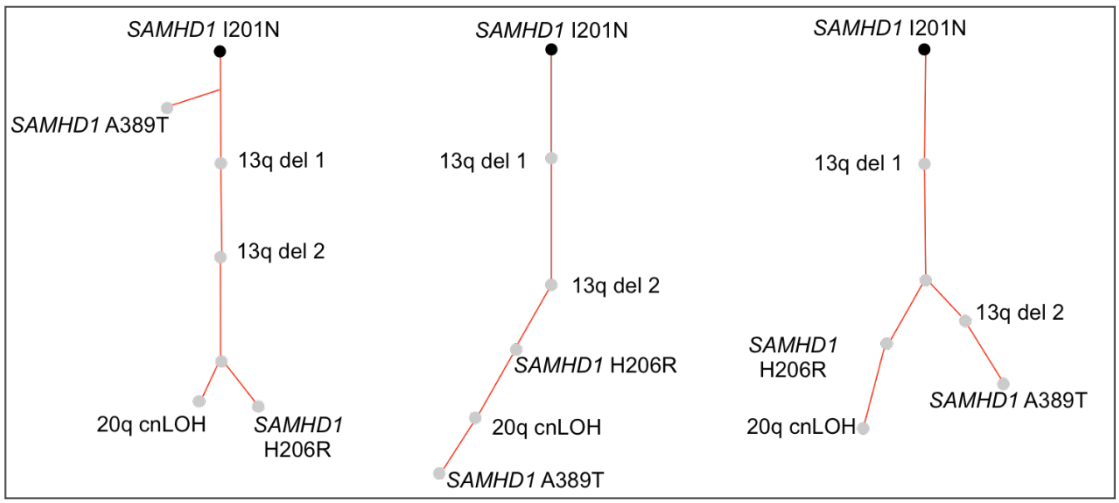


Figure 3.11 Phylogenetic trees for CLL007

Potential linear and branching phylogenetic trees for patient CLL007. Three mutations with VAF <10% (R226H, D120A, L369S) could be present in any of the clones illustrated above. Black dot denotes germline mutation present in all cells, grey dot denotes acquired mutation in CLL cells.

This example highlighted a very aggressive CLL disease course in a patient with an isolated 13q deletion.

3.4.7 Expression of SAMHD1 in mutated CLL samples

To interrogate the functional impact of the *SAMHD1* variants, SAMHD1 expression was analysed by both mRNA and SAMHD1 protein expression where available. *SAMHD1* is expressed in normal tissues including normal B

cells as shown in **Figure 3.12** and **Figure 3.13** A. *SAMHD1* mRNA expression from highly purified samples on 11 *SAMHD1* mutated CLL patients was compared with that of normal B cells. Using RT-qPCR statistically significant lower expression levels of mRNA in *SAMHD1* mutated CLL cells were found using the Mann-Whitney test for unpaired samples (p=0.0159) (**Figure 3.13** A). Where possible Sanger sequencing was performed on cDNA extracted from *SAMHD1* mutated CLL cells using primers designed to cross intron-exon boundaries (Chapter 2 Section 2.12). This confirmed that mutant *SAMHD1* transcripts in CLL cells were expressed for the three patients tested (**Figure 3.13** B). Western blots performed on protein extracts from CLL samples with both mutated and wt *SAMHD1* revealed reduced protein expression in the *SAMHD1* mutated CLL samples (**Figure 3.13** C).

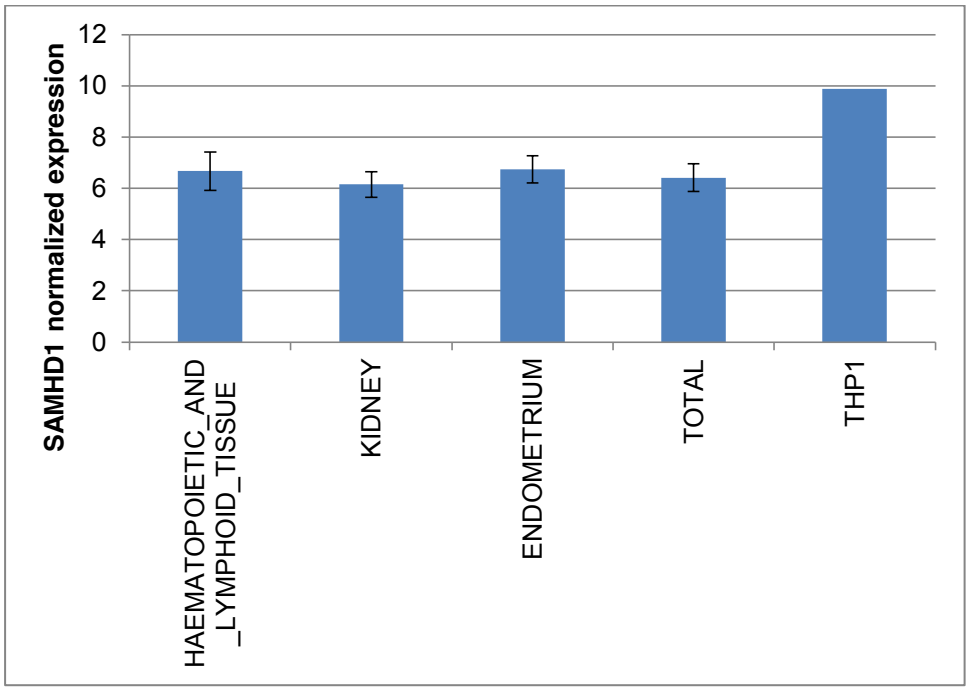
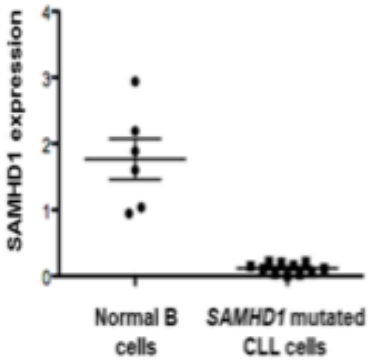


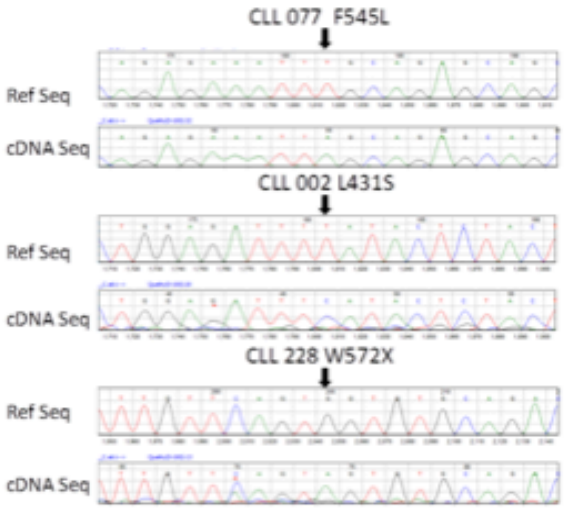
Figure 3.12 *SAMHD1* expression in various tissues

SAMHD1 expression in different normal tissue types including epithelial cells. Gene expression data for *SAMHD1* taken from from CCLE_Expression_Entrez_2012-10-18.res:Gene-centric RMA-normalized mRNA expression data.

A



B



C

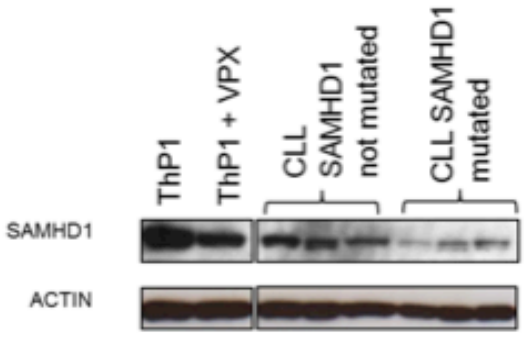


Figure 3.13 SAMHD1 expression in Primary CLL cells

A cDNA pools from B cells (CD19+, CD5-) sorted from peripheral blood of healthy donors and SAMHD1 mutated CLL cells (CD19+, CD5+), were subjected to quantitative PCR analysis using ABL as a control gene.

B Sanger Sequencing of cDNA from *SAMHD1* mutated cells

Electropherogram of cDNA obtained from SAMHD1 mutated CLL cells from 3 patients using primers that cover intron/exon boundaries of SAMHD1. Arrows indicate mutation compared to reference sequence. RNA expression data (relative values calibrated against normal B cell control expression of 1) for SAMHD1 mutated samples.

C SAMHD1 protein expression profile

Western Blot of SAMHD1 in *SAMHD1* mutated, wt *SAMHD1* CLL and ThP1= monocytic cell line as control.

3.4.8 SAMHD1 regulation of cell proliferation and survival

To begin to understand the function of SAMHD1 further *in vitro* experiments were undertaken by Dr. Tania Louis and Prof. Monsef Benkirane at the Institute of Human Genetics, Montpellier, France. There is evidence that an imbalance in the dNTP pool produces a hypermutator phenotype, and that increased levels of dNTPs interfere with proper DNA replication and repair (Chabosseau et al. 2011; Bester et al. 2011). Therefore, the potential role of SAMHD1 in cellular proliferation and DNA damage response was evaluated. When HeLa cells were transduced with lentiviral vector expressing, under the control of an inducible promoter (**Figure 3.14**), either wild type (wt) SAMHD1, a dNTP TPH-inactive HD mutant, or a C-terminal deletion mutant which has been characterized as a constitutively active form (Laguette et al. 2011; Goldstone et al. 2011; Cribier et al. 2013). Expression of wt SAMHD1, but not HD mutant, resulted in slow proliferation compared to cells transduced with empty vector (**Figure 3.15** A). The anti-proliferative activity of SAMHD1 was amplified when the constitutively active mutant (SAMHD1ΔCter) was expressed, suggesting a key role for SAMHD1 enzymatic activity. When HeLa cells expressing wt *SAMHD1* were treated with three DNA damaging agents, camptothecin, mitomycin C and etoposide (known to induce DNA double strand breaks (DSBs)), and with a methylating agent, methyl methanesulfonate, increased cell death after treatment with DNA DSB agents was noted in the Hela cells expressing wt SAMHD1 (**Figure 3.15** B). However, no effect on cell survival was seen with the methylating agent, suggesting that SAMHD1 is involved specifically in DNA DSB repair.

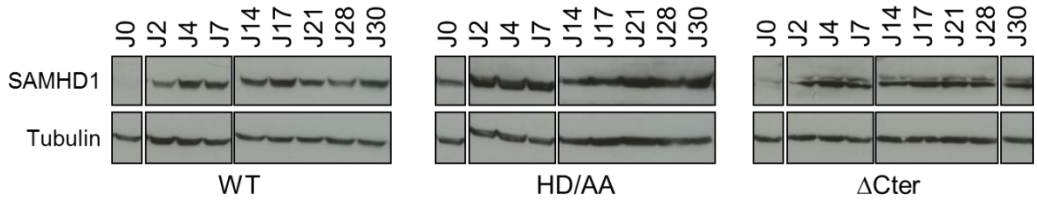


Figure 3.14 SAMHD1 expression in transduced HeLa Cells

HeLa cells were transduced with lentiviral vector expressing, under the control of an inducible promoter, either wild-type (wt) SAMHD1, a dNTP TPH-inactive HD mutant (HD/AA) or a C-terminal deletion mutant (ΔCter) (Laguette et al. 2011; Goldstone et al. 2011; Cribier et al. 2013). J = number of days. Tubulin is the control.

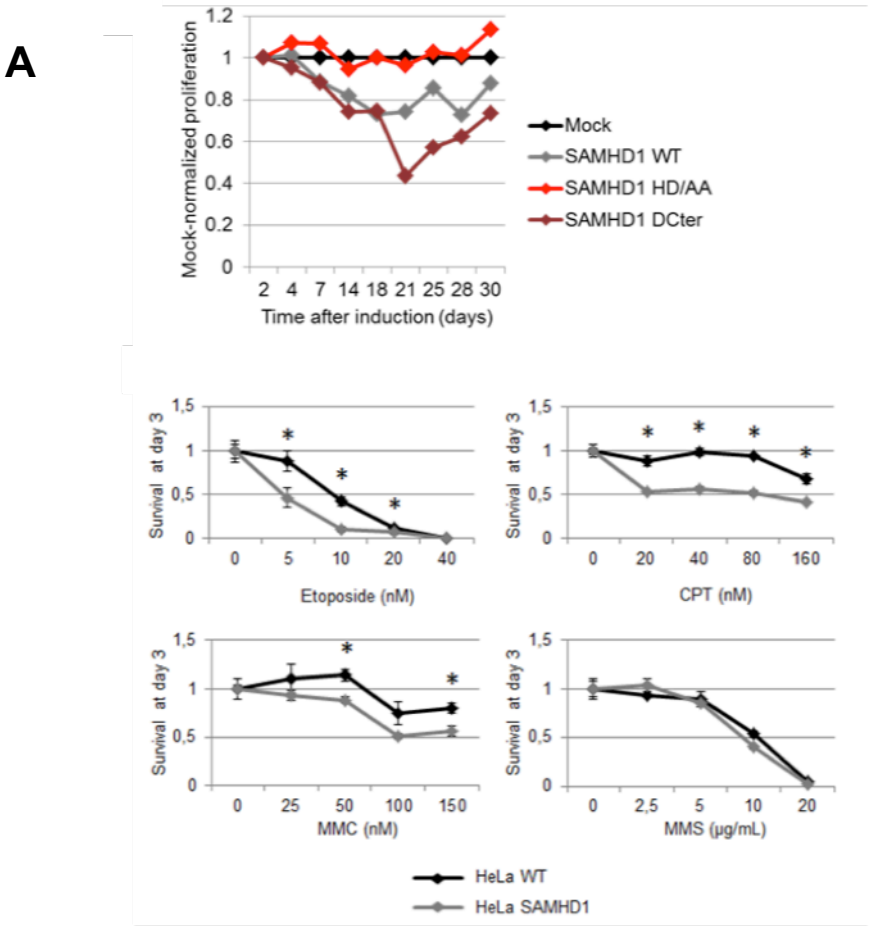


Figure 3.15 SAMHD1 effects on cell proliferation and cell viability in response to DNA damaging agents

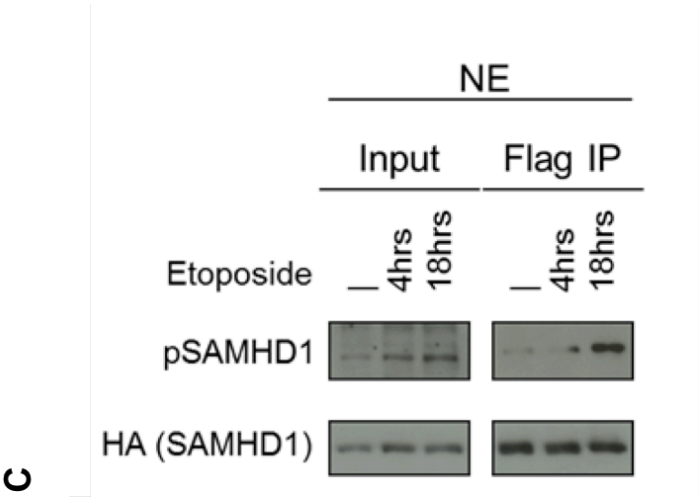
A HeLa cells stably expressing an inducible empty vector (Mock) or a vector encoding for SAMHD1 wild type (wt), SAMHD1 mutated in the HD domain (HD/AA) or truncated at amino acid 575 (ΔCter) were induced with doxycycline. Results are expressed as proliferation ratio of SAMHD1 transduced cells divided by mock-transduced cell. A representative experiment out of 3 is shown.

B HeLa and HeLa cells stably expressing SAMHD1 were treated with increasing concentrations of DNA damaging agents and the percentage of living cells was measured 3 days post treatment. Each graph shows the mean of three independent experiments. The asterisk indicates a t-test p-value of <0.05.

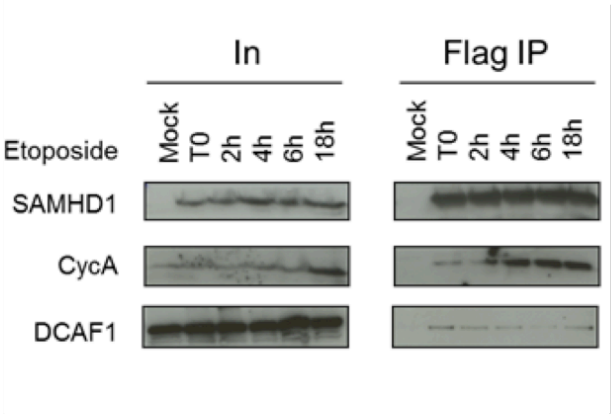
CPT= camptothecin, MMC= mitomycin C, MMS= methyl methanesulfonate

3.4.9 SAMHD1 involvement in DNA repair

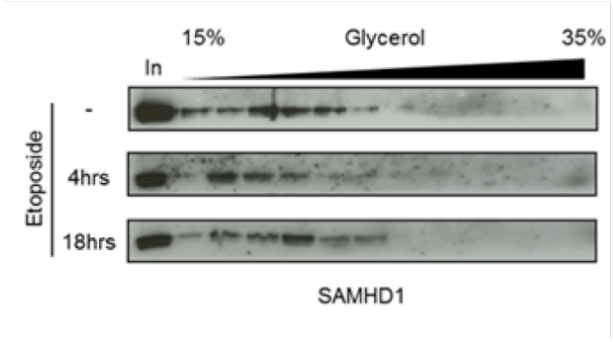
To explore specific protein interactions of SAMHD1 in the setting of cellular DNA damage, the distribution of SAMHD1 was analysed in nuclear extracts (NE) prepared from mock- or etoposide-treated HeLa SAMHD1 cells as described in Chapter 2 Section 2.28. The results showed that SAMHD1 resides in a high-molecular-weight complex (HMW) in untreated cells and converts to low-molecular-weight complex (LMW) after 4hrs treatment with etoposide (**Figure 3.16** A). Conversion from HMW to LMW complex is transient, since SAMHD1 was recovered in HMW complex 18hrs post etoposide treatment. Treatment of cells with etoposide increased the pool of SAMHD1 interacting with CyclinA (**Figure 3.16** B) and induced SAMHD1 phosphorylation at the Thr592 (**Figure 3.16** C). The phosphorylation of SAMHD1 at Threonine 592 (Thr592) by CyclinA2/CDK1 is known to be a key regulatory mechanism of its antiviral activity (Cribier et al. 2013; White et al. 2013).


NE
Input
Flag IP
Etoposide
_
4hrs
18hrs
_
4hrs
18hrs
pSAMHD1
HA (SAMHD1)

C


In
Flag IP
Etoposide
Mock
T0
2h
4h
6h
18h
Mock
T0
2h
4h
6h
18h
SAMHD1
CycA
DCAF1

B


15%
Glycerol
35%
In
Etoposide
-
4hrs
18hrs
SAMHD1

A

Figure 3.16 SAMHD1 expresion from nuclear extracts

A. Western Blot.showing SAMHD1 expression from nuclear extracts of Hela cells stably expressing SAMHD1 treated with 5nM of etoposide along a glycerol gradient at time points of 0hrs, 4hrs and 18hrs. B. Western Blot showing expression of SAMHD1 that was FLAG-HA-tagged from HeLa cells that were mock-treated or treated with 5nM of etoposide.. Flag-IPs were assessed at the time intervals shown and SAMHD1 interactions with CyA and DCAF1 were assessed. FLAG=polypeptide protein tag HA=Human influenza haemagglutinin CyA=Cyclin A, DCAF1 Drosophila Melanogaster chromatin Assembly Factor 1 C. Western blot analysis showing SAMHD1 and SAMHD1 phosphorylated on Threonine 592 (pSAMHD1) levels from nuclear extracts (NE) of HeLa cells stably expressing Flag-HA-tagged SAMHD1 after treatment with 5nM etoposide.

3.4.10 SAMHD1 is recruited to site of DNA damage

To further explore the involvement of SAMHD1 in DNA damage response, whether SAMHD1 was recruited to the site of DNA damage was investigated. HeLa cells expressing SAMHD1 were mock-treated or treated with camptothecin (CPT) and stained for both p53 binding protein-1 (53BP1) and SAMHD1. 53BP1 foci were seen in untreated cells presumably reflecting double stranded breaks that resulted from replication stress. An increase in number and size of 53BP1 foci was observed upon camptothecin treatment. SAMHD1 co-localized with 53BP1 foci in both untreated and in CPT treated cells (Figure 3.17).

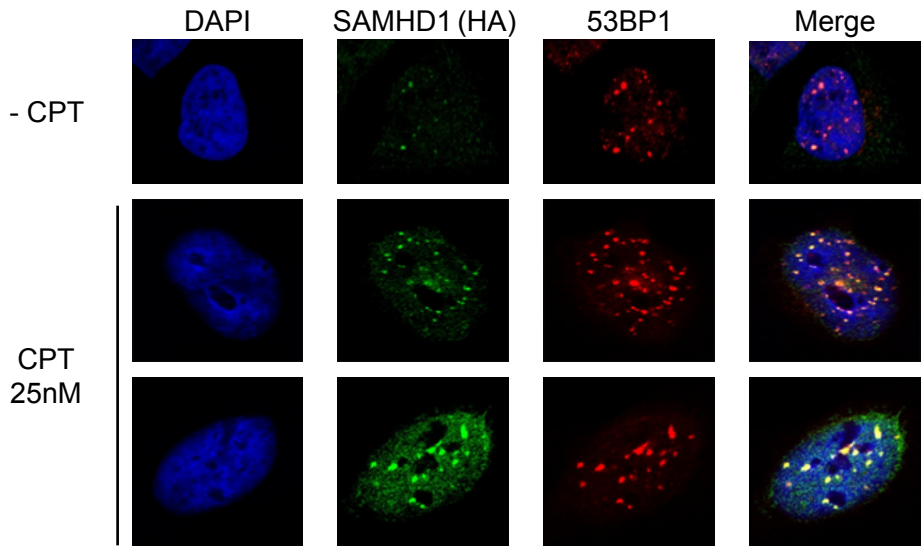


Figure 3.17 SAMHD1 and DNA damage response

SAMHD1 co-localizes with 53BP1 at the site of DSBs. HeLa cells expressing SAMHD1-HA were mock-treated or treated with 25nM CPT. SAMHD1 and 53BP1 intranuclear localizations were determined by immunofluorescence using specific antibodies and confocal microscopy with DAPI staining the live nucleus. DAPI=4',6-diamidino-2-phenylindole

3.5 Discussion

The focus of this chapter was to investigate CLL patients for previously unidentified recurrent mutations to try to explain mechanisms of chemo-resistance in patients without *TP53* disruption. Over the course of this chapter the finding of recurrent mutations in *SAMHD1* in CLL patients, the molecular features associated with these mutations, and the early exploration of the role of SAMHD1 in CLL pathogenesis have been outlined.

The application of NGS technologies to cancer samples has facilitated the identification of candidate genes involved in cancer pathogenesis or disease progression (Pleasance et al. 2010; Campbell et al. 2008; Puente et al. 2011; Chapman et al. 2011; Roberts et al. 2012). Sequencing samples at different time-points reveals the clonal dynamics of the disease and permits features of candidate genes, such as founder or driver mutations, to be studied. Using WGS to identify potential candidate genes of interest from sequential studies, *SAMHD1* was found to be mutated in all CLL cells in one patient at all time-points after extrapolating from variant allele frequencies.

The molecular characterization of an AGS patient in whom a homozygous germline mutation in *SAMHD1* had been identified previously highlighted this gene as potentially involved in leukaemogenesis (Rice et al. 2009). Remarkably, this patient developed CLL at only 24 years of age; only 0.2% of patients diagnosed with CLL are aged 20-34 years and the median age at presentation is 72 years (Howlader 2016). There have been no other reports of CLL in AGS patients, although only a small number of patients beyond the age

of 20 years are known. Using targeted NGS, WES and genome-wide SNP array to test sorted CLL B-cells from this patient, the *SAMHD1* mutation was confirmed and a large region of homozygosity was revealed. This was consistent with inheritance from consanguineous parents carrying shared haplotypes. Importantly, no further acquired CLL-related sequence changes (Zenz et al. 2010; Skowronska et al. 2012; Puente et al. 2011; Quesada et al. 2013; L. Wang et al. 2011a) nor structural variants (Döhner et al. 2000; Edelmann et al. 2013) were present in the leukaemia cells. This data supported the notion that a homozygous *SAMHD1* germline mutation was involved in CLL development in this particular patient. Alternatively, other currently unknown acquired or germline driver mutations could be involved in the pathogenesis of his disease.

Both *TP53* and *ATM* genes have a well-documented association with CLL and high-risk aggressive disease (Thorsten Zenz, Eichhorst, et al. 2010; Stephan Stilgenbauer et al. 2014; Austen et al. 2005; Skowronska et al. 2012). Both are associated with hereditary cancer syndromes and behave as tumour suppressor genes. In the case of *TP53* the Li-Fraumeni syndrome amongst others, and Ataxia Telangiectasia in the case of *ATM* (Srivastava et al. 1990; Malkin et al. 1990; Mavrou et al. 2008). These inherited cancer predisposition syndromes have mutations in common with CLL. In addition, the finding of germline pathogenic *ATM* mutations in CLL patients was found more frequently in a cohort of patients with 11q deletions, and was associated with rapid disease progression (Skowronska, Austen, et al. 2012). The authors hypothesise that germline *ATM* mutations are involved in disease progression rather than CLL

initiation. Germline samples were not available for all patients therefore the possibility that germline pathogenic variants are included in this study was not completely excluded. For the case study CLL007 while this patient had two acquired SAMHD1 variants of >0.05 VAF a germline variant (I120N) was also present at VAF of 0.50 and this mutation is found in AGS also. Larger studies with germline and tumour samples will be required to draw definitive conclusions on individual mutations in *SAMHD1* akin to those findings of Skowronska et al. (2012) in the setting of CLL.

The findings of a germline *SAMHD1* mutation in this AGS patient presenting with CLL at such a young age, together with the longitudinal CLL WGS results in which an acquired homozygous *SAMHD1* mutation was identified, led us to perform further molecular studies of SAMHD1. The mutation frequency was tested using well-characterized samples from patients referred to a tertiary referral centre, and those recruited into national clinical trials in both the UK and France. This sequencing data revealed recurrent *SAMHD1* mutations in all cohorts, increasing in frequency at relapse. NGS analyses have been used to document recurrent mutations in CLL and pertinently, *SAMHD1* mutations have subsequently been noted at a statistically significant frequency (Puente et al. 2011; L. Wang et al. 2011a; Quesada et al. 2011; Landau et al. 2013). Enrichment of *SAMHD1* mutations in the relapsed/refractory setting have recently been noted in two additional cohorts, leading to the belief that *SAMHD1* is involved in driving relapse (Amin et al. 2016; Burger et al. 2016).

While this is examined in more detail in the following chapter, the enrichment of *SAMHD1* mutated samples at relapse may be partly explained by the significant association with IGHV-UM status. This is analogous to the finding of high risk *TP53* and *ATM* disruption in the IGHV-UM group (Gladstone et al. 2012). This finding suggests that *SAMHD1* mutations track with a specific IGHV mutation status and potentially these *SAMHD1* mutated samples account for a subset of IGHV-UM. In the case study of patient CLL007 with six *SAMHD1* mutations, the very short PFS after CIT, points towards a far more aggressive case than is generally seen with IGHV-UM status. The median PFS of patients with IGHV-UM treated with FCR CIT is 4.2 years, far longer than the PFS for this patient (Thompson et al. 2016). It is possible that the aggressive clone harbouring at least one *SAMHD1* mutation explains this patient's chemo-refractory status with a PFS of only 15 months after gold standard CIT. Analysis of a follow up sample at disease progression could confirm this theory, but was unavailable. Despite the finding that certain patients with IGHV-M have curable disease, there remain subsets of patients in typically good or poor risk groups that do not conform to these predictions (Fischer et al. 2016; Thompson et al. 2016). Furthermore, having molecularly defined risk groups in addition to IGHV mutation status may aid the development and selection of therapies as well as aiding risk stratification.

Genome-wide array data highlighted that *SAMHD1* mutations associate with cnLOH or monoallelic loss across chromosome 20q encompassing *SAMHD1,* a feature of tumour suppressor genes. This leads to homozygous mutations in *SAMHD1*, and may involve all CLL cells. However, when studying the VAF of

all samples in this study, it was clear that wild-type *SAMHD1* was also present for many of these samples. Expression data reveals decreased *SAMHD1* mRNA and protein levels in these mutated CLL samples. It is accepted that previously held definitions of oncogenes and tumour suppressor genes were oversimplified and that hemizygous mutations can be involved in tumorigenesis (Hohenstein 2004). It appears that tumour formation and/or progression can arise from various degrees of mutation within the same gene (Venkatachalam *et al.*, 1998). In this *SAMHD1* mutated CLL population various levels of *SAMHD1* dosage can be present. Rentoft *et al., 2016* showed that hemizygous *SAMHD1* mutations lead to sufficient alterations in dNTP levels that allowed an increased rate of mutagenesis in mouse embryos (Rentoft et al. 2016). Regarding cases with multiple *SAMHD1* mutations, one can hypothesise that at the time of sampling numerous clones were present (eg CLL007, CLL092). Alternatively, compound heterozygous mutations could account for multiple mutations of similar VAF seen in this study (eg LEMJac, CLL2198). In the absence of single cell analysis solely an inference of the subclonal architecture can be discussed. Based on the varied VAFs seen in this study it is clear that *SAMHD1* is not a founder mutation for all patients. This pattern points towards the acquisition of mutations as disease progresses, or at least of the outgrowth of clones to become more dominant. This is consistent with the increased frequency of *SAMHD1* mutations seen in relapsed/refractory groups compared to pre-treatment suggesting an innate high-risk disease phenotype, perhaps due to treatment resistant features that are supported by *SAMHD1* disruption.

In addition, the functional consequence and allelic burden of the *SAMHD1* mutation may influence the clonal dynamics. It has been shown for *TP53*, an extensively studied tumour suppressor gene, that varying effects on subsequent disease relapse are dependent on both of these factors (Kruiswijk et al. 2015). For CLL patients with *ATM* disruption, mutation burden correlated with varying levels of *ATM* function, severity of downstream apoptosis and p53 inactivation (Stankovic et al. 2002). The co-occurrence of cnLOH over chromosome 20q encompassing *SAMHD1* may be necessary for an impact to be made in the setting of a mutated *SAMHD1* gene and subsequent outgrowth of this clone.

To strengthen the genomic studies, SAMHD1 mRNA and protein expression in CLL and solid tumours was examined. SAMHD1 is widely expressed across most normal tissues (Descours et al. 2012; Coon et al. 2012). While *SAMHD1* mutations were expressed and clearly associated with markedly reduced but not absent mRNA and protein, a mutated gene may indeed lead to a dysfunctional SAMHD1 protein, as has been seen in studies using samples from AGS patients (Goncalves et al. 2012).

The series of experiments presented here provides evidence for the potential involvement of SAMHD1 in the regulation of cell proliferation and response to DNA damage inducing agents. The role of SAMHD1 in the regulation of cell proliferation has been confirmed and expanded by SAMHD1 RNA interference studies (Elisa Franzolin et al. 2013). SAMHD1 engages in specific interactions upon DNA damage and its expression regulates cell survival in response to

DNA damage inducing agents. Importantly, SAMHD1 co-localizes with 53BP1 in DNA repair foci after induction of DSB in these co-localising experiments. Together, these data suggest that SAMHD1 plays a role in the cellular response to DNA damage.

From a clinical perspective, it is possible that *SAMHD1* mutations in CLL lead to chemotherapy resistance. This is supported by the fact that loss of SAMHD1 expression results in enhanced resistance to DSB-inducing agent and also the finding that *SAMHD1* mutations are enriched in the relapsed/refractory cohort without concurrent increase of potential poor risk confounding factors such as del11q and del17p. Alternatively, *SAMHD1* mutations might be associated with disease aggressiveness, defined by a short time to first treatment or transformation to Richter's Syndrome. Analyses of larger patient cohorts are required to address this question fully.

In conclusion in this chapter we have shown that recurrent mutations in *SAMHD1* are newly identified molecular events in CLL and the protein SAMHD1 engages in specific interactions in response to DNA damage. More work is needed to fully elucidate the role of SAMHD1 both in regulation of DNA damage response and in leukaemogenesis and/or disease progression.

CHAPTER 4 : DEFINING THE GENOMIC LANDSCAPE OF PRE-TREATMENT CLL

4.1 Introduction

Acquired copy number alterations (CNAs) and gene mutations are pivotally important in CLL pathogenesis as outlined in Chapter 1. CNAs detectable by FISH occur in 80% of patients, defining a hierarchy of prognostic groups in chemotherapy treated patients (Döhner et al. 2000). The immunoglobulin heavy chain variable (IGHV) region mutation status alone defines two broad risk groups following distinct clinical courses (Hamblin et al. 1999; Damle et al. 1999). Long-term follow-up data suggests that a significant number of patients with hypermutated IGHV remain in remission for many years following FCR treatment (Fischer et al. 2016; Tam et al. 2014).

Disruption of *TP53 (*either through deletions of 17p13.1 or mutations of *TP53)* and deletions of 11q22.3 are well described in CLL and impact on survival across clinical trials (Zenz et al. 2010; Zenz et al. 2008; Jitka Malcikova et al. 2009; Gonzalez et al. 2011; Michael Hallek et al. 2008). However, the addition of Rituximab to the chemotherapy-backbone led to significant improvements in the prognosis of patients with deletions of 11q22.3. Currently, the presence of *TP53* disruption alone indicates the need for alternative non-CIT treatments, in particular targeted agents such as Ibrutinib and Idelalisib and Venetoclax (Hallek *et al*, 2008; Oscier *et al*, 2012; Byrd *et al*, 2014; Furman *et al*, 2014)

More recently, whole genome and exome sequencing (WGS, WES) studies have revealed recurrent mutations involving novel pathways in CLL pathogenesis e.g. mRNA processing (*SF3B1*), NOTCH signalling (*NOTCH1*) and the innate immune response (*SAMHD1*) (Puente et al. 2011; Quesada et al. 2012; Landau et al. 2013; Schuh, Becq, Humphray, et al. 2012). To examine the prognostic impact of these previously reported CLL driver mutations (Rossi, Rasi, et al. 2012; Rossi et al. 2011; Weissmann et al. 2013; Del Giudice et al. 2012; Quesada et al. 2012; Rossi et al. 2012; Jeromin et al. 2014; Baliakas et al. 2014; Oscier et al. 2013), studies using primarily Sanger sequencing have led to the proposal of a refined integrated prognostic model at diagnosis (Rossi et al. 2013).

Compared to Sanger sequencing, MPS allows inference of the subclonal architecture of individual CLL samples (Schuh et al. 2012; Landau et al. 2013; Mardis et al. 2009; Shah et al. 2012). The presence of a subclonal driver mutation, in particular in *TP53*, may have a negative impact on outcome similar to clonal mutations (Rossi et al. 2014). Besides, genome-wide array analyses have confirmed the presence of recurrent CNAs in CLL that evolve over time in addition to chromothripsis, constituting markers of genomic complexity (Gunnarsson et al. 2010; Kujawski et al. 2008; Knight et al. 2012; Ouillette et al. 2010) that are associated independently with aggressive disease and reduced overall survival.

4.2 Aims and Objectives

The identification of numerous genetic features including novel single gene mutations, CNAs, sub-clonal intra-tumour heterogeneity and genomic complexity in clinically heterogeneous CLL cohorts at various disease time-points warrants an integrated analysis in a prospective clinical trial. The overall aim of the work presented in this Chapter was to define the mutational landscape of pre-treatment CLL samples by investigating the frequency and interactions of key CLL driver mutations in CLL and evaluating the intra-tumour heterogeneity of samples by interrogating them for both genomic complexity and subclonal and clonal populations of mutations. This work was performed using a combination of massively parallel sequencing (MPS) and high-resolution genome-wide array tests of 250 CLL DNA samples from frontline FCR-based UK clinical trial patients.

4.3 Materials and Methods

Materials and methods relevant to this Chapter are given in Chapter 2

4.4 Results

4.4.1 Patient characteristics

The study of 250 frontline treated patients in a clinical trial setting with well-annotated clinical outcomes was undertaken. Patient characteristics were representative of the entire cohort without any statistically significant difference in characteristics between the 250 cases in this project and all patients enrolled in the clinical trials (**Table 4.1**). The median age of CLL onset was 61 and predominately male (73%). Of the patients with IGHV mutation status available (n=212), 57% had an unmutated IGHV status. Genetic features previously associated with high-risk disease, detected by FISH, were present in 13/250 (5%) for 17p deletions and 46/250 (18%) for 11q deletions.

Table 4.1 Patient characteristics of all ARCTIC and ADMIRE patients

	Analysed (n=250)	Not analysed (n=165)

Characteristic	Median		Median		P value
Age	61		62.7		0.092
WBC	113.6		109.9		0.149
Lymphocyte Count	105.2		99.23		0.096
Haemoglobin	11.94		12.18		0.512
Platlelets	149.2		153.5		0.667
	No.	**%**	**No.**	**%**	
Sex					0.58
M	182	72.8	116	70.3	
F	68	27.2	49	29.7	
Binet stage					0.436
A progressive	34	13.6	27	16	
B, C	216	86.4	138	84	
Trial, Treatment arm					
ADMIRE					0.636
FCR	57	22.8	50	30	
FCMR	61	24.4	47	28	
ARCTIC					0.765
FCR	65	26	35	21	
FCM-miniR	67	26.8	33	20	
IGHV mutation status					0.659
Mutated	91	36.4	51	31	
Umutated	121	48.4	75	45	
N/A	38	15.2	39	24	
FISH Del17p					0.229
Del17p	13	5.2	4	2	
Non del17p	219	87.6	134	81	
Missing	18	7.2	27	16	
FISH Del11q					0.178
Del11q	46	18.4	20	12	
Non Del11q	188	75.2	121	73	
Missing	17	6.8	24	15	
CD38					0.643
Positive	100	40	60	36	
Negative	139	55.6	92	56	
Missing	11	4.4	13	7.8	

4.4.2 Validation of Next Generation Sequencing method

4.4.2.1 TSCA quality metrics and depth of coverage

Nine sequencing runs of multiplexed TSCA samples with an average of 28 samples per run resulted in a mean of 15.9 million reads per run. Of these reads 96.6% passed the Illumina chastity filter (PF) meaning they presented an acceptable signal to noise ratio. This is defined as the ratio between the brightest base intensity divided by the sum of the brightest and second brightest base intensities. Reads do not pass the quality filter if there are two or more base calls with less than the Threshold of Chastity of 0.6 in the first 25 cycles. The yield was 4.4 Gigabases and 92.3% of reads were above a Phred score of Q30 across nine MiSeq runs. This score of Q30 is equivalent to the probability of an incorrect base called 1 out of every 1,000 base calls. Across all samples the average depth of coverage per amplicon was 2032X. Median coverage of the genes was 2012X. The mean coverage across all genes was greater than 100X and 500X, for 98.8% and 98% patients respectively (Figure 4.1). *NOTCH1* had a less uniform coverage across targets. However the performance of the amplicon covering the 7541_7542delCT mutation hotspot was consistent with the other target genes; where 98.8% of samples yielded at least 100X depth and 93.6% at least 500X depth.

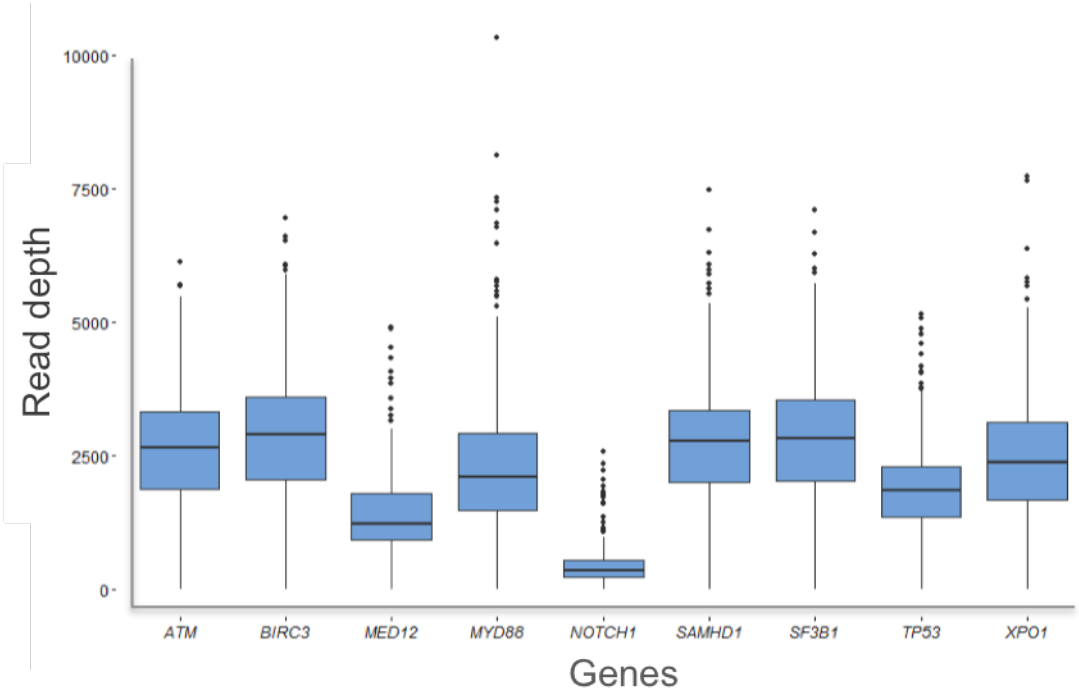


Figure 4.1 Distribution of read depth per gene.

The blue boxes represent the lower and upper quartile ranges and the black lines represent the median read depth for each given gene.

4.4.2.2 Filtering process

To validate the filtering pipeline 16 germline samples with matched tumour samples were sequenced using the TSCA Miseq panel. The analysis generated 494 variants with 18 true somatic variants confirmed from 19 previously identified somatic variants, revealing a sensitivity of 95%. 466 true germline variants were confirmed out of 475 variants previously annotated as non-somatic mutations, revealing a specificity of 98%.

4.4.2.3 Sensitivity analysis of TSCA panel

A mutation detection sensitivity analysis was performed to assess the capacity of the MiSeq instrument and bioinformatics pipeline to detect mutations with low variant allele frequencies (VAF) of <20%. The test DNA harboured a heterozygous somatic K700E mutation in *SF3B1* (VAF=0.49) in addition to seven other mutations or SNPs as illustrated in Figure 4.2.

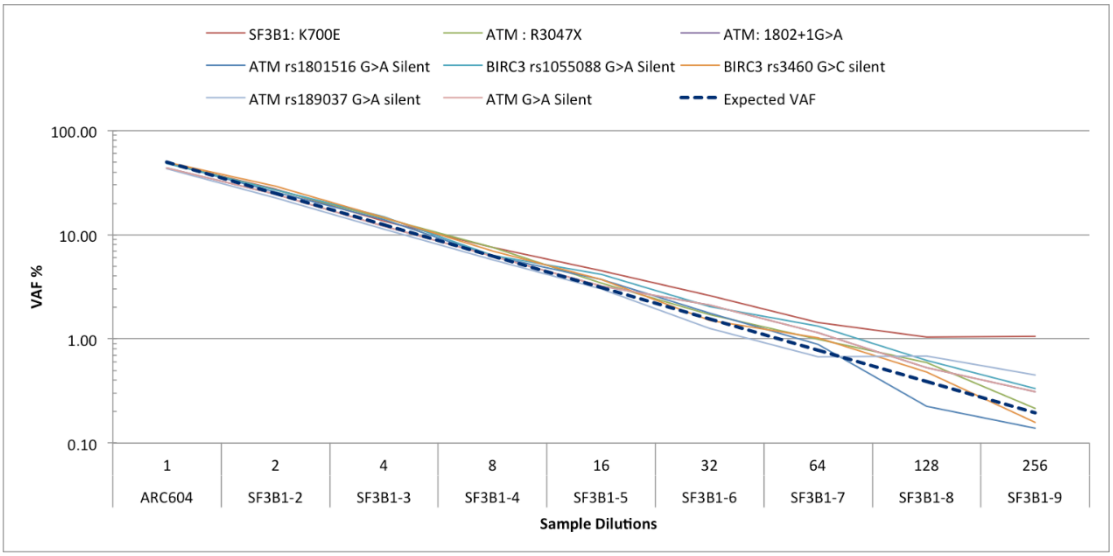


Figure 4.2 Expected versus Observed VAFs for sample ARC604
One sample containing the *SF3B1* mutation hotspot K700E and seven other variants/SNPs underwent serial dilutions, x-axis, and was sequenced on the MiSeq in duplicate. The blue dotted line reveals the calculated expected VAF for each dilution.

The correlation between expected VAF (based on sample dilutions) and the observed VAF obtained from visual inspection of IGV (number of altered reads / total number of reads) for the results of the *SF3B1* VAFs at various dilutions was high ($r=0.9983$ $p<2.2e-16$) using a Pearson's correlation (**Table 4.2**, Figure 4.3).

Table 4.2 Dilution experiment results for the *SF3B1* mutation

Dilution*	Total number of reads	Number of ref reads	Number of alt reads	Measured VAF (%)	Expected VAF (%)
1:1	3247	1654	1592	49.03	50
1:2	2831	2052	773	27.3	24.51
1:4	2785	2403	378	13.57	12.26
1:8	1933	1787	146	7.55	6.13
1:16	2610	2489	117	4.48	3.06
1:32	3622	3527	94	2.6	1.53
1:64	5604	5514	80	1.43	0.77
1:128	3460	3419	36	1.04	0.38
1:256	2720	2691	29	1.07	0.19

*The Dilution column shows the dilution factor. Ref = reference reads, alt = alternative reads, VAF = variant allele frequency. The measured VAF is calculated from reference and alternative reads. The expected VAF is calculated from the first measured VAF of 49.03% using the dilution factor.

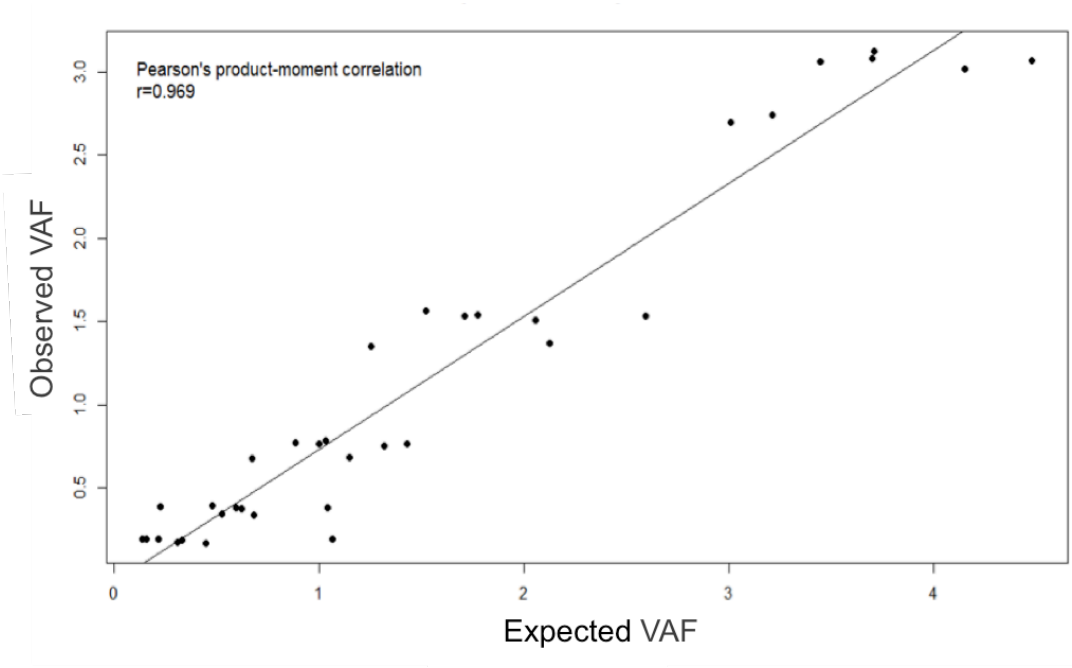


Figure 4.3 Graph showing correlation between expected versus observed VAFs for *eight* mutations in sample ARC604

Results from dilution experiment illustrating only the VAFs of ≤5%. Pearson's correlation between expected VAFs (x axis) and measured VAFs (y axis) for a spiked mutation sample containing five mutations in total in *ATM, BIRC3* and *SF3B1* and a sample Wild type for each mutation.

4.4.2.4 Sanger sequencing validation of mutations

For all CLL samples with novel mutations where matched germline samples were available, validation by Sanger sequencing was performed. Within the detection limits of Sanger sequencing, all mutations found using the MiSeq instrument were also seen by Sanger sequencing to a level of 10% VAF (**Table 4.3**).

Table 4.3 Variants detected by NGS and validated by Sanger sequencing

Sample ID	Gene	Chr	Start (hg19)	VAF	Depth (X)	Sanger Confirmed
ADM007	*SAMHD1*	chr20	35539726	0.081	885	Y
ADM007	*SAMHD1*	chr20	35559171	0.176	267	Y
ADM007	*SAMHD1*	chr20	35559186	0.492	267	Y
ADM029	*BIRC3*	chr11	102206910	0.354	2012	Y
ADM041	*ATM*	chr11	108218083	0.534	4006	Y
ADM045	*MED12*	chrX	70339253	0.492	1268	Y
ADM068	*BIRC3*	chr11	102207657	0.253	715	Y
ADM069	*SAMHD1*	chr20	35563507	0.958	3162	Y
ADM106	*ATM*	chr11	108164095	0.302	1838	Y
ADM210	*SAMHD1*	chr20	35526266	0.935	1697	NA
ADM242	*TP53*	chr17	7577027	0.68	432	Y
ADM287	*ATM*	chr11	108181006	0.189	4361	Y
ADM295	*ATM*	chr11	108201102	0.928	1765	Y
ADM324	*ATM*	chr11	108186605	0.493	796	Y
ADM340	*ATM*	chr11	108121753	0.511	1767	Y
ADM366	*ATM*	chr11	108201039	0.59	1642	Y
ADM373	*BIRC3*	chr11	102206865	0.09	2981	Y
ADM373	*BIRC3*	chr11	102207736	0.476	679	Y
ADM427	*ATM*	chr11	108175504	0.716	443	Y
ADM446	*MED12*	chrX	70339254	0.215	5000	Y
ARC026	*ATM*	chr11	108160384	0.098	2766	N†
ARC026	*SF3B1*	chr2	198266834	0.489	3013	Y
ARC042	*SF3B1*	chr2	198267361	0.429	641	Y
ARC043	*BIRC3*	chr11	102201946	0.122	1744	Y
ARC043	*BIRC3*	chr11	102207657	0.067	1216	Y
ARC104	*ATM*	chr11	108139215	0.216	706	Y
ARC138	*ATM*	chr11	108121676	0.892	557	Y
ARC143	*SF3B1*	chr2	198266833	0.062	1763	N†
ARC143	*SF3B1*	chr2	198267481	0.202	1193	Y
ARC143	*ATM*	chr11	108224493	0.076	1838	Y‡
ARC201	*ATM*	chr11	108151824	0.763	457	Y
ARC292	*ATM*	chr11	108158429	0.138	551	Y
ARC554	*ATM*	chr11	108196251	0.814	553	Y
ARC554	*ATM*	chr11	108225582	0.078	204	Y‡
ARC566	*ATM*	chr11	108121753	0.49	923	Y
ARC604	*ATM*	chr11	108122759	0.461	2500	Y
ARC604	*SF3B1*	chr2	198266834	0.462	2596	Y
ARC621	*SAMHD1*	chr20	35545421	0.448	2010	Y
ARC627	*BIRC3*	chr11	102201929	0.351	3269	Y
ARC631	*BIRC3*	chr11	102207792	0.267	327	Y
ARC631	*BIRC3*	chr11	102201967	0.117	435	Y‡
ARC646	*ATM*	chr11	108216529	0.188	2208	Y
ARC646	*ATM*	chr11	108206684	0.092	685	Y
ARC679	*ATM*	chr11	108115595	0.921	1463	NA
ARC689	*ATM*	chr11	108150267	0.2	2474	Y

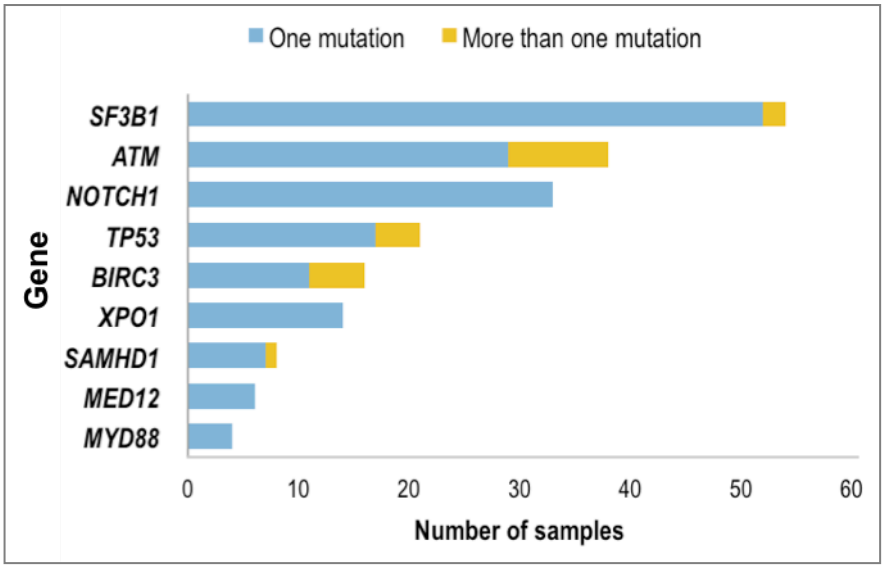
† below detection limit of Sanger sequencing
‡ at detection limit of Sanger sequencing
Y: Yes, N: No, NA: Not done

4.4.3 Results of TSCA panel tests: Mutation frequency and distribution

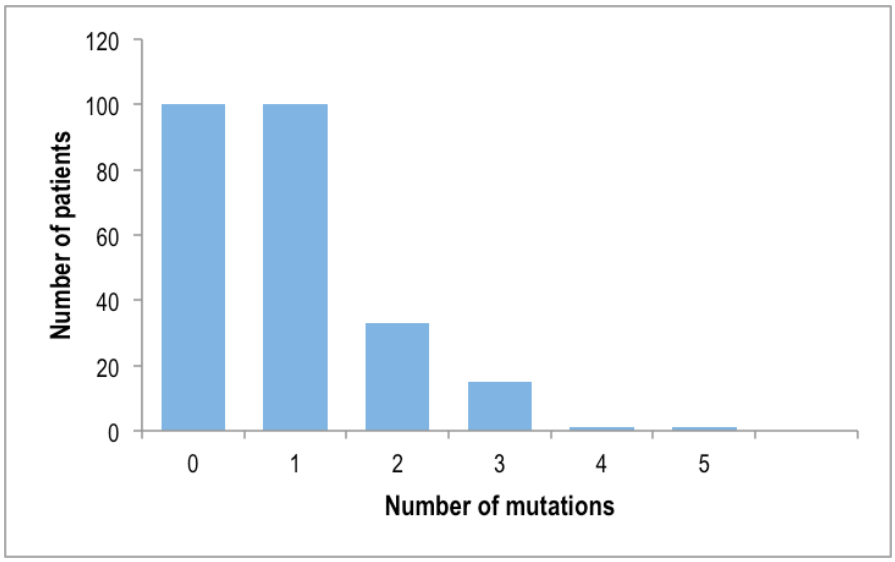
MPS yielded an average depth of 2032X for the 250 CLL samples. High-confidence pathogenic mutations, as described in Chapter 2 Section 2.15.2.4, with a VAF of ≥5% were identified in 150 (60%) patients, totalling 220 mutations (Figure 4.4 A, Appendix 1). Multiple mutations within the same gene for the same patient were seen most frequently for *ATM* but were also present in *BIRC3, TP53, SF3B1* and *SAMHD1* genes. 100 (40%) patients harboured one mutation and 50 patients (20%) had two or more mutations present (Figure 4.4 B). 100 patients did not have any pathogenic mutation detected using the TSCA panel.

The distribution and functional consequence of mutations is shown in Figure 4.4 C. Nonsynonymous SNVs accounted for the majority of mutations: 137 out of 220 (62%), with indels accounting for 68 (31%) mutations found more commonly in *NOTCH1* (n=26/32, 81%) and *BIRC3* (n=18/21, 87%) genes. Stop gain and loss mutations were found in 15 (7%) patients. A small number of mutations in both *ATM* and *TP53* genes were found at splice sites (n=8, 3.6%); however this is not representative, as not all genes were sequenced in their entirety (Chapter 2 Table 2.3).

A


One mutation
More than one mutation
Gene
SF3B1
ATM
NOTCH1
TP53
BIRC3
XPO1
SAMHD1
MED12
MYD88
0
10
20
30
40
50
60
Number of samples

B


120
100
80
60
40
20
0
Number of patients
0
1
2
3
4
5
Number of mutations

C

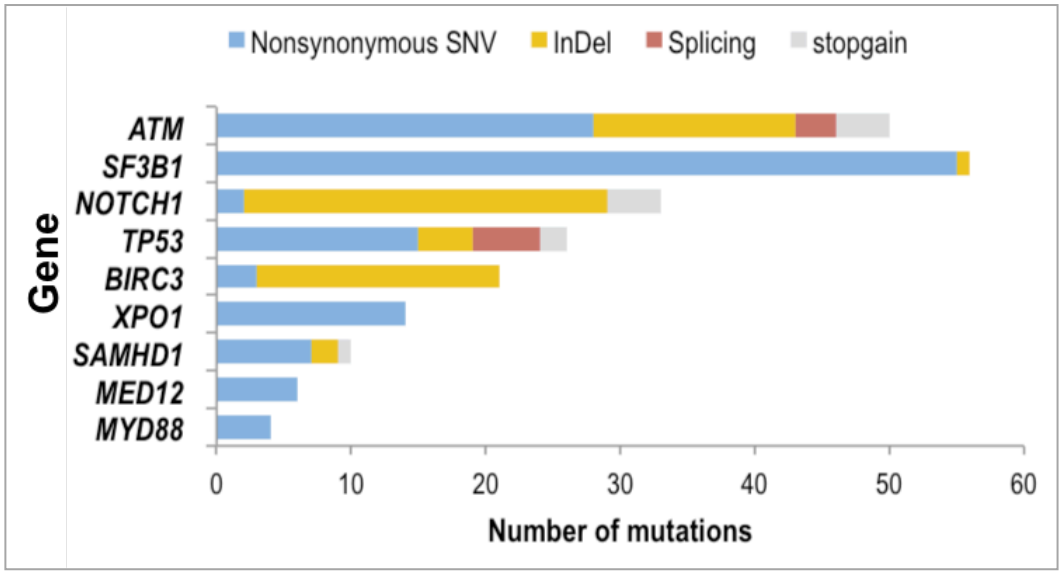


Figure 4.4 TSCA results: Number of mutations and functional consequence

A Frequency of gene mutations per sample. The blue bars represents the number of samples with one mutation in the gene. The yellow bars shows the number of samples with more than one mutation in the gene.

B Total number of mutations per patient. The distribution per patient of the 220 mutations found using the Truseq custom amplicon panel. Median mutation number = 1.

C Distribution and functional consequence of all mutations (n=220) across the nine genes tested. The blue bars represent non-synonymous point mutations; the yellow bars show small indels (less than 50bp); the red bars symbolise the mutations in splicing sites; and the grey bars represent stop gains.

4.4.4 Genes with a mutation frequency greater than 10%

4.4.4.1 SF3B1 mutations

SF3B1 was affected most by recurrent mutations (n=54; 21.6%). Almost all mutations were non-synonymous with the exception of one non-frameshift deletion at c.698-700 (Figure 4.4C). Mutations were located within the HEAT domains 2-8 (Figure 4.5 A), which is consistent with findings in other CLL publications and haematological diseases (Quesada et al. 2011; L. Wang et al. 2011a; Papaemmanuil et al. 2011). The three most frequently mutated amino acids were K700E, K666E and G742D accounting for 39/56 (70%) of all *SF3B1* mutations. Within this cohort, a number of other recurrent mutations (n=8), and non-recurrent (n=9) mutations were identified. Only two patients harboured more than one *SF3B1* mutation. In the case of ARC143, two low level *SF3B1* mutations were found; Q698_700del within exon 15 and N626H within exon 14 at VAFs of 20% and 6% respectively. This sample also harboured a subclonal *ATM* mutation of 7% VAF and copy number losses at both 13q and 11q loci at a similarly low level. Patient ADM321 had two subclonal *SF3B1* mutations, G742D and P780R, with VAFs of 7% and 10% respectively. This patient had a clonal 13q deletion and an *XPO1* mutation at 39% VAF.

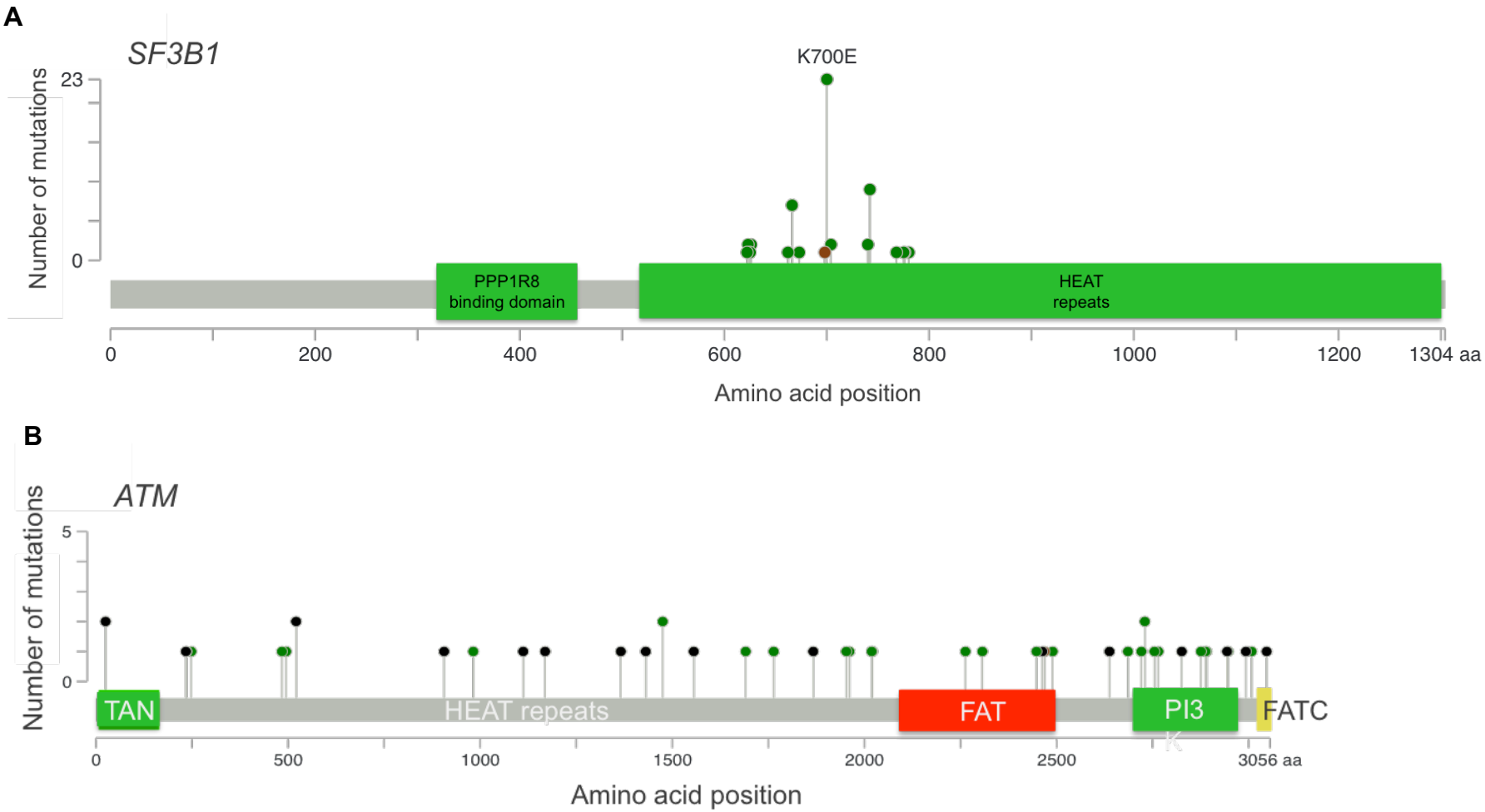


Figure 4.5 Schematic of *SF3B1* and *ATM* protein domains and mutations

A: PPP1R8 protein phosphatase 1 regulatory subunit 8 ; HEAT repeats Huntingtin, elongation factor 3, protein phosphatase 2A, yeast PI3 kinase TOR1

B: TAN Tel1/ATM N-terminal motif; the FAT domain is named after FRAP, ATM and TRRAP; PI3K phosphoinositide kinase domain; the FAT-C domain is named after FRAP, ATM, TRRAP C-terminal

4.4.4.2 ATM mutations

ATM was the second most frequently mutated gene with 50 mutations found in 38 (15.2%) patients. Mutations were located around the C-terminal protein kinase domain but were also dispersed across the entire coding region from exon 3 to exon 63 (Figure 4.5B). The majority were missense mutations (n=29/50, 58%) with nonsynonymous mutations accounting for 28/50 and one non frameshift deletion (Figure 4.4C). Truncating mutations (n=18/50, 36%) including four stop gain mutations and 14 frameshift deletions across the *ATM* gene were also discovered. Outside the coding frame of exons 11, 19 and 20 splice site mutations were seen in three cases.

Based on the mutation selection criteria outlined in Chapter 2 Section 2.16.2.4 germline mutations described in the Ataxia Telangiectasia database were reported here. For two cases, ARC566 and ADM340, the *ATM* mutation was confirmed in the germline. These samples contained an identical frameshift deletion, R521fs, with a VAF of 0.49 and 0.51. A further 10 pathogenic *ATM* mutations with VAFs between 0.45-0.55 and >0.95 did not have paired germline samples and therefore the possibility of a germline origin cannot be ruled out despite stringent reporting criteria.

Table 4.4 Details of somatic and germline *ATM* mutations

Sample ID	Exon	Exonic function	Nucleotide	Amino Acid	VAF	Depth (X)	Germline sequenced	Present in Germline
ADM326	3	Fdel	73_74del	K25fs	0.394	1055	N	NA
ARC528	3	Fdel	73_74del	K25fs	0.455	2099	N	NA
ADM333	7	Fdel	702delA	A234fs	0.322	1685	Y	N
ADM333	7	NS	T707G	L236R	0.324	1683	Y	N
ARC679	7	NS	G743A	R248Q	0.921	1463	Y	N
ADM340	10	Fdel	1561_1562del	R521fs	0.511	1767	Y	Y
ARC091	10	NS	T1450C	W484R	0.5	1407	N	NA
ARC138	10	NS	T1484A	I495K	0.892	557	Y	N
ARC566	10	Fdel	1561_1562del	R521fs	0.49	923	Y	Y
ARC604	11	Splice site	1802+1G>A	NA	0.461	2500	Y	N
ARC104	18	Fdel	2717_2718del	L906fs	0.216	706	Y	N
ARC574	19	Splice site	2921+1G>A	NA	0.805	1241	N	NA
ADM107	20	NS	C2944T	R982C	0.088	692	N	NA
ARC689	20	Splice site	3077_3077+18del	NA	0.207	972	N	NA
ARC689	23	Fdel	3334delC	P1112fs	0.2	2474	Y	N
ARC201	24	Sgain	G3505T	E1169X	0.763	457	Y	N
ARC292	27	Fdel	4096_4097del	C1366fs	0.138	551	Y	N
ADM080	29	NS	A4424G	Y1475C	0.474	587	N	NA
ARC026	29	Fdel	A4292fs	N1431fs	0.098	2766	Y	N
ARC245	29	NS	A4424G	Y1475C	0.509	277	N	NA
ADM106	31	Fdel	4667delA	Y1556fs	0.302	1838	Y	N
ARC176	34	NS	A5071C	S1691R	0.529	276	N	NA
ADM224	35	NS	T5291C	L1764P	0.857	252	Y	N
ADM427	37	Fdel	5599delC	Q1867fs	0.716	443	Y	N
ADM287	39	NS	A5882G	Y1961C	0.189	4361	Y	N
ARC104	39	NS	C5858A	T1953K	0.448	1852	Y	N
ADM272	41	NS	A6056G	Y2019C	0.12	937	N	NA
ADM324	41	NS	G6062A	C2021Y	0.493	796	Y	N
ARC554	46	NS	A6787G	R2263G	0.814	553	Y	N
ARC091	47	NS	C6919T	L2307F	0.418	409	N	NA
ADM295	50	NS	T7469A	L2490H	0.928	1765	Y	N
ADM333	50	Fdel	7388-7393del	L2463fs	0.099	1542	Y	N
ADM366	50	nFdel	7406_7408del	N2469fs	0.59	1642	Y	N
ARC254	50	NS	G7342A	D2448N	0.961	1027	N	NA
ADM169	53	Sgain	G7913A	W2638X	0.95	1041	Y	N
ADM326	55	NS	T8058A	F2686L	0.185	271	N	NA
ADM333	56	NS	C8188A	Q2730K	0.359	468	Y	N
ARC185	56	NS	G8161A	D2721N	0.941	665	N	NA
ARC241	56	NS	A8189T	Q2730L	0.308	474	N	NA
ARC646	56	NS	A8264T	Y2755F	0.092	685	Y	N
ARC042	57	NS	G8293A	G2765S	0.973	667	N	NA
ARC646	58	Fdel	8478delT	N2826fs	0.188	2208	Y	N
ADM041	59	NS	A8662C	I2888L	0.534	4006	N	NA
ARC646	59	NS	T8627A	I2876N	0.119	1381	N	NA
ARC143	60	NS	G8672A	G2891D	0.076	1838	Y	N
ARC528	61	NS	C8840A	T2947N	0.432	1134	N	NA
ARC554	61	Fdel	8831_8832del	T2944fs	0.078	204	Y	N
ARC125	62	Sgain	C8977T	R2993X	0.502	4212	N	NA
ADM117	63	NS	C9022T	R3008C	0.722	870	N	NA
ARC604	63	Sgain	C9139T	R3047X	0.442	481	N	NA

NS=Non synonymous; Sgain= stop gain; Fdel=frameshift deletion; nFdel=non frameshift deletion; Y yes; N no; NA not available

Multiple *ATM* mutations were noted in nine cases, seven cases with two *ATM* mutations, one case contained three *ATM* mutations and a further sample had four *ATM* mutations. The VAFs of these mutations are illustrated in Figure 4.6.

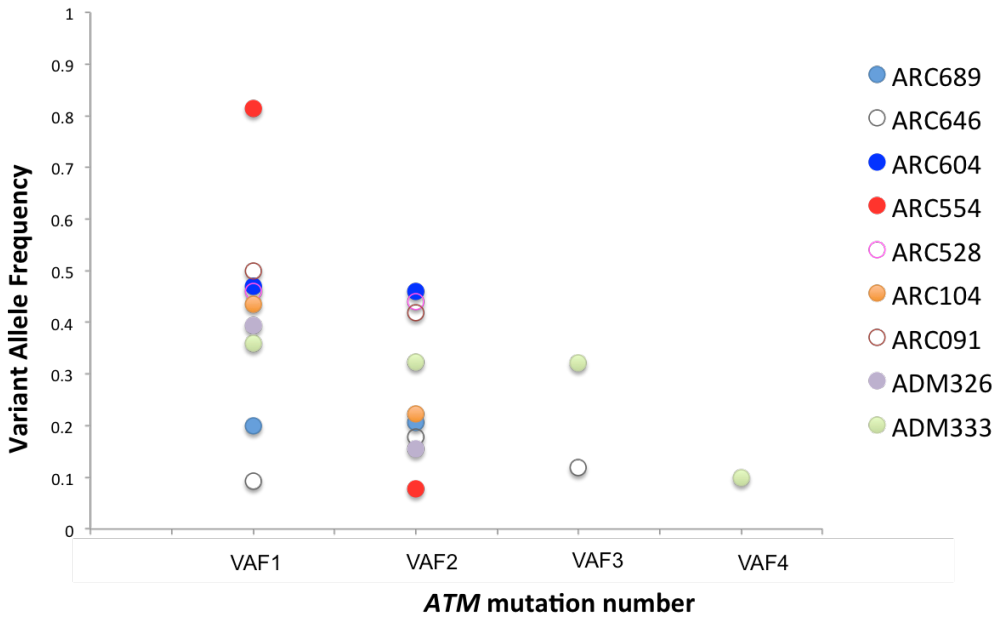


Figure 4.6 Variant allele frequencies of samples with more than one *ATM* mutation

Each sample is represented by a different coloured circle. Each sample has between 2 and 4 *ATM* mutations with VAFs depicted by separate circles of the same colour per sample

4.4.4.3 NOTCH1 mutations

NOTCH1 exon 34 mutations were found in 32 (13.2%) patients. The frame-shift deletion at c.2514_2514 in exon 34 accounted for 26 mutations. A further four stop gain mutations in exon 34 were found (Figure 4.7 A). Only two non-synonymous mutations were found at c.G2186S and c.V2536I, also in exon 34 (Figure 4.4C). Both have previously been described in the COSMIC database (Forbes et al. 2009). Multiple *NOTCH1* mutations in individual patients were not a feature in this cohort.

4.4.5 Genes with a mutation frequency less than 10%

4.4.5.1 TP53 mutations

TP53 mutations were found in 21 patients (8.1%). All exons were sequenced and all mutations were found between exons 4 and exon 9 (Figure 4.7B). Of these 17 were missense mutations including two inframe deletions. A further nine truncating mutations consisting of two frameshift indels, five splice site mutations and two stop gain mutations were identified. Three cases harboured two *TP53* mutations, and a further case carried three *TP53* mutations.

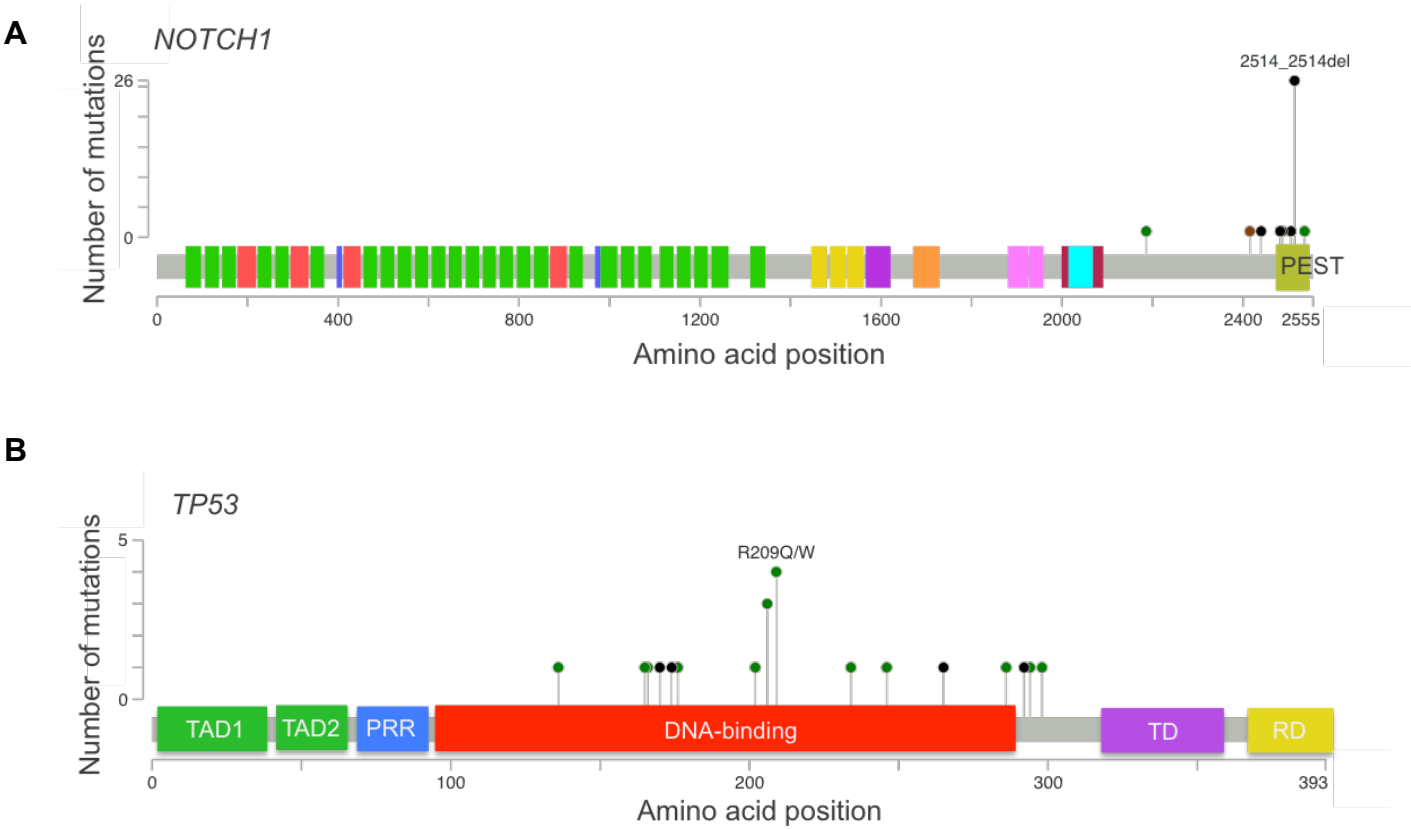


Figure 4.7 Schematic of *NOTCH1* and *TP53* protein domains and mutations

A: PEST protein domain enriched in proline, glutamic acid, serine, and threonine residues

B: TAD transactivation domain; PRR proline rich region; TD tetramerisation domain; RD reguatory domain; green circle: missense; black circle: truncating; brown circle: inframe

4.4.5.2 SAMHD1 mutations

Eight cases contained a total of 10 *SAMHD1* mutations located across the entire gene (Figure 4.8 A). One of these cases contained three *SAMHD1* mutations and has been described in more detail in Chapter 3 Section 4.11. Germline samples were available for four of the eight cases (**Table 4.5**). Of these, three were confirmed to be somatic. The fourth case, ADM007, had one germline mutation and two somatic mutations. Clonal mutations were a feature of these samples and VAFs were in the range of 0.45-0.51 or >0.93 for all mutations. ADM007 was an exception, containing three mutations; two subclonal mutations (VAF 0.17, 0.081) and a germline mutation with a VAF of 0.49%.

Table 4.5 Features of *SAMHD1* mutations

Sample ID	Nucleotide	Amino Acid Change	VAF	Depth	Present in Germline
ADM007	G1165A	A389T	0.081	885	N
ADM007	A617G	H206R	0.176	267	N
ADM007	T602A	I201N	0.492	267	Y
ADM069	G434A	R145Q	0.958	3162	N
ADM210	1694_1705del	565_569del	0.935	1697	N
ADM228	G1715A	W572X	0.982	2327	NA
ADM354	T1478G	L493R	0.511	1137	NA
ARC241	G1112A	R371H	0.464	3633	NA
ARC621	884delG	S295fs	0.448	2010	N
ARC663	G1753A	D585N	0.513	4998	NA

Y yes; N no, NA not available

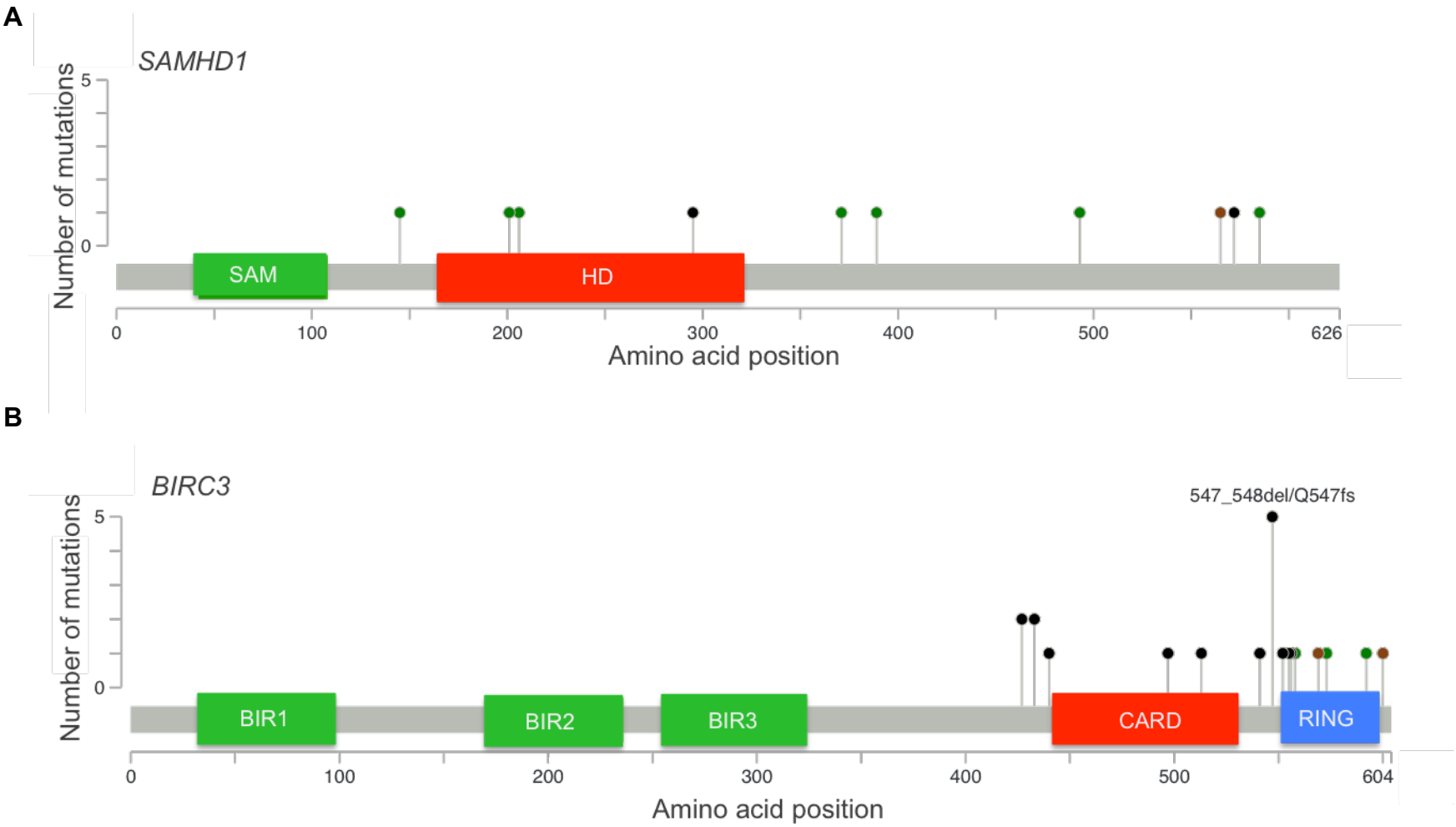


Figure 4.8 Schematic of *SAMHD1* and *BIRC3* protein domains and mutations

A: SAM sterile alpha motif; HD histidine-aspartic domain-containing protein 1

B: BIR baculoviral inhibitor of apoptosis repeat; CARD caspase activation and recruitment domain; RING really interesting new gene type zinc-finger like domain; green circle: missense, black circle: truncating, brown circle: inframe

4.4.5.3 BIRC3 mutations

BIRC3 mutations were found in 16 (6.4%) patients and involved exons 6, 7, and 9 (Figure 4.8 B). All *BIRC3* exons were sequenced and no new hotspot regions were identified in this cohort. Recurrent mutations were seen for three particular mutations found in eight patients (**Table 4.6**). Five patients harboured more than one *BIRC3* mutation as illustrated by the VAFs shown in Figure 4.9.

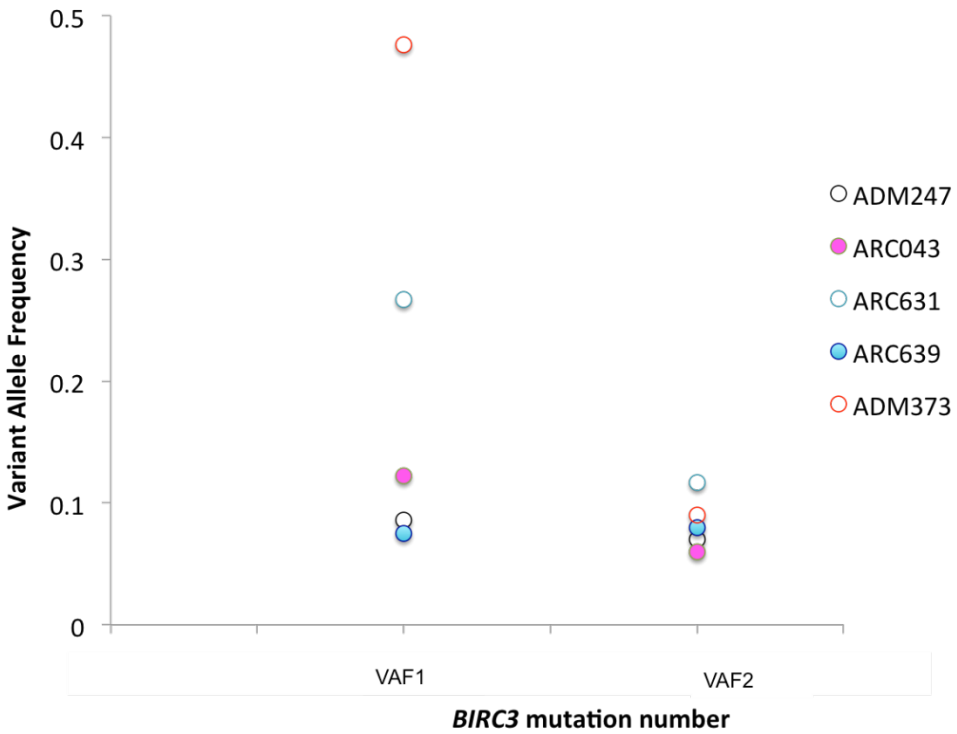


Figure 4.9 Variant allele frequencies of cases with more than one *BIRC3* mutation

Compared to the other mutations across the sequencing panel a high proportion of insertions and deletions were found accounting for 18 out of 21 (87%) *BIRC3* mutations. This compared to a total of 69 out of 220 (31%) mutations overall in the cohort. A number of these mutations were validated by

conventional Sanger sequencing. Six tumour samples containing nine mutations (seven frameshift indels and two missense mutations) were validated and present on Sanger sequencing (**Table 4.6**). Germline material was available for four samples and confirmed the somatic nature of seven of the *BIRC3* mutations including a mutation with a VAF of 0.47.

Table 4.6 Details of *BIRC3* mutations

Sample ID	Exonic function	Amino Acid Change	VAF	Depth	Sanger Confirmed	Present in Germline
ADM029	Fins	S513fs	0.354	2012	Y	NA
ADM068	Fdel	Q547fs	0.253	715	Y	NA
ADM141	Fdel	Q547fs	0.227	916	N	NA
ADM247	nFdel	569_570del	0.066	258	N	NA
ADM247	Fdel	555_555del	0.086	244	N	NA
ADM279	Fdel	Q547fs	0.098	245	N	NA
ADM373	Fins	D497fs	0.09	2981	Y*	N
ADM373	NS	G573D	0.476	679	Y	N
ADM456	Fdel	T556fs	0.148	1046	N	NA
ARC043	Fdel	Q547fs	0.07	1216	Y *	N
ARC043	Fdel	433_433del	0.122	1744	Y	N
ARC095	Fdel	427_429del	0.071	1780	N	NA
ARC502	Fdel	541_547del	0.24	560	N	NA
ARC627	Fdel	427_429del	0.351	3269	Y	N
ARC631	Fdel	440_441del	0.117	435	Y	N
ARC631	NS	R592W	0.267	327	Y	N
ARC639	Fdel	433_433del	0.075	1879	N	NA
ARC639	nFdel	600_603del	0.075	1315	N	NA
ARC643	Fdel	Q552fs	0.279	936	N	NA
ARC679	NS	K558E	0.404	312	N	NA

*below standard detection limit but visually present in electropherogram, Y=yes; N=no; NA=not available; Fins=Frameshift insertion; NS=Non synonymous; Fdel=frameshift deletion; nFdel=non frameshift deletion

4.4.5.4 XPO1 mutations

All 14 mutations in *XPO1* were found at codon 571 with various nucleotide substitutions (Figure 4.10 A). Four mutations affected the nucleotide c.1712 leading to one of E571A/G/V amino acid substitutions. The majority of mutations involved c.1711 and led to the E571K amino acid substitution. The burden of this mutation was illustrated by a range of variant allele frequencies from 0.08 to 0.53.

4.4.5.5 MYD88 mutations

MYD88 mutations were found in four patient samples and consisted of a missense mutation at codon 160 (Figure 4.10 B). The mutation variant allele frequency ranged from 0.37 – 0.51.

4.4.5.6 MED12 mutations

Six mutations in *MED12* were found. Four of these were within exon 2 and two further novel mutations in exon 23 (L1114P) and exon 43 (P2129L) (Figure 4.10). A potential hotspot in the leucine-rich protein domain at codon G44 harboured three of the six mutations in this cohort.

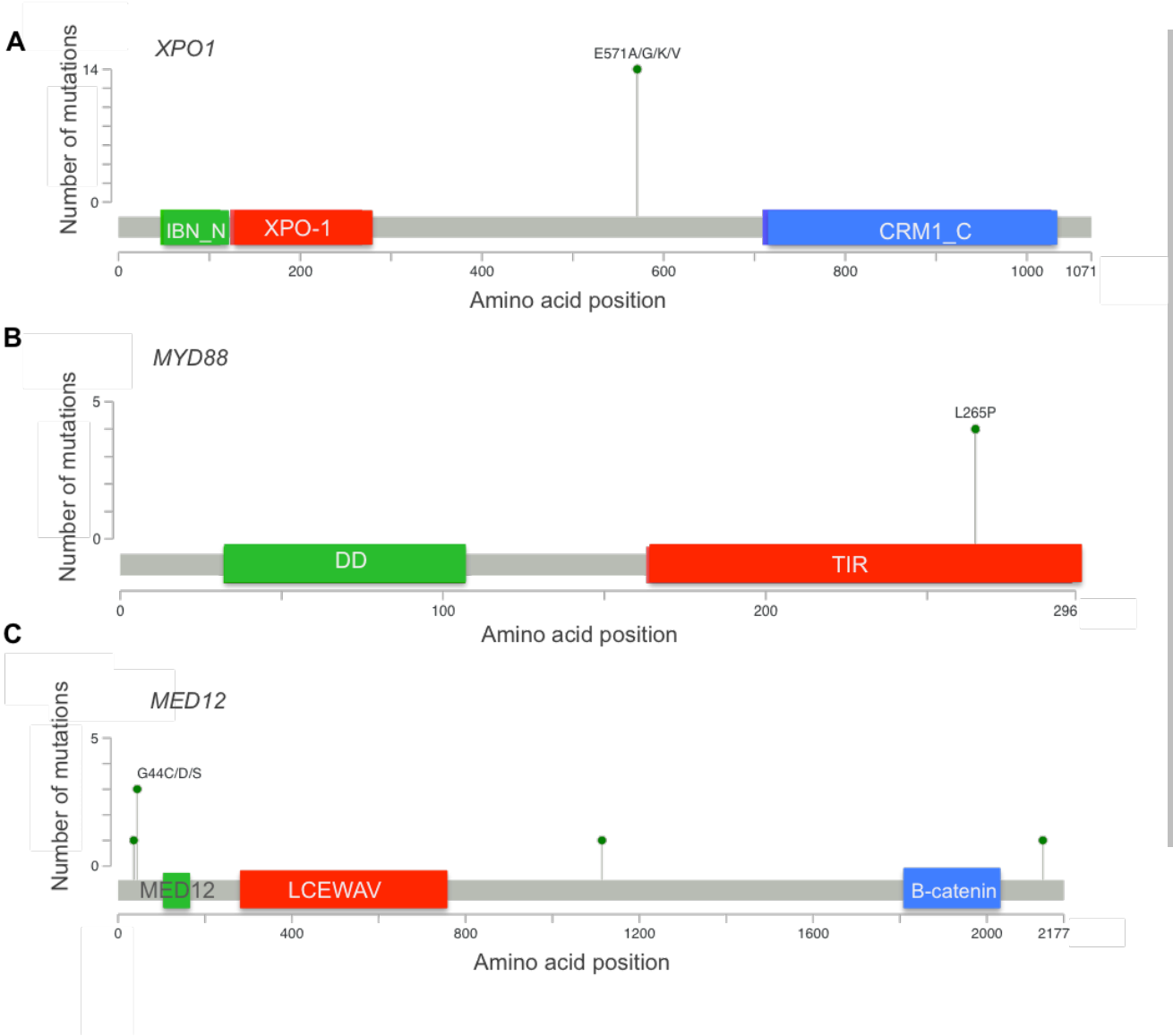


Figure 4.10 Schematic of *XPO1*, *MYD88* and *MED12* protein domains and mutations

A: IBN_N Importin-β N-terminal domain; XPO-1 Exportin-1/Importin-beta-like; CRM1_C CRM1 (also known as Exportin)-C terminal

B: DD death domain; TIR Toll/Interleukin-1 receptor ; C: MED12 mediator of RNA polymerase II subunit 12; LCEWZV conserved sequence; green circle: missense, black circle: truncating, brown circle: inframe

4.4.5.7 Minor TP53 subclones

Recent data have highlighted the negative impact on prognosis of minor subclonal *TP53* mutations at diagnosis that expand post-treatment (Rossi et al. 2014; Nadeu et al. 2016; Malcikova et al. 2014).This deep sequencing technology facilitated the investigation of the incidence of low-level (VAF 0.01-0.05) *TP53* mutations in this cohort. Eleven additional *TP53* mutations in eight patients were discovered. The VAF ranged from 0.015 to 0.045. Two of these eight patients also carried a *TP53* mutation with a VAF>0.05 leaving six patients with a newly identified low level *TP53* mutation (Figure 4.11). The mutations found were those previously described at higher VAFs in the literature and were present in the IARC database (Bouaoun L 2016). Two patients had more than one minor *TP53* subclone present.

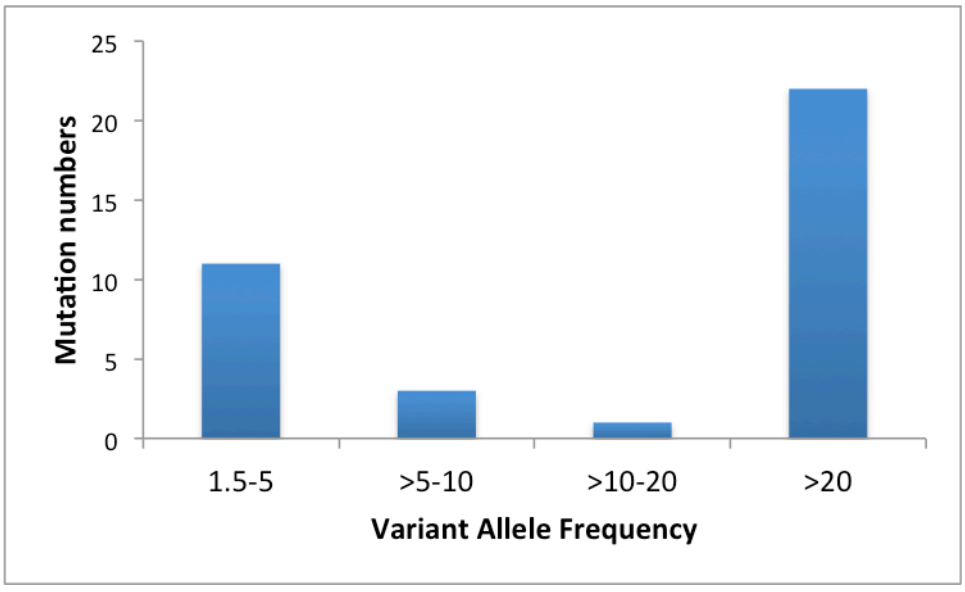


Figure 4.11 *TP53* mutations grouped by variant allele frequency

Table 4.7 Low level *TP53* mutations with VAF <0.05

Sample ID	Exonic function	Nucleotide	Amino Acid Change	VAF	Depth
ARC592	Sgain	C553T	Q278X	0.015	6708
ADM012	NS	G317C	C199S	0.022	2912
ARC017	NS	C346T	R209W	0.022	787
ARC079	Fdel	230_231del	R170del	0.023	1819
ARC079	NS	G121T	V134L	0.023	822
ARC393	NS	G347A	R209Q	0.026	4835
ARC079	NS	C346G	R209G	0.029	787
ARC683	NS	T433G	C238G	0.030	1757
ARC683	NS	G317T	C199F	0.033	5609
ARC304	NS	G64A	G115S	0.036	1057
ARC091	NS	A182T	H154L	0.045	815

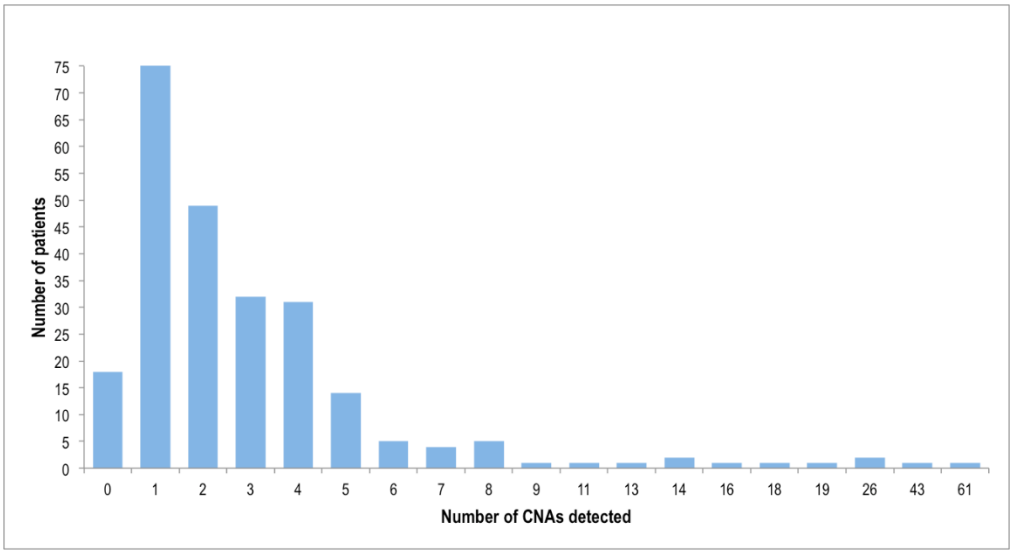
Sgain= stop gain; NS= Non synonymous; Fdel= Frameshift deletion

4.4.6 CNAs identified using high resolution genome-wide array

The majority of patients harboured at least one CNA (n=232/250; 93%). The median number of CNAs was two with a wide range from 0-61 (Figure 4.12 A). Copy number losses (n=633) were more frequent than copy number gains (n=221). In total, 854 CNAs fulfilled the reporting criteria outlined in Chapter 2 Section 2.19.3 (Appendix 2). Recurrent CNAs in critical regions in CLL accounted for 281 of these alterations (Figure 4.12 B). Of these 226 occurring in 190 (76%) cases are routinely tested for by FISH as part of the original hierarchical prognostic model described (Döhner et al. 2000). The remaining 573 alterations were present in less frequently reported or novel regions with recurrences of <5% (Figure 4.13).

4.4.6.1 CLL-associated CNAs

Deletions of 13q14 were the most abundant (n=133, 53%). Deletions of 11q encompassing the *ATM* gene (n=44, 18%) and trisomy 12 (n=40,16%) were the next most common alterations. Recurrent lesions occurring at lower frequencies of 3-5% included deletions of 17p (n=10, 4%), 14q32 (n=9, 3.6%) and 6q21 (n=9, 3.6%). Amplifications of 8q24 (n=8, 3.2%), 2p (n=13, 5%), trisomy 19 (n=7, 2.8%) and trisomy 18 (n=6, 2.4%) featured in this group (Figure 4.12).



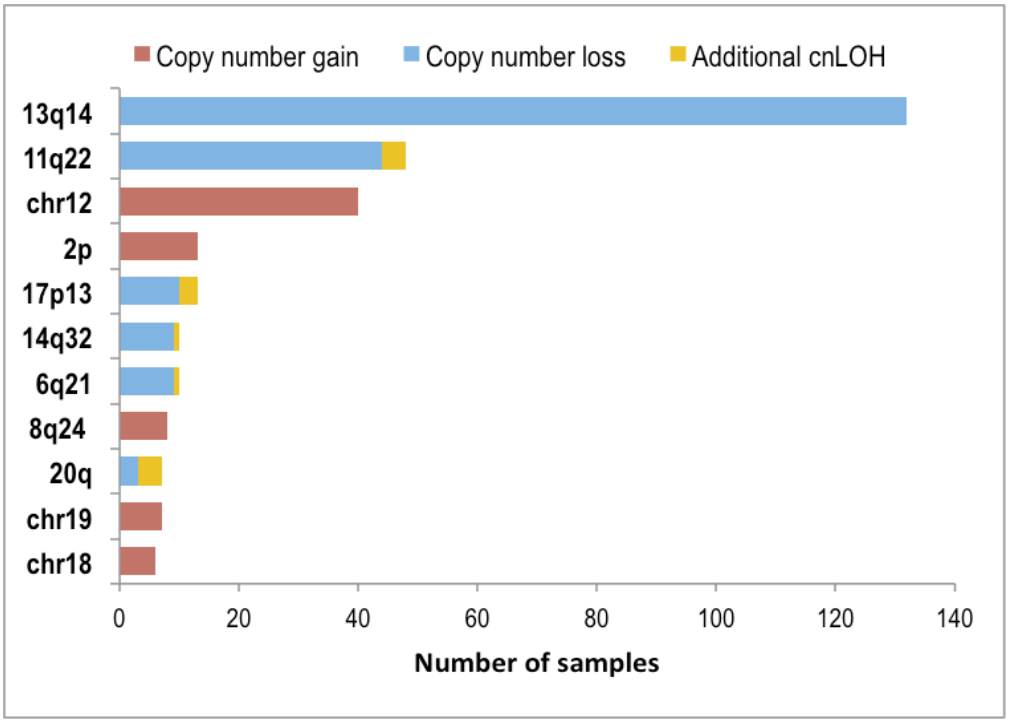


Figure 4.12 Summary of CNA results from genome-wide array on 250 samples

A Total number of CNAs detected per patient; the median number of CNAs was two. B Frequency of CNAs and cnLOH in recurrent CLL regions. The blue and red bars represent, copy number loss or gain respectively; the yellow bars reflect cnLOH.

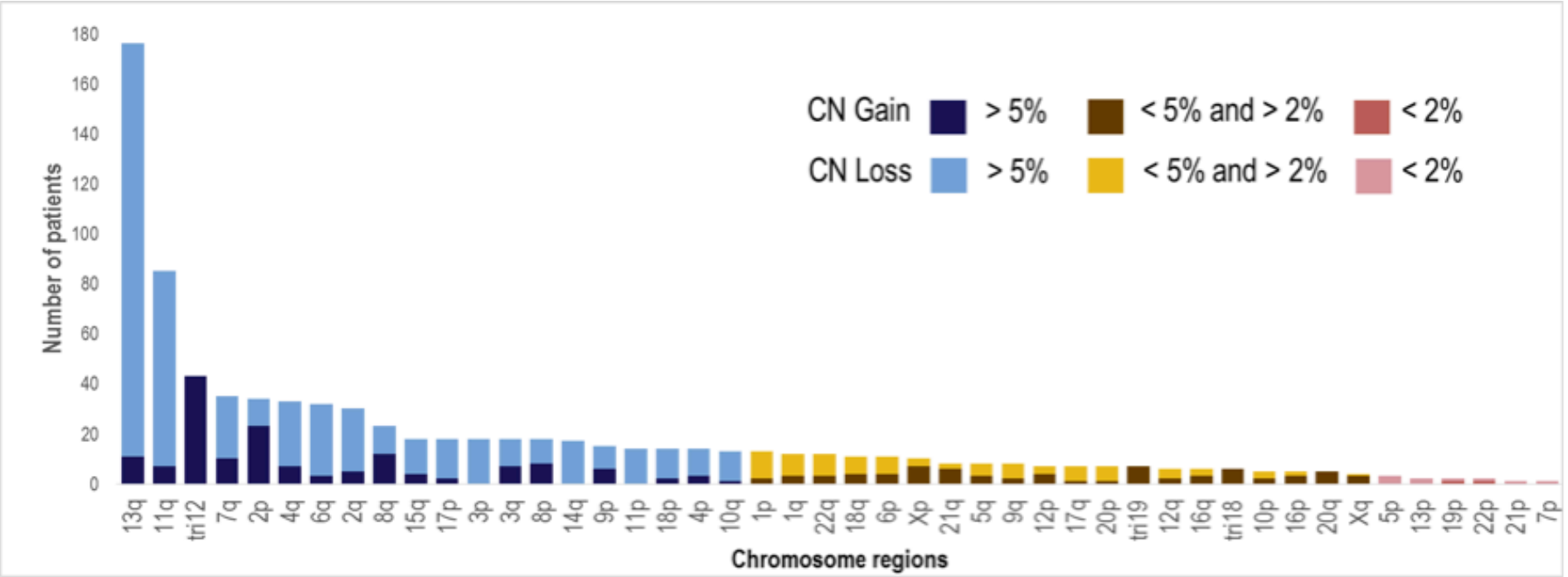


Figure 4.13 Frequency of all CNAs

Distribution of CN gains and losses per chromosome arm. The dark and light blue bars represent CN losses and CN gains, respectively, with a frequency >5% in the cohort. The yellow and brown bars: losses and gains with a frequency of 2%-5%, pink and red bars show losses and gains with a frequency <2%.

4.4.6.2 Chromothripsis

Three patients in this cohort had >10 CNAs on at least one chromosome fulfilling the definition of chromothripsis. A further six patients had at least five CNAs on the same chromosome with a median of eight regions of alternating CNAs (range: 5-29). As outlined in **Table 4.8** six cases had more than one chromosome affected by alternating CNAs. In the example shown in Figure 4.14 when viewed at higher resolution 15 separate regions of copy number loss are seen.

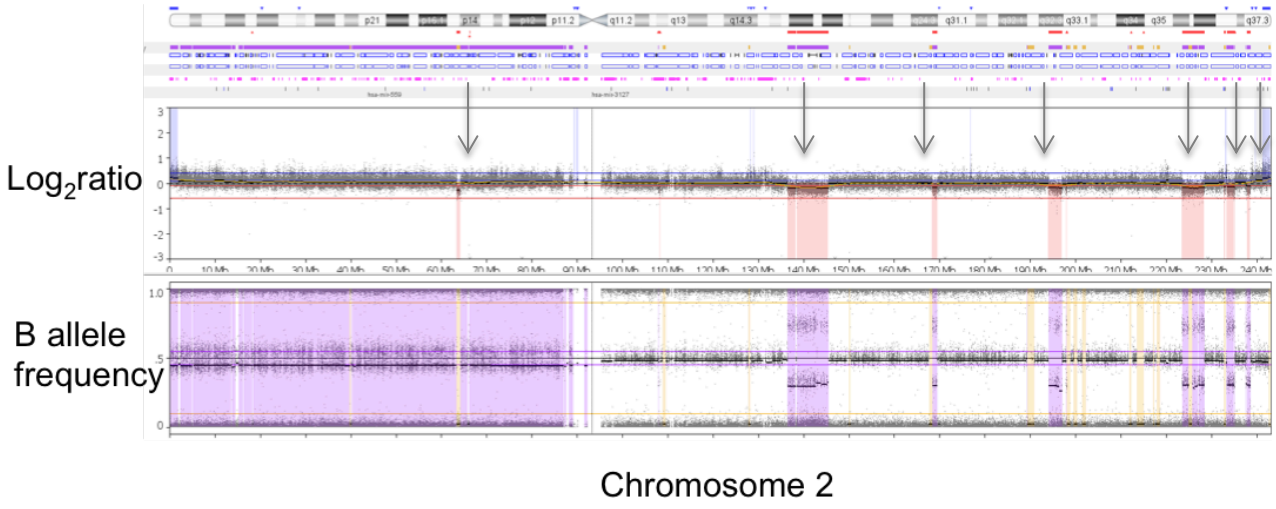


Figure 4.14 Genome-wide array of CLL sample depicting chromothripisis of chromosome 2

Array data visualised using Nexus software, Biodiscovery v6.1, illustrates intermittent copy number loss across chromosome 2 accounting for 15 lesions in total when anaysed in detail. The log_2 ratio plot on top displays the fluorescent signal intensity per probe. A diploid signal has a log_2 ratio result = 0. The B allele ratio plot below reveals the genotype, with a homozygous SNP giving a result of 0 (BB) or 1 (AA) and a heterozygous SNP a result of 0.5 (AB). The yellow line outlines the length of the regoin of cnLOH. Seven grey arrows highlight the regions seen at this resolution, highlighted in red on the log_2 ratio plot. The B allele frequency plot at the regions corresponding to the copy number deletions reflects a subclonal population, these regions are highlighted in purple.

This case, ADM028, also has a subtle subclonal region of LOH across the entire 2p arm reflected by the shaded purple area and tiny separation of the probes on the 0.5 axis of the B allele frequency plot.

Table 4.8 Samples with Chromothripsis

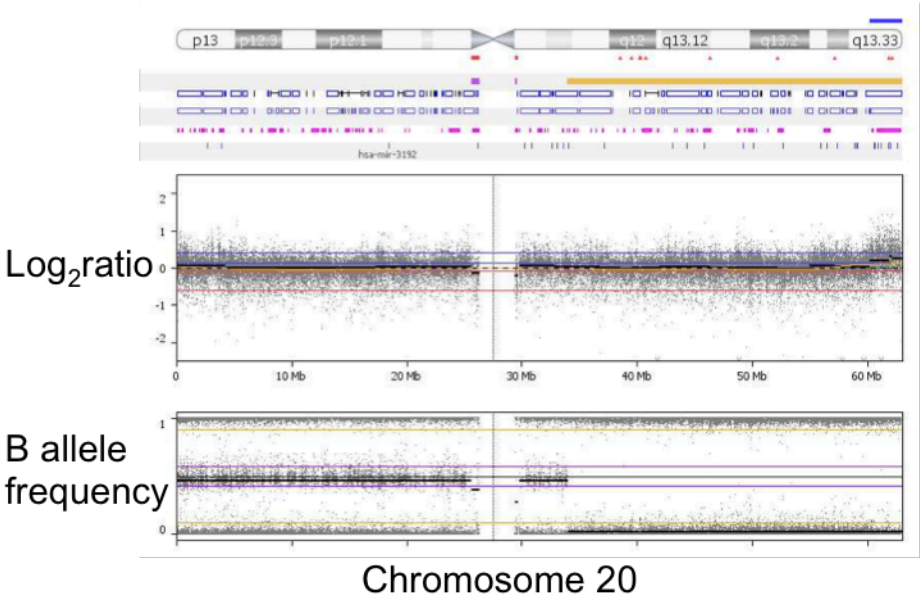
Sample	Chromosome	Number of CNAs
ADM028	2	15
ADM242	2	8
ADM318	6	11
	7	10
ARC241	2	8
	7	6
	8	12
	9	5
ARC304	3	6
	4	7
	13	7
ARC393	6	8
	11	8
ARC484	11	7
	13	9
ARC579	6	6
ARC679	4	21
	7	8
	11	29
	22	10

4.4.6.3 Copy neutral loss of heterozygosity frequently involves chromosomes within critical CLL lesions

The use of genome-wide array technology with analysis of the B allele frequency plot facilitates the identification of cnLOH, where previous technologies would not have recognised this alteration. Regions of cnLOH of 11-115 Mb (median 60.6Mb) were identified in 51 (20.4%) patients. An example of a large region of cnLOH across chromosome 20q is depicted in Figure 4.15

A. This area of cnLOH is greater than 29Mb long and is identified by the absence of a CNA on the log_2 ratio plot while the B allele frequency plot reveals a loss of heterozygosity.

A



B

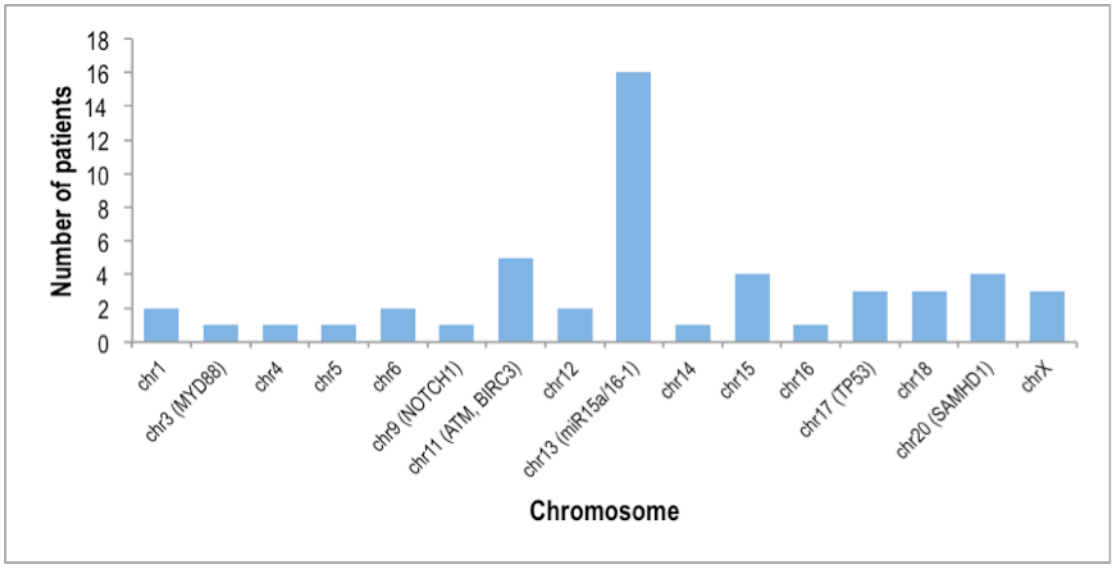


Figure 4.15 SNP array analysis reveals regions of cnLOH

A An example of cnLOH highlighted by the yellow line on patient ADM069. This is visualised using Nexus BioDiscovery v6.1 software. The entire chromosome 20 is shown.

B Distribution of cnLOH events across the genome. The total number of cnLOH regions >10Mb (n=52) reported in the cohort studied. Genes or microRNA within known minimally deleted regions in CLL encompassed by the region of cnLOH are highlighted on the X-axis within brackets.

Though cnLOH events occurred across the genome, these events were more concentrated in regions affected frequently by CLL-associated CNAs and mutations (n=30, 60%) (Figure 4.15). The region most frequently affected by cnLOH was chromosome 13 with 18 samples harbouring a large cnLOH of between 20.6Mb and 97Mb (median 95Mb). Other regions affected by cnLOH included chromosome 17p involving the *TP53* locus and chromosome 11 involving the loci of both *ATM* and *BIRC3* genes.

4.4.6.4 Features of heterogeneity over the 13q14.2-14.3 locus

Details of the 13q deletions encompassing the minimally deleted region (MDR) of 13q14.2-14.3 found in 133 (53%) cases are depicted in Figure 4.16. Varying lengths of 13q deletions were noted ranging from 107.6kb to 95Mb (median = 2.38Mb). 58 cases (44%) involved the *RB1* gene (Figure 4.17).

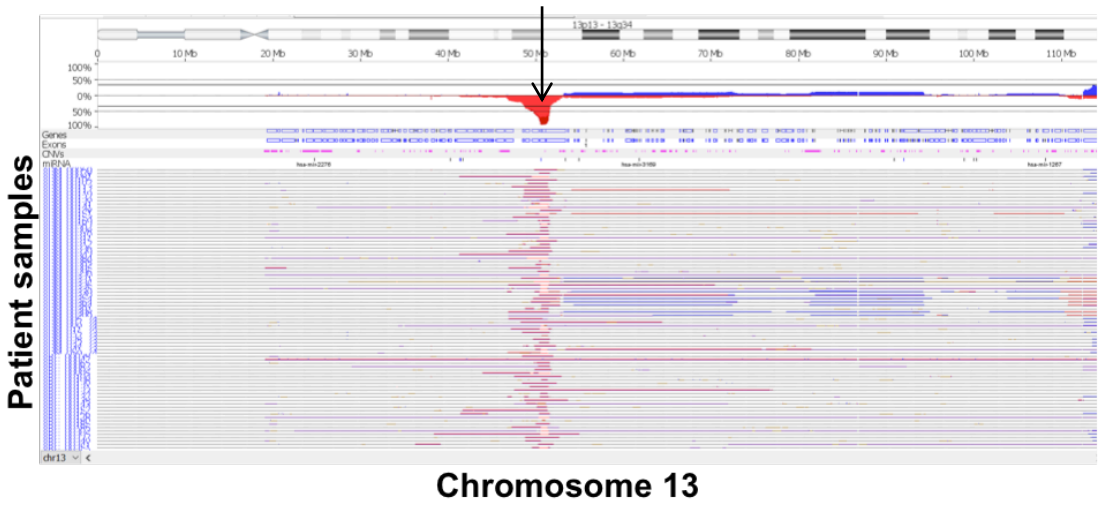


Figure 4.16 Heterogenous 13q deletions on 132 cases

Overview of all 13q deletions illustrated by Nexus software (BioDiscovery, Inc., El Segundo, CA, USA). Red line - CNA loss, black arrow – 13q minimally deleted region (MDR)

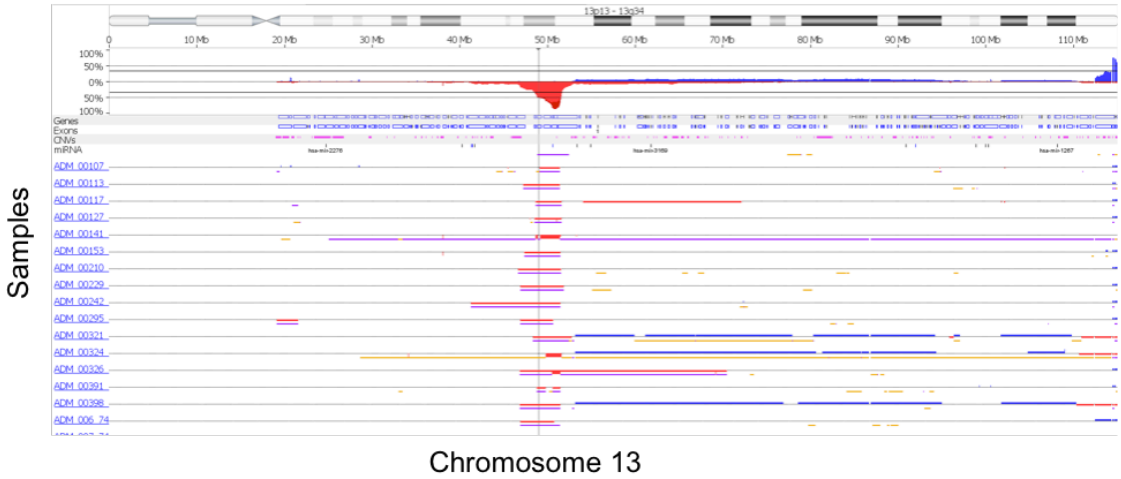


Figure 4.17 Genome-wide array showing 13q deletions involving *RB1*

Overview of longer 13q deletions involving *RB1* illustrated by Nexus software (Biodiscovery) Red line - CNA loss, grey vertical line intersecting *RB1*

Of these 133 cases, 27 carried homozygous deletions ranging from 160kb to 4.6Mb in length (median 877kb). In addition 18 patients had large regions of cnLOH (median 95Mb, range 20.66-97.01Mb). 10 patients had both a homozygous 13q deletion and a large region of cnLOH. While 17 cases had a homozygous 13q deletion without evidence of cnLOH, all of these cases were longer 13q deletions encompassing the *RB1* locus (Figure 4.18 A). All of the homozygous 13q deletions that were associated with 13q cnLOH were shorter deletions and did not overlap the *RB1* locus (Figure 4.18 B,C).

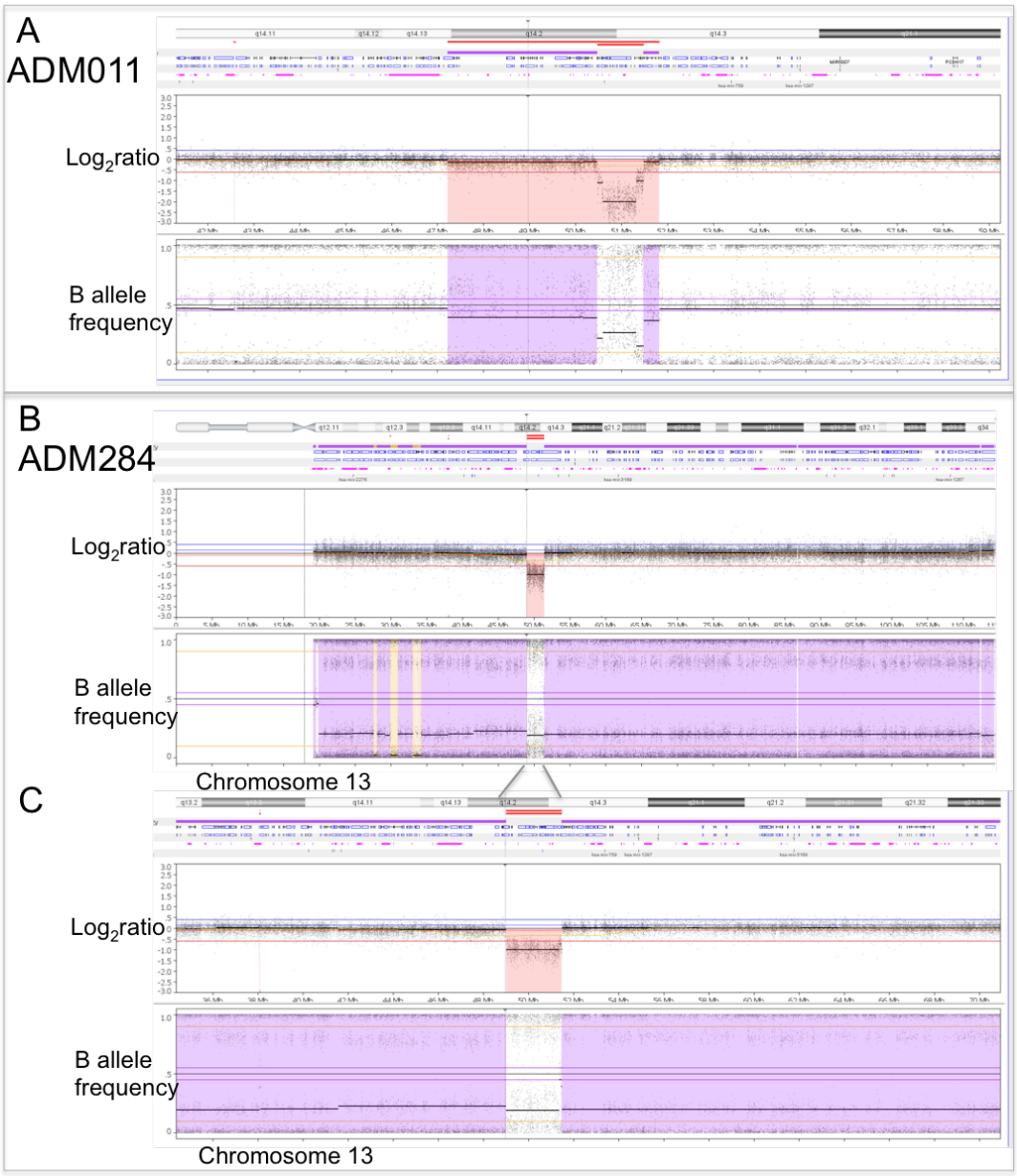


Figure 4.18 Gemone-wide array reveals 13q heterogeneity

Two examples of chromosome 13 deletions involving the critical MDR as illustraed using Nexus v6.1. A Sample ADM011, the red line represents a copy number loss, single line is a heterozygous loss, double ine is a homozygous loss. The purple shading represents a subclonal lesion and the heterozygous copy number loss is in a subclone only. The fine vertical grey line is the *RB1* gene which is only involved in the heterozygous loss. B: Sample ADM284 contains a short homozygous copy number loss (red region) and also a subclonal 13q cnLOH (purple region) not including the *RB1*, breakpoint is telomeric of the *RB1* gene as seen in C at higher resolution as the vertical grey line pinpoints the *RB1* gene.

13q locus are depicted in Figure 4.19.

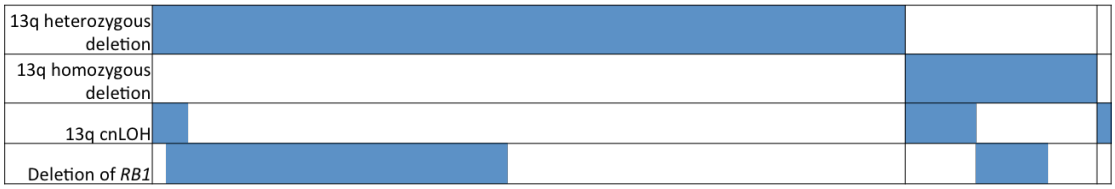


Figure 4.19 Deletions and cnLOH of 13q

Association of findings at the 13q cluster from SNP array. The 135 informative patients in this cluster were sorted depending on the length of heterozygous deletion, presence of homozygous deletion and the presence of cnLOH. In the chart, The light blue bar represents presence of each finding and each column represents one patient. A differentiation was made between the presence of a shorter deletion (excluding the *RB1* gene) and a longer deletion (including the *RB1* gene).

One case, ADM391, harboured two 13q deletions of 0.9Mb and 1.0Mb in size

Figure 4.20.

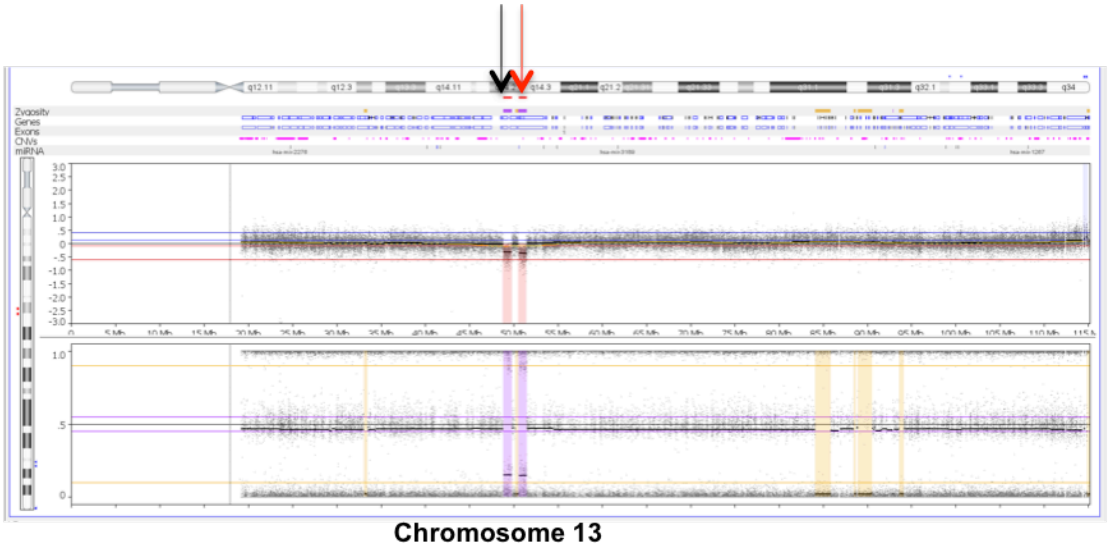


Figure 4.20 Case of two 13q deletions: ADM391

Two heterozygous 13q deletions in one patient illustrated by Nexus software (Biodiscovery). Black arrow involves RB1, Red arrow – 13q MDR over MIR15a/16.2

A further case, ADM141, contained two subclonal homozygous 13q deletions in addition to subclonal cnLOH of the 13q arm (Figure 4.21). The larger homozygous deletion was 2.2Mb and encompassed the known 13q MDR of *DLEU 1*, *DLEU 2* and MIR15a/16.2 locus. The smaller homozygous deletion was 271.5kb and the telomeric breakpoint terminated just at the *RB1* gene (Figure 4.21B).

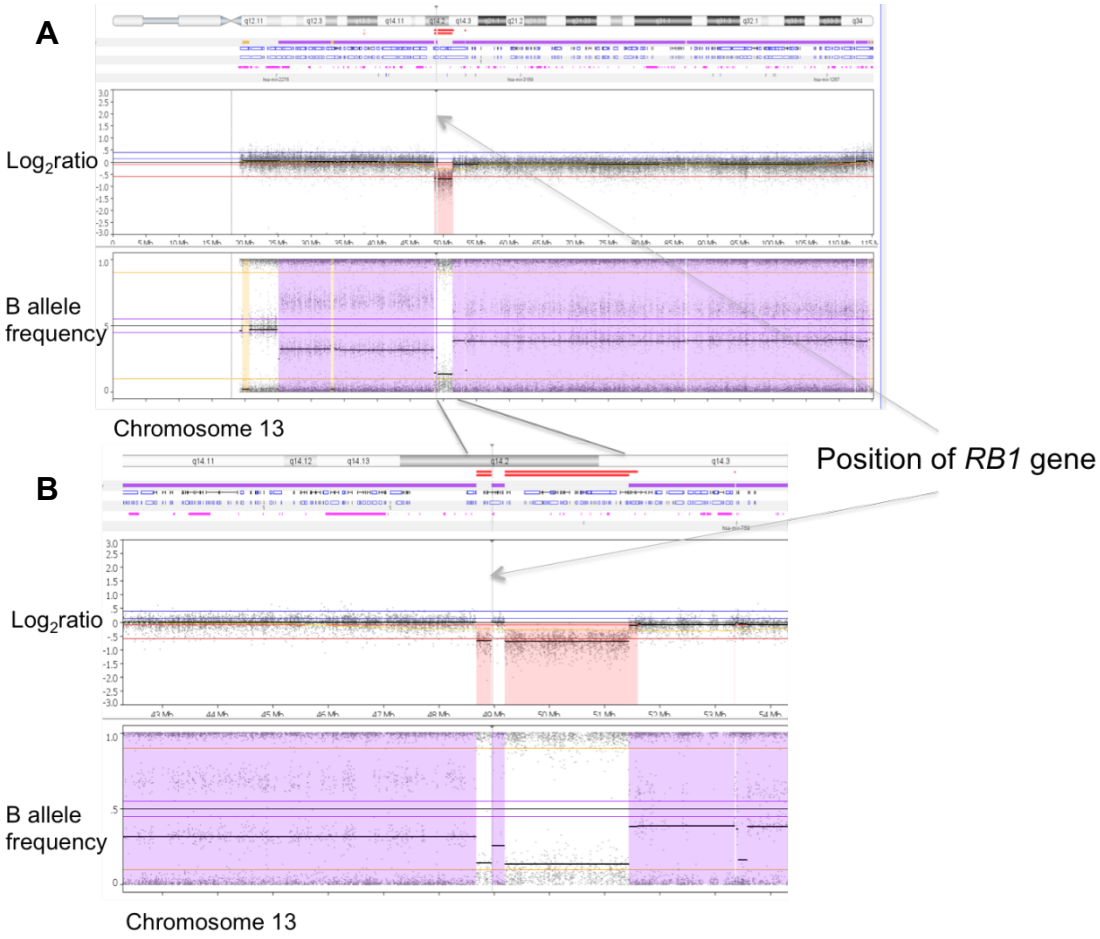


Figure 4.21 Separate deletions across the 13q locus illustrated by whole-genome array

A The entire chromosome 13 is depicted, the purple region outlines a subclonal cnLOH based on lack of a change from 0 of the log2ratio plt while probes sit between 0-0.5 and 0.5-1 of the B allele frequency plot. The red bar relates to a homozygous deletion, also a subclone based on the level at the log2ratio plot
B Zoom in to the regions of 13q homozygous deletions reveal the small and larger areas. Telomeric end of shorter deletion contains *RB1* gene, vertical line highlights the position of *RB1* gene.

4.4.7 Integrated sequencing and copy number analysis

When combined, the mutation and CNA data revealed that 231 (92.4%) cases had at least one CLL driver lesion (Figure 4.22). Overall, only 19 patient samples did not reveal a CLL driver, 78 had one CLL driver affected and 153 patients had two or more drivers affected (Figure 4.22).

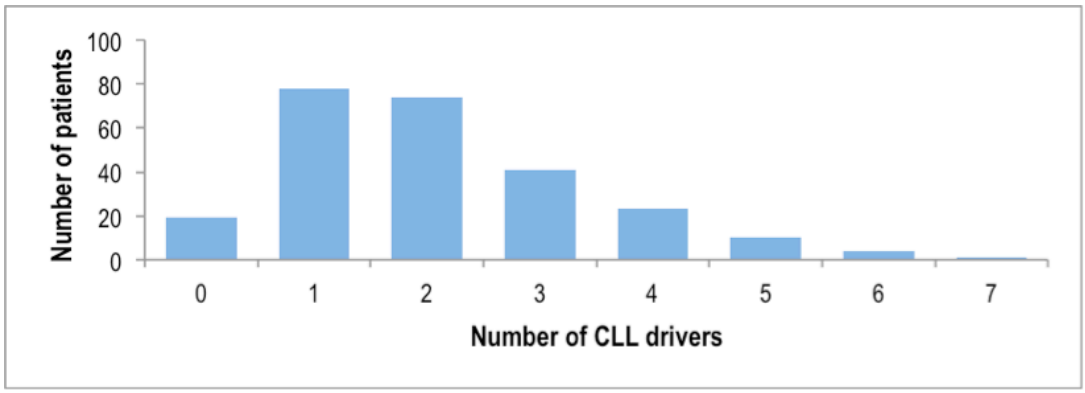


Figure 4.22 Number of CLL driver lesions

Distribution of the number of CLL drivers in the cohort.

A further 3% of cases harboured low-level recurrent lesions (Figure 4.23 'other CNVs, not recurrent'). Only six patients (2.4%) lacked any detectable genomic alteration using this combined testing approach.

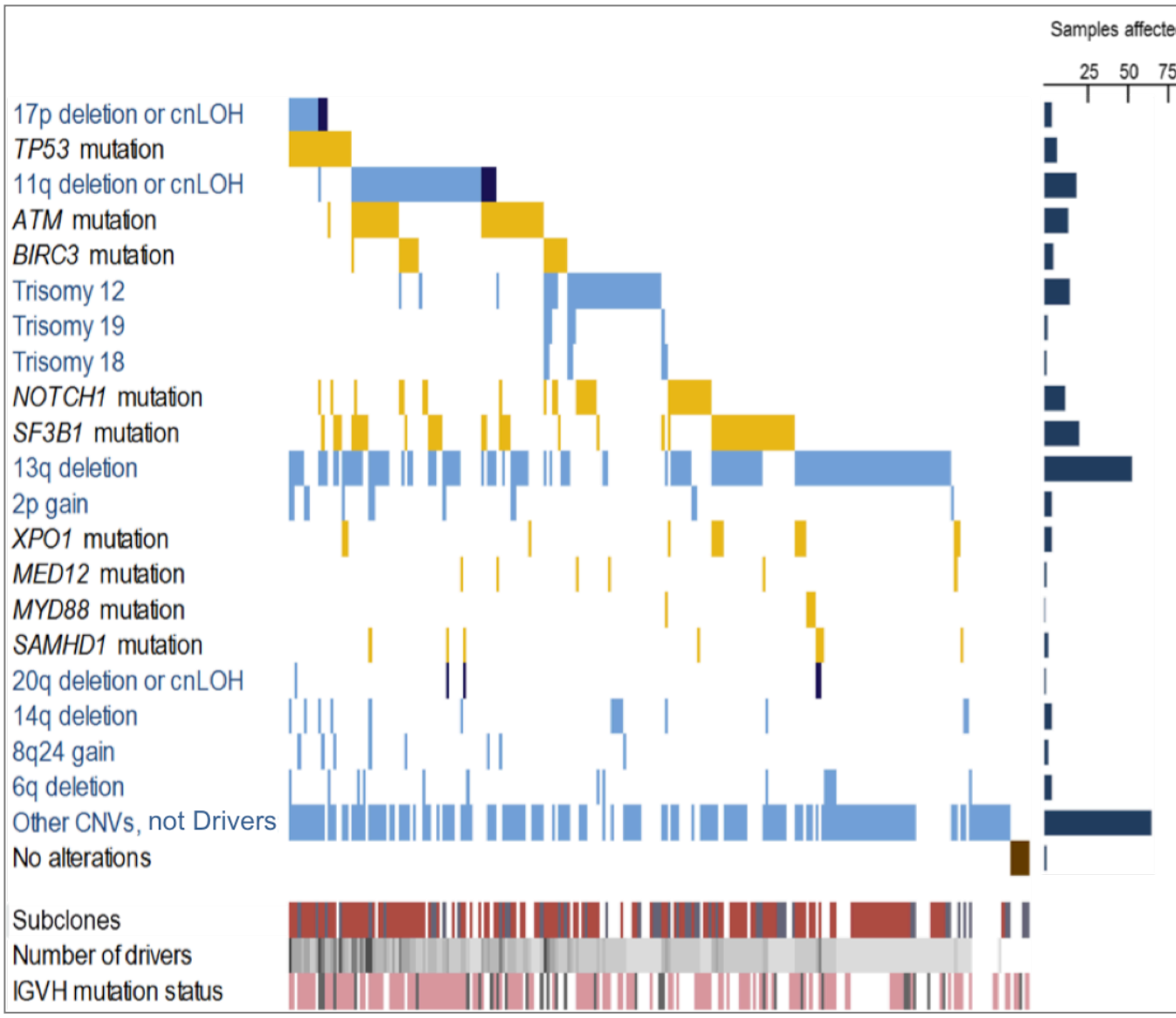


Figure 4.23 Plot integrating sequencing and SNP array data

Association of mutations and copy number alterations along with additional findings in the 250 patient cohort. Each column represent a patient. Each row represents a unique or multiple events; CNAs in light blue, additional cnLOH regions in dark blue, SNVs and indels in yellow. Apparent absence of any event is represented by the row 'no alterations' in brown. At the bottom of the figure other features are represented including the presence/absence of subclones (dark red/white, grey not available), increasing number of CLL drivers (colour coded from light grey: no drivers to dark grey: 6 or more CLL drivers) and *IGVH* mutation status (unmutated in pink, mutated in white, not available in dark grey).
To the right of the figure the percentage of cases affected by each feature is represented in dark blue.

4.4.7.1 Pairwise gene associations

Testing for pairwise gene associations, either co-occurring or mutually exclusive lesions, revealed that 17 pairs were significant (Figure 4.24). Mutations in *TP53* associated strongly with *TP53* deletions and cnLOH at this locus. *ATM* mutations co-occurred with 11q deletions and with 11q cnLOH. Furthermore *SAMHD1* mutations associated with cnLOH across chromosome 20q. From the integrated genomics plot (Figure 4.23) it is clear that few cases of *ATM* disruption (either a mutation or a deletion) are seen in cases with *TP53* disruption. In cases with *SAMHD1* mutations, there were no findings of *TP53* mutation or deletion. This finding will become more significant later in survival analyses in Chapter 5.

13q deletions were highly mutually exclusive to trisomy 12. In this cohort only five cases had both 13q deletions and trisomy 12. Interrogating the findings for these cases revealed that all five had subclonal 13q deletions present.

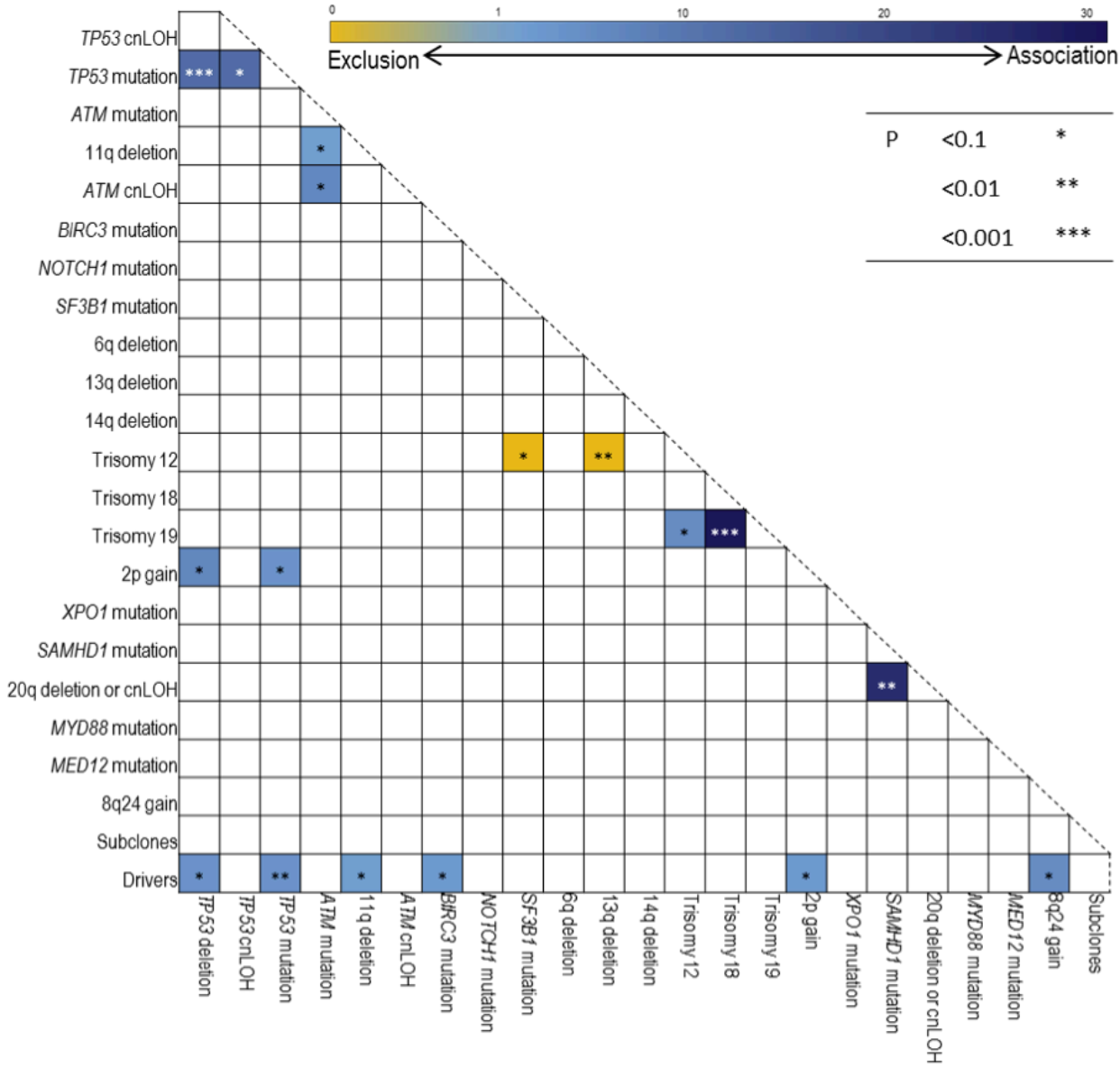


Figure 4.24 Heatmap of genetic lesions outlining significant associations

Heatmap showing significant pairwise gene associations, either co-occurring or mutually exclusive lesions. The strength of the association is represented by different intensities of colour: from light blue to strong is weak to strong association; from light yellow to strong yellow is weak to strong mutual exclusion. It is measured by the relative difference between the number of observed and expected events. Only statistically significant associations and exclusions are presented. Subclones refers to the presence of 1 or more. Drivers refers to CLL Drivers of ≥4 (Corrected Fisher's exact test with a False Discovery Rate of 10%).

Cases with *SF3B1* mutations were mutually exclusive to trisomy 12. Only one case had both a *NOTCH1* mutation and a *SF3B1* mutation. This patient had a low level *SF3B1* mutation at a variant allele frequency (VAF) of only 7%.

Trisomy 12 was highly correlated with trisomy 19 and a non-significant trend to co-occur with trisomy 18 was observed. Trisomy 19 significantly associated with trisomy 18. No significant statistical relationship between trisomy 12 and *NOTCH1* was found.

While not significant it was noted that cases with 13q deletions can harbour many CLL drivers and there is a tendency for these cases to have an unmutated IGHV locus. Cases with 13q deletions and a mutated IGHV locus harboured fewer additional CLL driver mutations. This is illustrated in Figure 4.23 by comparing the left side of the waterfall plot with the right side. No significant difference between number of CLL drivers and heterozygous or homozygous 13q deletion was noted (Figure 4.25).

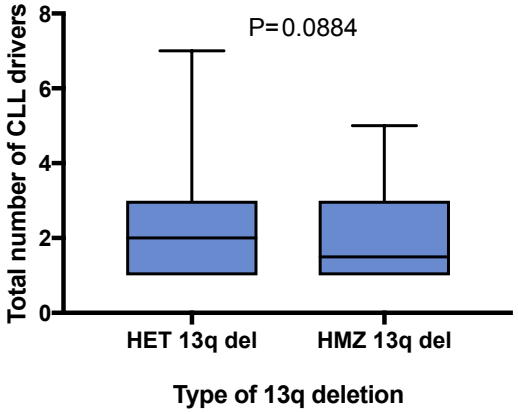


Figure 4.25 Total number of CLL drivers and 13q deletion type
All 13q deleted cases are represented here, divided by whether a heterozygous only or homozygous deletion was present. P value was calculated using an unpaired t-test.

A higher number of CLL driver lesions associated with *TP53* disruption and with deletions of 11q. Gains of 2p and 8p also associated with increased numbers of CLL drivers as did mutations of *BIRC3*. Gains of 2p were also associated with *TP53* deletions and mutations.

4.4.7.2 Chromosome 11q locus findings

Presented here are the interactions between cases with 11q deletions, cnLOH over the 11q locus, *ATM* mutations and *BIRC3* mutations. In Figure 4.26 bi-allelic events include both *BIRC3* and *ATM* mutations in cases with either deletions or cnLOH across the 11q locus. All cases of cnLOH of 11q were found to have *ATM* mutations. Deletions including the *BIRC3* gene always also included the *ATM* locus, however 10 patients with shorter 11q deletions encompassed *ATM*, but not *BIRC3* (Figure 4.26). In 50% of cases *BIRC3* or *ATM* mutations were not accompanied by an 11q deletion or cnLOH (8/15 and 22/45 respectively) leading to likely mono-allelic inactivation only. One case, ARC679, carried an 11q deletion and was affected by both *ATM* and *BIRC3* mutations. This 11q deletion did not encompass the *BIRC3* locus.

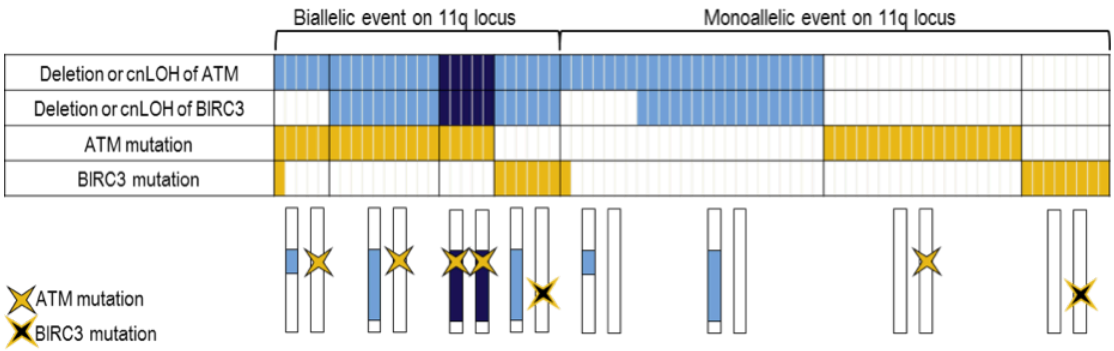


Figure 4.26 Association chart showing 11q locus findings

Association of mutations, CNAs and cnLOH on the chr11q cluster. The 75 informative cases were sorted based on a biallelic or monoalleic event. The light and dark blue bars represent CNAs and cnLOHs, respectively; the yellow bar represents mutations. Each column represents one patient. On the chromosome representation below the table, the same colour code was applied. A distinction is made between deletions including *ATM* only (short blue bar) and including both *ATM* and *BIRC3* (long blue bar). All cnLOH events encompassed both *ATM* and *BIRC3*. The *ATM* mutations are yellow stars and the *BIRC3* mutations and represented with black stars.

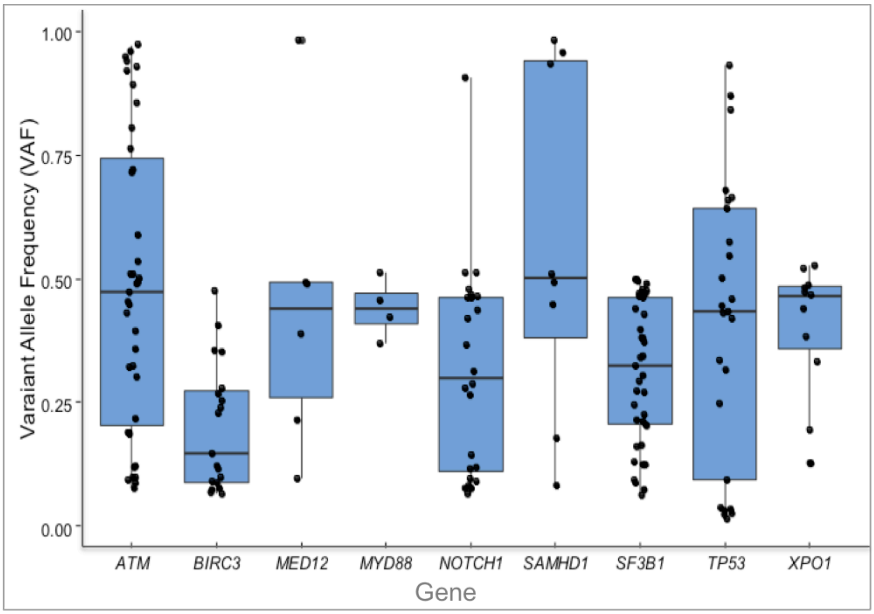
4.4.8 Quantification of subclones to define samples with intra-tumour heterogeneity

An impact of subclonal mutations on clinical outcome has been suggested for some time (Schuh et al. 2012; Landau et al. 2013; Knight et al. 2012; Guièze et al. 2015). In this study, subclonal populations were defined by both array and sequencing data as described in Chapter 2 Section 2.21.

A per gene analysis revealed that mutations in *SF3B1*, *NOTCH1* and *BIRC3* were predominantly subclonal, whereas *MYD88*, *XPO1, SAMHD1* and *MED12* mutation*s* were predominantly clonal (Figure 4.27A).

Both mutation and copy number data were included in the subsequent analysis. The majority of samples had at least one subclone present (n=133/203, 66%), with 75 samples (34%) harbouring solely clonal lesions (Figure 4.27B). Genome-wide array findings confirmed subclonal alterations in 105 cases. Mutation data from the TSCA panel accounted for 28 additional subclonal cases. Of these 15 were defined by point mutations and 13 by indels.

A



B

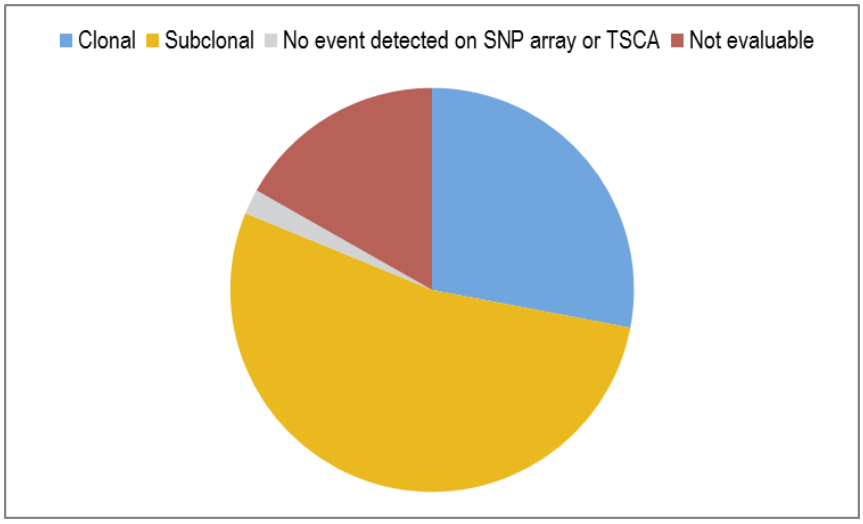


Figure 4.27 Subclonal analysis

A Distribution of Variant Allele Frequencies (VAFs) per gene. The black dots represent the individual VAFs per mutation. The blue box represents the lower and upper quartile ranges. The horizontal black line represents the median VAF per gene and the vertical line the distribution

B Distribution of clonal and subclonal alterations in the cohort (n=203). The proportion of patients with only clonal or subclonal events (CNAs, cnLOH and mutations) is shown in blue and yellow, respectively.

4.5 Discussion

Here we studied somatic SNVs and indels using MPS and integrated the results with CNAs from genome-wide arrays. As different genomic features can be found at various periods of the CLL disease course it is important to study a clearly defined well annotated homogenously treated group of patients to draw meaningful conclusions regarding frequency of genomic events and more importantly when reporting clinical outcome (Guièze & Wu 2015). In total, 250 samples were available for study and this was well representative of the entire clinical trial cohorts without any remarkable differences in patient characteristics. The cases studied were sampled at the time prior to first treatment in two randomised controlled clinical trials (Howard *et al*, 2017; Hillmen 2013).

4.5.1 MPS using TSCA panel is a reliable and sensitive platform

Overall the results from the TSCA panel were robust and validated using conventional Sanger sequencing on a subset of samples. The same methodologies have recently been validated on a cohort of AML samples where the TSCA panel with 25 genes targeted was found to be reliable and accurate when compared with both Sanger and pyrosequencing (Fernandez-Mercado et al. 2013). While the VAF limit of 5% was selected for our dataset initially, the sensitivity analysis highlighted that it is possible to identify a VAF as low as 1% provided adequate sequencing depth had been achieved.

In the absence of germline samples it is essential to filter pathogenic variants and report only the true somatic mutations. The TSCA panel was applied to

paired somatic and germline samples from 16 patients to test the stringency of the variant selection pipeline. This was a highly selected group as only a limited number of germline samples were available to our laboratory from the UK CLL biobank. In this group only a small number of false positives were identified, perhaps not surprising given that it was a group selected also on the basis of harbouring *ATM* mutations and often with more than one additional mutation of uncertain pathogenicity.

For both *ATM* and *SAMHD1* genes an exception to the filtering criteria was applied. Germline mutations were accepted if previously reported in patients with Ataxia Telangiectasia or Acardi Goutieres Syndrome, the inherited genetic diseases associated with germline *ATM* and *SAMHD1* mutations respectively (Tachibana 2009; Rice et al. 2009). These germline mutations accounted for only three out of all mutations found in this cohort.

The distinction between somatic and inherited mutations in cancer genomics is a very helpful tool, aiding the selection of pathogenic variants through subtraction methods, especially in the era of next generation sequencing and Big Data to filter out private SNPs (Brown et al. 2016). For CLL bioinformatics pipelines, and possibly more broadly for cancer genomics, the relevance of inherited gene mutations such as *ATM* and *TP53* may need to be considered. This has implications for the standard subtraction approaches currently applied. More recently, germline pathogenic mutations have been included in larger CLL genomic projects highlighting the importance of this point in CLL cases with potential *ATM* germline mutations (Nadeu et al. 2016).

4.5.2 Gene mutations and associations from MPS

The mutation frequency for the commonly mutated genes in CLL was within ranges expected in a pre-treatment cohort of CLL patients. While the *SF3B1* and *NOTCH1* mutation frequencies were higher than in other publications, many of these are population-based series and samples were taken at any clinical stage of CLL (Jeromin et al. 2014; Baliakas et al. 2015). For a comparable cohort the German CLL 8 clinical trial reported mutation frequencies of 18.4%, 10% and 11.8% for *SF3B1*, *NOTCH1* and *TP53* respectively. In keeping with our findings a recent review of various next generation sequencing projects confirmed that mutation frequencies of *SF3B1* and *NOTCH1* mutations in addition to T*P53* and *ATM* increased over time and with clinical stage (Guièze & Wu 2015). This however is not the case for predominantly clonal lesions such as MYD88, trisomy 12 and 13q deletions as these lesions are thought to be present early in the disease course as putative founder mutations and therefore the frequency of these lesions is relatively static over time (Landau et al. 2013).

After applying a stringent filtering pipeline 150 cases had at least 1 mutation present and 50 cases had two or more mutations. More than one mutation per gene was a feature of certain genes tested by the TSCA panel. Of the 38 cases with *ATM* mutations 9 had two or more *ATM* mutations. It is possible that these mutations are present in different CLL clones based on the VAFs seen and that this is reflecting convergent evolution of the CLL clone towards ATM inactivation. Similarly with *BIRC3* mutations five out of 16 cases carried more

than one *BIRC3* mutation and again the VAFs were at levels where a second CLL clone could be present. Only single cell analyses are likely to be able to dissect the clonal architecture of these patients conclusively.

Our knowledge of CLL driver mutations has rapidly expanded since this project was undertaken and panels with increasing numbers of targets to interrogate these drivers further will increase the number of mutations found per case (Landau et al. 2013; Landau et al. 2015; Puente et al. 2015).

4.5.3 Whole genome array reveals common and novel aberrations in the majority of cases

CNAs were present in almost all cases tested (93% of cohort) either the more common CNAs that are tested for routinely by FISH (76% of all lesions found) or less common lesions of undetermined significance in CLL. A high number of genomic alterations was not a feature in this cohort with on average two CNAs per case. This confirms previous reports in comparable populations (Gunnarsson et al. 2011; Ouillette et al. 2008; Edelmann et al. 2013). The absence of germline samples for subtraction analysis can lead to overcalling aberrations however stringent criteria were applied in the analysis phase (Chapter 2 Section 2.19.3) to correct for this limitation. This finding of low CNAs is consistent when compared with a similar pre-treatment cohort from the German CLL8 trial and is therefore reassuring that the reporting criteria applied are capturing significant aberrations (Edelmann et al. 2013).

4.5.4 Heterogeneity at 13q locus points towards convergent evolution

Previous groups have outlined features of 13q deletions that correlate with an inferior outcome such as longer length of deletion or inclusion of *RB1* within the 13q deletion Interrogating samples with 13q deletions further has revealed a more complex picture. Here, examples with homozygous 13q deletions with and without cnLOH are also presented.. In cases of homozygous deletions without cnLOH, the 13q deletion was longer and involved the *RB1* gene. One can speculate that these observations suggest a process of convergent evolution across the 13q locus. Possibly pointing towards functional redundancy existing between the biological process of a large region of cnLOH equating to duplication of all or part of one 13q chromosome and a longer 13q deletion including *RB1*. *RB1* has not been identified in CLL as a recurrently mutated gene. Different genomic mechanisms may be involved but the overall consequence is that key genes/regulatory elements may be deleted with either direct or downstream perturbance of gene expression. The spectrum of findings at this locus points to a dynamic region undergoing evolution in this cross-sectional study. A limitation of this study is that chromosomal translocations have not been excluded. While previously in CLL these genomic features have been infrequently reported, as technology improves increasing frequencies of CLL translocations are reported, and can include the 13q MDR region (DE Braekeleer et al. 2016; Burns et al. 2017; Puente et al. 2016; Haferlach et al. 2007).

4.5.5 Integrating genomics reveals CLL driver in 95% of cases

Combining both array and sequencing technologies enables a more accurate picture of the genomic landscape of CLL. Almost all cases had at least one CLL driver affected with only 18 cases without a CLL driver. This compares favourably with reports from a larger dataset tested by whole exome sequencing. In that study of 538 cases, 278 from the frontline treatment setting, all but 8% of cases were defined by at least one CLL driver (Landau et al. 2015). It is notable that multiple driver mutations were not a feature in this cohort acknowledging the limitation that since the time of this study greater numbers of CLL drivers have been defined (Landau et al. 2013; Landau et al. 2015; Puente et al. 2016; Ljungström et al. 2016; Wang et al. 2017) Specifically the finding of *TP53* disruption occurring in tandem with *ATM* disruption was not common compared to a study on CLL in the relapsed refractory setting where cases with mutations in all three of the CLL drivers *TP53*, *ATM* and *SF3B1* genes were seen (Guièze et al. 2015). This complex evolution is seen more frequently in relapsed refractory CLL (Knight et al. 2012), however certain cases in this cohort displayed features of genomic complexity and cases with chromothripsis (n=3) or chromothripsis-like events (n=6) were seen.

4.5.6 Pair-wise associations

Patterns of co-occurring or mutual exclusive lesions bring us closer to understanding the phylogenetic order of mutation acquisition. For mutually exclusive lesions such as trisomy 12 and 13q deletions, as outlined in this study, this points either towards functional redundancy or possibly that initial clonal lesions such as these influence subsequent genetic lesions, rather than pointing towards a common pathogenic mechanism which seems unlikely for

these large alterations. For trisomy 12, a subgroup of these cases carried trisomy 19 and 18. *NOTCH1* and *BIRC3* mutations were not seen in this subgroup of trisomy 12 but were a feature of trisomy 12 cases. *SF3B1* was mutually exclusive to trisomy 12 and also to *NOTCH1*. A small number of cases carrying both trisomy 12 and 13q deletions (n=5) or *NOTCH1* and *SF3B1* (n=1) were all found at subclonal levels indicating that these lesions were likely to be in different CLL clones and maintaining the mutual exclusivity associated with these lesions. To confirm this finding single cell analysis of these samples would be required.

Statistical conclusions regarding the hierarchy of mutation acquisition was not possible in this group despite the large number of cases tested. It has been calculated that over 2000 CLL samples would be required to provide robust statistically meaningful interactions of all CLL drivers now known to us (C Holmes, 2014, personal communication). This underlines the need to study greater numbers of patients in international collaborations (Lazarian et al. 2017; Guièze & Wu 2015)

4.5.7 Clonal and subclonal mutations discovered both on array and sequencing panel

It is possible to infer the subclonal architecture of an individual's cancer cell and from this the timing of mutation acquisition from MPS and genome-wide array data (Papaemmanuil et al. 2013; Landau et al. 2013). The frequency of subclonal populations present in 66% of patients studied here is comparable to that reported in the more extensively studied cases of Landau *et al.* (2013) (Landau et al. 2013). Additionally the only solely clonal lesion from MPS was

MYD88, again in agreement with Landau's work. In this study a targeted panel was used, and as there was a low number of targets on the TSCA panel and a low mutation frequency present per sample clustering analysis could not be robustly performed. While bioinformatics algorithms have been specifically developed to interrogate the subclonal composition of samples using WGS and WES, TSCA panel subclonal data analysis required bespoke methodology as outlined in Chapter 2 Section 2.21 that included 203/250 samples to improve the stringency of the analysis (Carter et al. 2012). It is notable that VAFs between WGS and targeted deep sequencing are comparable as has been validated by R. Alsolami (2015, personal communication) in addition to work published on the pilot WGS data (Schuh et al. 2012).

4.5.8 Conclusion

Here, the genomic landscape of CLL at the frontline treatment time-point in 250 clinical trial patients is outlined utilising MPS and high-resolution genome-wide array. The features of somatic SNVs, indels and CNAs and their interactions are described. The next Chapter will address the clinical utility of measuring these parameters and whether there is clinical value in progressing from FISH and *TP53* sequencing to a broader panel testing for more CLL driver mutations.

CHAPTER 5 : TOWARDS RESPONSE PREDICTION IN CLL - INTEGRATED GENOMICS USING CLINICAL TRIAL SAMPLES

5.1 Introduction

CLL is clinically and biologically heterogeneous. Patients with progressive disease requiring treatment have variable responses to chemo-immunotherapy (CIT). Whilst 90% of patients achieve a clinical response to the gold standard treatment of Fludarabine, Cyclophosphamide and Rituximab (FCR), 76% of patients suffer Grade 3/4 toxicity and 25% relapse within two years (Hallek et al. 2010; Tam et al. 2014). Patients relapsing within 24 months of treatment have been shown in a Phase II study using CIT to have an overall life expectancy of 27 months and represent an unmet clinical need (Tam et al. 2014). Highly-active small molecules targeting different CLL survival pathways have proven efficacy irrespective of *TP53* status such as Ibrutinib and Idelalisib (Byrd et al. 2014; Furman et al. 2014). However, side-effects, potential long-term toxicity and cost remain obstacles to adoption as frontline therapy. Developing a risk-based genotype-specific model for frontline therapy is a desirable goal in the emerging era of precision medicine.

As outlined in Chapter 4 the genomic landscape of frontline treated CLL reveals a CLL driver lesion in the majority of cases. Previous efforts at genetic risk stratification were heralded by the German model with four separate risk groups defined by FISH testing in the setting of combined Fludarabine and

Cyclophosphamide treatment (Döhner et al. 2000). More recently, when FCR was defined as the new standard of care in the frontline treatment of CLL patients the FISH-defined hierarchical model retained prognostic significance (Hallek et al. 2010). *TP53* mutations were also included in this model. Cases bearing *TP53* mutations by Sanger sequencing or *TP53* deletions detected by FISH continued to show poor outcomes in the era of CIT (Hallek et al. 2010).

Many studies since the application of next generation sequencing have discovered that novel genetic markers are also associated with poor outcome in CLL. These markers included *NOTCH1*, *SF3B1* and *BIRC3* mutations in addition to standard FISH testing and *TP53* sequencing (Rossi et al. 2012; Rossi et al. 2011; Rossi et al. 2012; Rossi et al. 2013; Jeromin et al. 2014; G. Fabbri et al. 2011; L. Wang et al. 2011a; Quesada et al. 2011). The studies were usually undertaken on diagnostic samples or a heterogenous group of patients at various stages of their disease course and patients who required treatment were usually not treated with CIT.

To truly define prognostic markers, the sample of patients studied should ideally be treated homogenously to derive meaningful conclusions. Clinical trial cohorts have predefined inclusion criteria i.e. first-line treatment and will have robust clinical outcome collated for accurate analysis of survival outcomes. The CLL4 clinical trial enrolling patients in the frontline treatment setting has outlined genomic features and clinical outcome of patients treated with Chlorambucil, Fludarabine or Fludarabine and Cyclophosphamide (Catovsky et al. 2007). In this cohort poor outcomes for biallelic *ATM* disruption, *SF3B1*,

NOTCH1, and *TP53* disruption have been confirmed (Rose-Zerilli et al. 2014; Gonzalez et al. 2011; Oscier et al. 2013; Skowronska et al. 2012).

The molecular eradication of CLL is possible with modern CLL treatment (Ghia 2012). Minimal residual disease negativity, defined by guidelines as a level of detection below 10^{-4}, correlates with prolonged progression free and overall survival (Kwok et al. 2016). The CLL8 trial confirmed that in the era of FC and FCR therapy, MRD negativity after three months of treatment associates with longer progression free survival (PFS) and overall survival (OS) independent of IGHV status and 17p deletion status (Böttcher et al. 2012). While this useful variable can predict long-term outcome, a pre-treatment biomarker is preferable.

5.2 Aims and Objectives

The aim of the work described in this Chapter is to examine if an association exists between genetic measures reported in Chapter 4 and clinical outcome in the setting of frontline treated patients in clinical trials. The outcome measures evaluated include clinical response, minimal residual disease (MRD) response and progression free survival (PFS). The specific objectives were (i) to examine global genetic changes and single genetic markers in combination with clinical outcome using robust statistical tests, including survival analyses , to evaluate whether they might be informative and aid prognostication (ii) to test the validity of significant findings both statistically and in an independent clinical trial cohort of 247 frontline CLL samples.

5.3 Materials and Methods

Materials and methods relevant to this Chapter are outlined in Chapter 2.

5.3.1 Patient samples

The patient samples and clinical outcome results are those described in Chapter 4; 250 frontline CLL patients enrolled in UK clinical trials treated with chemo-immunotherapy. The validation cohort consisted of 247 CLL patients from the UK LRF CLL4 trial (Catovsky, Richards, Matutes, Oscier, M. Dyer, et al. 2007). Full details of both cohorts are outlined in Chapter 2 Section 2.1.

5.3.2 Statistical analysis

Univariate and multivariate analyses were performed using the Chi-square test in IBM SPSS v22. Cox regression analysis with backward selection was selected for model development. Corrections for multiplicity testing were applied (Šidák 1967). Kaplan and Meier survival analysis with log-rank test was applied for survival data using GraphPad Prism V5. Significance was defined as <0.05. To validate our statistical findings, a penalised Cox proportional hazards (PH) regression approach was applied. Further details of the statistical methods are given in Chapter 2 Section 2.28.1.

5.4 Results

5.4.1 Patient Characteristics

The characteristics of 250 patients from two clinical trials on frontline treated CLL are outlined in **Table 5.1**. Patient characteristics were representative of the entire cohort as presented in Chapter 4. The median white cell count at trial enrolment was 114 x 10^9/L. Patients were mostly Binet stage B and C (n=216, 86%).

Of the 250 cases tested 118 were enrolled in the ADMIRE trial and 132 the ARCTIC trial. Within the ADMIRE trial, 57 patients received FCR treatment and 61 FCM-R treatment. Within the ARCTIC trial, 65 patients received FCR treatment and 67 patients FCM-miniR.

Table 5.1 Patient characteristics including clinical outcome

	Analysed (n=250)		Not analysed (n=165)		
Characteristic	No.	%	No.	%	P value
Age					
Median	61		62.7		0.092
WBC					
Median	113.6		109.9		0.149
Lymphocyte Count					
Median	105.2		99.23		0.096
Haemoglobin					
Median	11.94		12.18		0.512
Platlelets					
Median	149.2		153.5		0.667
Sex					
M	182		116		
F	68		49		0.580
Binet stage					
A progressive	34	13.6	27	16	
B, C	216	86.4	138	84	0.436
Trial, Treatment arm					
ADMIRE					
FCR	57	22.8	50	30	
FCMR	61	24.4	47	28	0.636
ARCTIC					
FCR	65	26	35	21	
FCM-miniR	67	26.8	33	20	0.765
IGHV mutation status					
Mutated	91	36.4	51	31	
Umutated	121	48.4	75	45	0.659
N/A	39	15.6	39	24	
FISH Del17p					
Del17p	13	5.2	4	2	
Non del17p	219	87.6	134	81	0.229
Missing	18	7.2	27	16	
FISH Del11q					
Del11q	46	18.4	20	12	
Non Del11q	188	75.2	121	73	0.178
Missing	17	6.8	24	15	
CD38					
Positive	100	40	60	36	
Negative	139	55.6	92	56	0.643
Minimial Residul Disease					
Negative	110	44	76	46	
Positive	106	42.4	59	36	0.322
Final Response					
CR/PR	220	88	145	88	
PD/SD/Early death	8	3.2	4	2	0.656
Missing	22	8.8	16	10	
Overall follow up (months)					
Median	27		25		0.142
Progression Status					
Not Progressed	194	77.6	120	73	
Progressed	56	22.4	45	27	0.258
Time to progression (months)					
Median	22		21		0.803

5.4.2 Predictive markers of clinical response

The overall response rate; the sum of complete responses (CR) and partial responses (PR) was 88% (n=220/250) (**Table 5.1**). Of these, 22 patients had non-evaluable data meaning that 95% (220/228) of evaluable patients responded with either a CR or a PR. *TP53* mutations were the only significant genetic determinant of clinical response when tested in univariate analyses (p=0.029) and after correction for multiplicity (**Table 5.2**). None of the biological features measured including sex, age or gender were associated with lack of response to treatment. In addition neither IGHV mutational status or deletion or mutation at the 11q locus were associated with a lack of response, with patients achieving at least a PR if not a CR. A trend towards a significant association was seen for 14q deletions with a corrected p value of 0.068 (**Table 5.2**).

Table 5.2 Univariate regression analysis on Clinical Response

	Clinical response		
Variable	Control	Affected	Corrected p value
TP53 mutation	211	16	0.029
IGHV mutation (mut, unmut)	79	122	0.802
ATM deletion	186	41	1.000
Monoallelic 11q deletion	205	22	0.834
ATM ALL	165	62	1.000
Clinical response	NA	NA	NA
SAMHD1 mutation	219	8	0.196
Trisomy 12	191	36	1.000
TP53 deletion	222	5	0.802
Trisomy 19	220	7	1.115
NOTCH1 mutation	197	30	0.829
Trisomy 18	221	6	1.208
Monoallelic *BIRC3* mutation	218	9	1.000
Biallelic *ATM*	210	17	1.000
CLL Drivers	198	29	0.464
2p gain	215	12	0.941
Age (<65, >65)	141	86	1.073
13q deletion	108	119	1.001
8q24 gain	220	7	1.000
BIRC3 mutation	212	15	1.000
MED12 mutation	222	5	1.036
ATM mutation	194	33	1.000
*SF3B1*mutation	179	48	0.862
Gender (F, M)	64	163	1.000
XPO1 mutation	214	13	1.000
Subclones	65	121	1.000
MYD88 mutation	224	3	1.000
6q deletion	214	13	0.921
14q deletion	214	13	0.068

Chi-square test was applied for significance testing and the Bonferroni correction was applied for multiplicity corrections (see Methods).

5.4.3 Predictive markers of minimal residual disease status

MRD monitoring of bone marrow samples was performed three months post-treatment in these clinical trials. 216 patients had MRD results available and 110 out of 216 (51%) patients achieved MRD negativity (**Table 5.1**). MRD positive status is associated with earlier progression than MRD negative status. In the absence of survival data MRD may be used as a surrogate marker of

PFS and OS as outlined in detail in Chapter 1. In this study the correlation of biological and genetic variables with MRD positive status was undertaken. The presence of unmutated IGHV genes was strongly associated with MRD positivity (p<0.001) (**Table 5.3**). As expected, *TP53* mutations (p<0.001) and *TP53* deletions (p=0.042) were strongly associated with MRD positive status. The presence of trisomy 12 (n=35, 14%; p=0.036) was also a significant predictor of MRD positivity. Cases with trisomy 19 revealed a trend towards an association with MRD negative status (n=6, %; p=0.084). Concerning novel CLL drivers tested in this cohort, an association with MRD positivity was seen for *SAMHD1* mutated samples alone (n=8, 3.2%; p=0.029 (**Table 5.3**). *SF3B1* mutated cases and *NOTCH1* mutated cases failed to show any association.

5.4.3.1 MRD and deletion or mutation at 11q locus

11q deletions (n=38, 15%; p=0.007) and the group of 11q deletions alone (includes *ATM* mutation if an 11q deletion also exists, n=23, 9.2%; p=0.001) showed the strongest relationship with MRD positive status of the groups within the 11q locus. The group defined as 11q deletion with or without an *ATM* mutation (n=58, 23.2%) is the largest genetically defined group in our cohort after 13q deleted cases and IGHV mutated/unmutated cases. A significant relationship was also seen here with MRD positivity (p=0.023). Whereas the group defined solely by biallelic 11q lesions (n=17, 6.8%) was not associated with MRD positive status (p=0.164), nor was the group defined as *ATM* mutations (n=32, 12.8%; p=0.587).

Table 5.3 Univariate Cox regression analysis on Minimal Residual disease positive status

Variable	Control	Affected	Corrected p value
TP53 mutation	200	16	<0.000
IGHV mutation (mut, unmut)	77	114	<0.000
11q deletion	178	38	0.007
11q deletion only	193	23	0.010
ATM deletion +/or mutation	158	58	0.023
Clinical response	7	206	0.025
SAMHD1 mutation	209	7	0.029
Trisomy 12*	181	35	0.036
TP53 deletion	210	6	0.042
Trisomy 19*	210	6	0.084
NOTCH1 mutation	188	28	0.152
Trisomy 18	211	5	0.158
Monoallelic *BIRC3* mutation	208	8	0.160
Biallelic *ATM*	199	17	0.164
CLL Drivers	187	29	0.295
2p gain	203	13	0.309
Age (<65, >65)	132	84	0.358
13q deletion	103	113	0.361
8q24 gain	209	7	0.417
BIRC3 mutation	204	12	0.518
MED12 mutation	212	4	0.525
ATM mutation	184	32	0.587
SF3B1 mutation	169	47	0.783
Gender (F, M)	61	155	0.785
XPO1 mutation	205	11	0.853
Subclones	63	115	0.875
MYD88 mutation	213	3	1.000
6q deletion	203	13	1.036
14q deletion	205	11	1.000

*associated with MRD negative status. Chi-square test was applied for significance testing, and the Bonferroni correction was applied for multiplicity corrections (see Methods)

Regarding *BIRC3* mutation status the group of *BIRC3* mutated patients (n=12, 4.8%; p=0.518) and the monoallelic *BIRC3* mutated group (n=8, 3.2%; p=0.16) failed to associate with MRD positive status.

5.4.3.2 Mutlivariate model using MRD as an outcome

In the multivariate testing all significant values from the univariate analysis were tested. Only *TP53* mutations and *SAMHD1* mutations along with clinical response retained significance indicating an independent relationship with MRD positive status (**Table 5.4**). A trend towards a significant association with MRD positivity was seen for unmutated IGHV status and 11q deletion and/or *ATM* mutation. A trend towards MRD negativity for trisomy 12 was suggested in the multivariate model.

Table 5.4 Multivariate Cox Regression Analysis using Minimal Residual Disease on genomic alterations.

Variable	χ2 test	P value
TP53 mutation	8.835	0.003
SAMHD1 mutation	6.764	0.009
Clinical response	5.349	0.021
IGHV mutation	3.055	0.080
Trisomy 12*	2.829	0.093
ATM deletion +/or mutation	2.826	0.093
11q deletion only	1.018	0.313
ATM deletion	0.052	0.819
TP53 deletion	0.000	0.995

*Trisomy 12 associated with MRD negative samples

5.4.4 Survival analysis reveals high-risk treatment response predictors

To identify markers of early relapse a Cox regression analysis for survival was performed using PFS data (**Table 5.5**). The median follow up time at the time of analysis was 27 months. Within the follow up period, 56 patients had progressed with a median time to progression of 22 months.

As expected clinical response and MRD positive status, variables measured after treatment has commenced, associated with shorter time to progression as outlined in **Table 5.5**. The mutation status of the IGHV locus did not impact significantly in this univariate analysis (Figure 5.1).

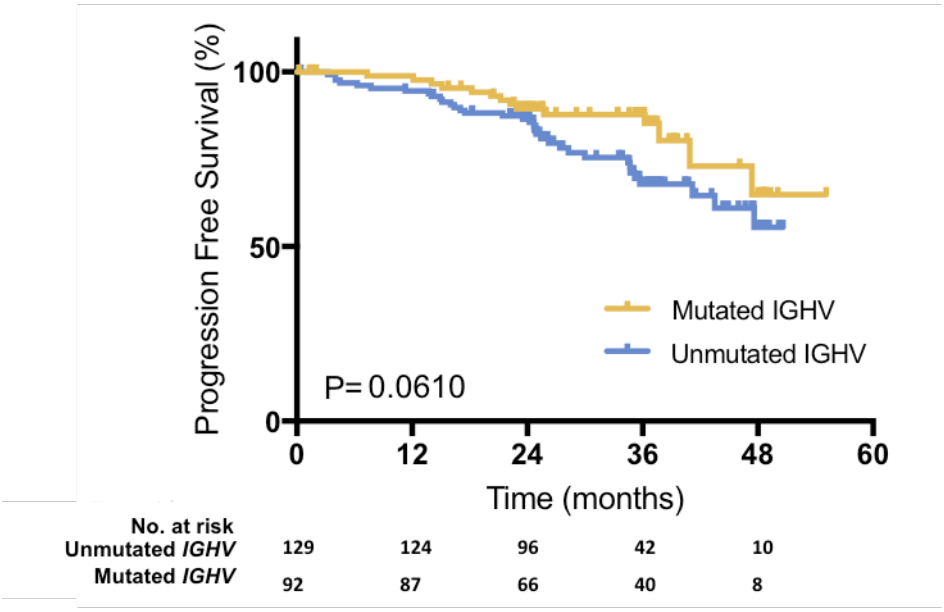


Figure 5.1 Kaplan-Meier estimate of survival for IGHV mutation status
Difference between curves was assessed by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown.

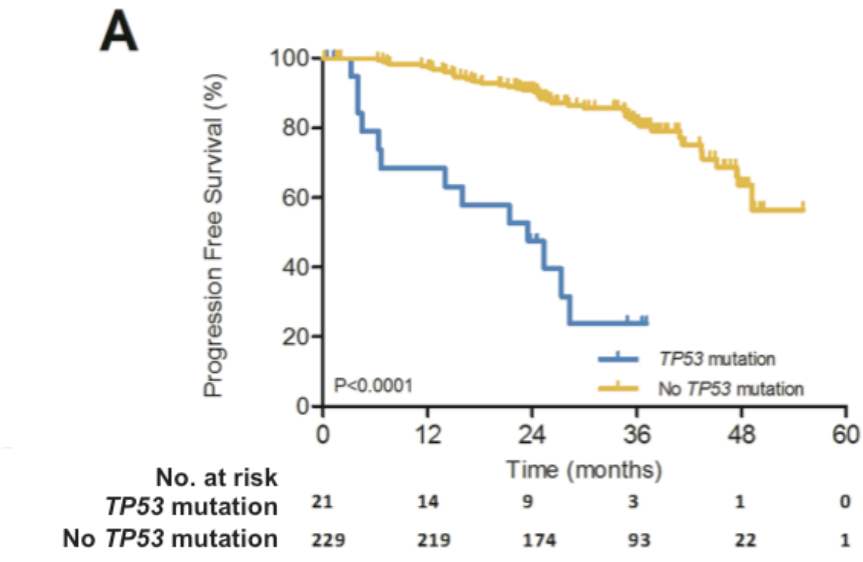
Table 5.5 Univariate Cox regression analysis on Progression Free Survival of all variables (n=250)

Progression Free Survival						
Variable	WT	Affected	Corrected p-value	HR	95% CI	
TP53 mutation	229	21	<0.001	7.719	4.048	14.720
Clinical response	8	219	<0.001	.070	0.030	.159
MRD assessment (Negative=WT)	110	106	<0.001	7.775	3.477	17.385
TP53 deletion	240	10	<0.001	7.404	3.127	17.529
SAMHD1 mutation	241	9	0.001	5.763	2.271	14.627
Chromothripsis	242	8	0.002	4.360	1.870	10.170
CLL Drivers (0-3, 4+)	211	39	0.005	2.793	1.507	5.175
14q deletion	236	14	0.231	2.492	0.985	6.306
IGHV mutation status	92	129	0.243	1.799	.965	3.357
2p gain	237	13	0.316	2.195	.873	5.520
Gender (M=1)	68	182	0.697	1.212	.661	2.220
Trisomy 12	210	40	0.718	1.248	0.629	2.475
ATM mutation	211	39	0.731	.731	0.331	1.615
ATM ALL	190	60	0.748	0.706	0.378	1.318
XPO1 mutation	236	14	0.750	1.392	0.502	3.860
MYD88 mutation	246	4	0.752	.048	00.000	333.123
13q deletion	116	134	0.758	1.368	0.800	2.341
8q24 gain	242	8	0.767	1.759	0.427	7.254
SF3B1 mutation	196	54	0.780	1.243	0.667	2.315
Biallelic *ATM*	230	20	0.788	1.255	0.498	3.162
Trisomy 18	244	6	0.793	0.047	0.000	82.271
BIRC3 mutation	234	16	0.818	0.552	0.134	2.265
Trisomy 19	243	7	0.825	0.047	0.000	46.440
Age (1=>65)	154	96	0.870	1.272	0.746	2.170
Subclones	70	133	0.904	0.761	0.422	1.369
MED12 mutation	244	6	0.962	0.776	0.107	5.641
Monoallelic 11q deletion	227	23	0.979	1.030	0.438	2.420
NOTCH1 mutation	217	33	0.980	0.926	0.418	2.051
6q deletion	236	14	0.993	0.995	0.310	3.190
ATM deletion	205	45	1.000	0.973	0.489	1.934
Monoallelic *BIRC3* mutation	240	10	1.000	0.917	0.223	3.770

5.4.4.1 Individual CLL drivers impact on Early Progression

TP53 mutations had the highest impact on PFS with a median PFS of 23.5 months (HR 7.72 [CI 4.04-14.72], p<0.001) (Figure 5.2 A). Cases with *TP53* deletions were also significantly associated with a shorter PFS (HR 7.4 [CI 3.12-

17.52], p<0.001). *SAMHD1* mutations impacted also on survival with a median PFS of 15 months (HR5.76 [CI 2.27-14.62], p<0.001) (Figure 5.2 B).



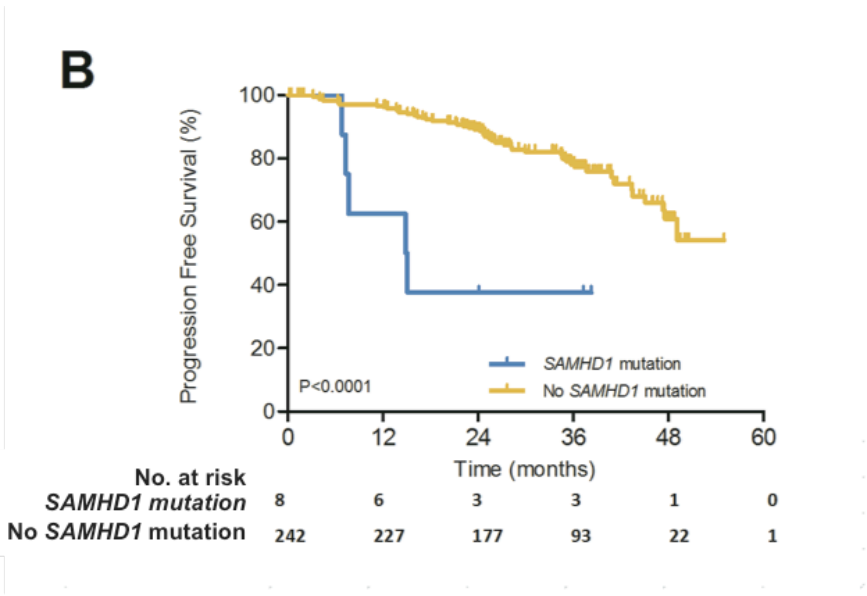


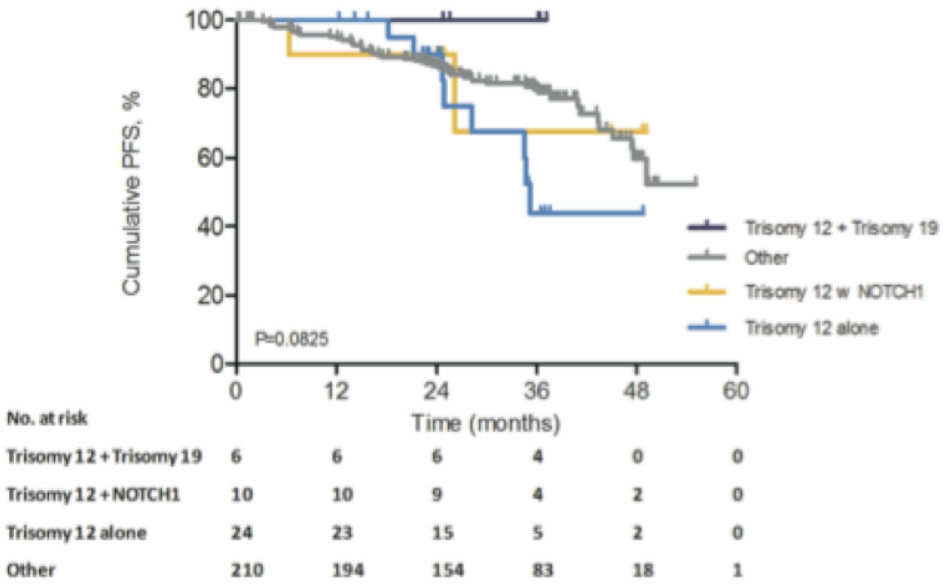
Figure 5.2 Kaplan-Meier estimates of survival for *TP53* mutations and *SAMHD1* mutations

A *TP53* mutation status B *SAMHD1* disruption status. Below the X-axis total numbers of affected cases at each timepoint are shown. The difference between curves was assessed using the Log rank test.

5.4.4.2 Early progression not defined by non-TP53 well known CLL drivers

5.4.4.2.1 Trisomy 12

As a group, trisomy 12 cases were not associated with shorter PFS compared to those without trisomy 12 (**Table 5.5**). The known association of trisomy 12 with trisomy 19 allowed further subgrouping that revealed a trend towards an improved outcome for trisomy 19 defined patients (HR 0.26 [0.061-1.18], p=0.083) (Figure 5.3). In total 40 patients in the cohort had a trisomy 12 detected, six of these also had trisomy 19. Within this group no patients had progressed within the follow up period. The trisomy 12 non-trisomy 19 group contained 10 patients who progressed or died despite 3 of these patients measuring as MRD negative 3 months post treatment. No difference in outcome was seen for patients with trisomy 12 and *NOTCH1* mutation compared to trisomy 12 cases without a *NOTCH1* mutation or compared to the non-trisomy 12 controls (Figure 5.3).



No. at risk						
Trisomy 12 + Trisomy 19	6	6	6	4	0	0
Trisomy 12 + NOTCH1	10	10	9	4	2	0
Trisomy 12 alone	24	23	15	5	2	0
Other	210	194	154	83	18	1

Figure 5.3 Kaplan-Meier estimates of PFS for trisomy 12

Trisomy 12 and trisomy 19, dark blue; trisomy 12 and *NOTCH1* mutated, yellow; trisomy 12 without trisomy 19 or *NOTCH1*, light blue; Other, non-trisomy 12 samples. P value calculated using the Log rank test relates to the difference in PFS between trisomy 12 and 19 group and all other cases with trisomy 12. Below the X-axis total numbers of affected cases at each timepoint are shown.

5.4.4.2.1 11q locus

Despite a significant impact on MRD outcome, no significant difference in PFS was noted for any of the 11q subgroups that were outlined in Chapter 4 Section 4.4.7.2 (**Table 5.5**). In particular the negative impact of biallelic *ATM* disruption seen in other data sets was not seen in this analysis despite an association with MRD positive status outlined earlier in this Chapter Section 5.4.3.1 (Figure 5.4).

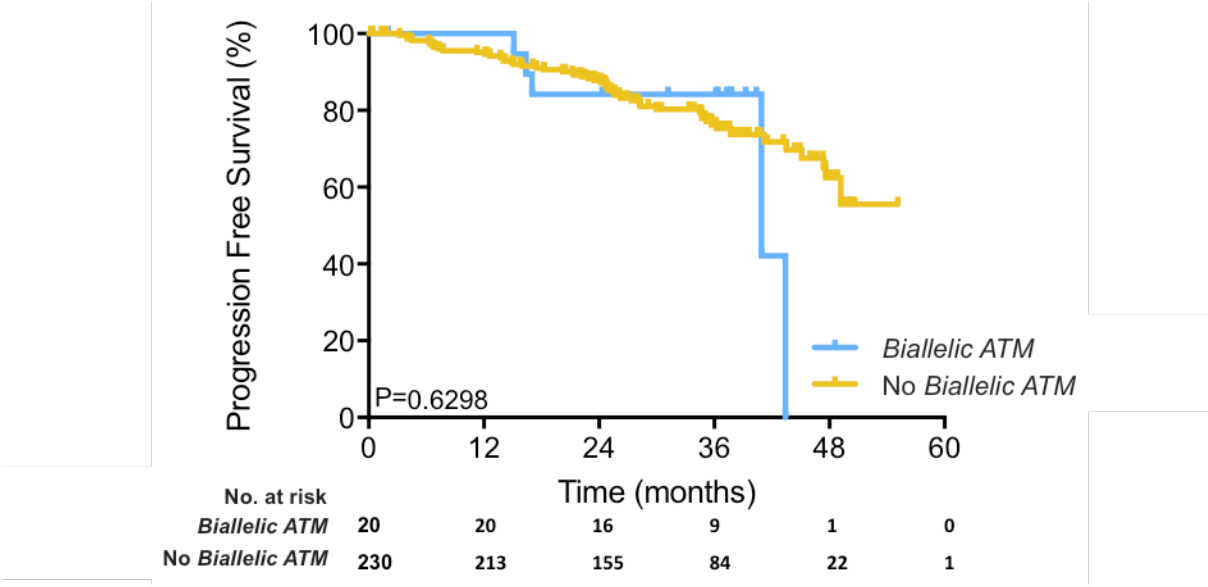


Figure 5.4 Kaplan-Meier estimates of Survival for Biallelic *ATM* lesions

The p-value relates to the overall significance of the findings between multiple curves defined by the Log rank test.

Below the X-axis total numbers of affected cases at each timepoint are shown.

5.1.1.1.1 13q14.2-14.3 locus

Following on from the description of heterogeneity within 13q deleted samples in Chapter 4 Section 4.4.6.4 the influence of these various subgroups of 13q deletions on PFS was evaluated. While a significant finding related to PFS in the 13q heterozygous deleted versus 13q homozygous deleted groups was not seen on standard Log rank testing, weighting for early progressors using the Gehan-Breslow-Wilcoxon test revealed significance with more patients in the 13q heterozygous deleted group progressing early (HR 0.963 [0.339-2.73], p=0.0358) (Figure 5.5). While the median time to progression between the groups was comparable (47.3 versus 49.1 months) it is worth remarking that within the 13q homozygous group not a single patient progressed before 36 months.

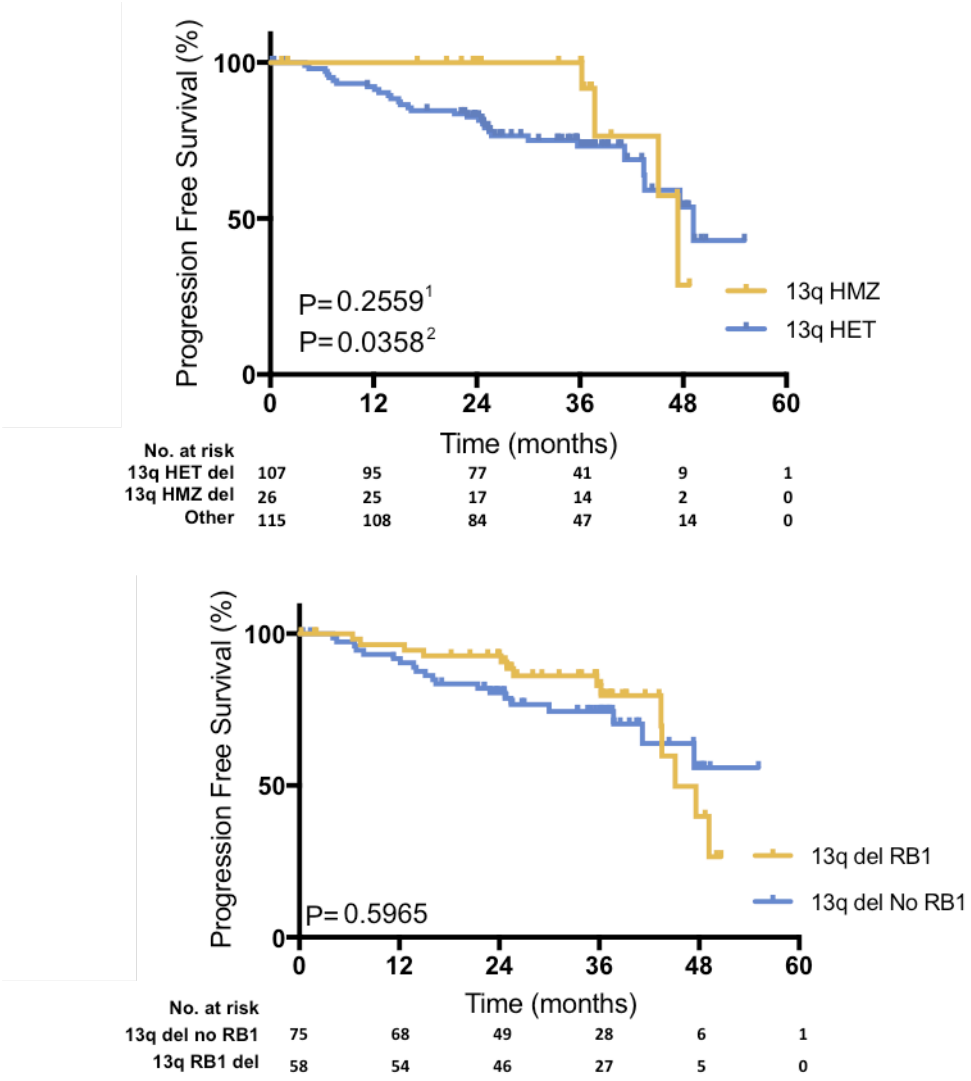


Figure 5.5 Kaplan-Meier estimates of Survival and 13q deletions

A 13q deletions both heterozygous (13q HET) and homozygous (13q HMZ) deletions are shown. The p- value is calculated by two methods: [1]standard Log rank test, [2]the Gehan-Breslow-Wilcoxon test.

B 13q deletions grouped by size with larger 13q deletions defined by deletion of the *RB1* gene.

Below the X-axis total numbers of affected cases at each timepoint are shown.

5.4.4.3 Global genomic measures to predict progression

5.4.4.3.1 Increasing genomic complexity defined by both CNA and cnLOH influences outcome

Genomic complexity defined by the total number of CNAs and cnLOH was evaluated. A linear relationship between increasing number of lesions and shorter PFS was seen. The number of patients defined by the total number of CNAs and cnLOH revealed six groups with between 0 and 5+ lesions. The median time to progression was not reached for 0 CNAs/cnLOH compared to 43 months for 5+ CNAs/cnLOH. The Kaplan-Meier curve demonstrates an overall significant value when comparing curves. The Logrank test for trend was highly significant ($p<0.001$). This indicates a significant linear decrease in PFS with increasing genomic complexity (Figure 5.6).

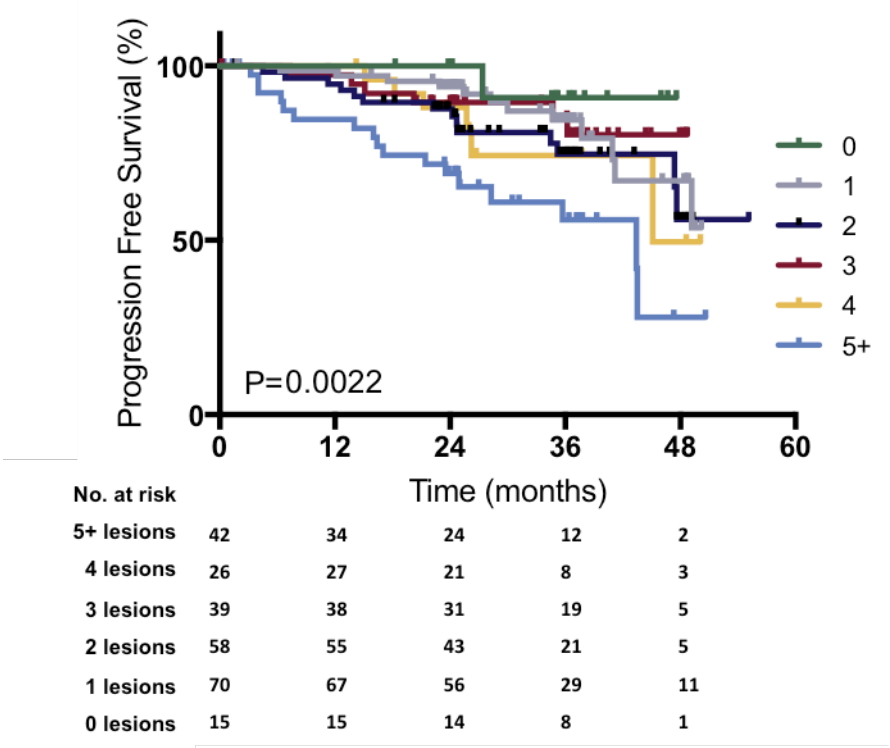


Figure 5.6 Kaplan-Meier estimate of survival based on total number of CNAs and cnLOH per sample

Total number of CNAs and cnLOH are presented and colour coded. The p- value for the overall logrank test is reported. Below the X-axis total numbers of affected cases at each timepoint are shown.

5.4.4.3.1 Chromothripsis and progression

It was noted that many cases with high levels of genomic complexity were cases with Chromothripsis or chromothripsis-like events. These events were present in nine cases where the total number of events per chromosome ranged from five to 29 as detailed in Chapter 4 Section 4.4.6.2.

Evaluating the impact of these findings in a univariate analysis on survival revealed a shorter PFS in these cases compared to cases without chromothripsis (HR4.36 [CI 1.87-10.17], p=0.002) (Figure 5.7).

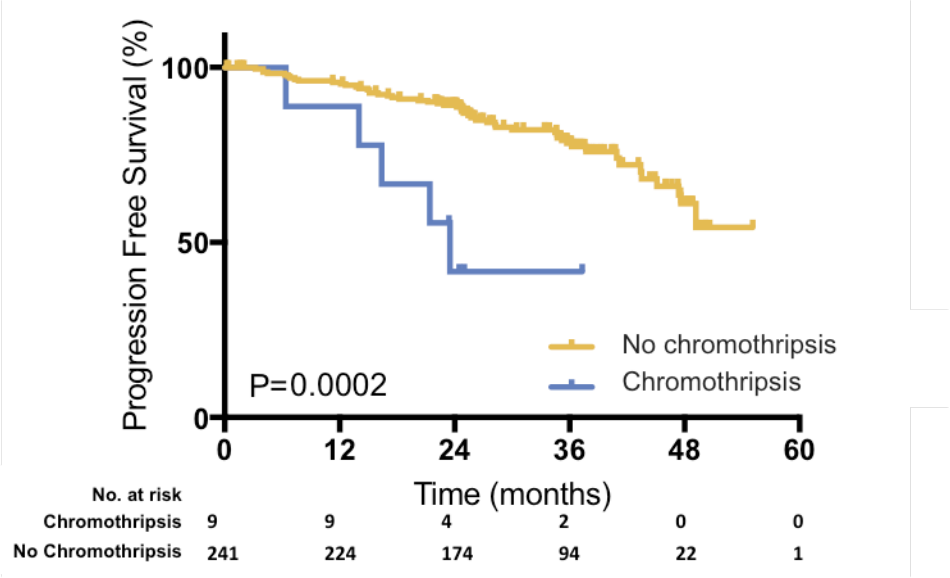
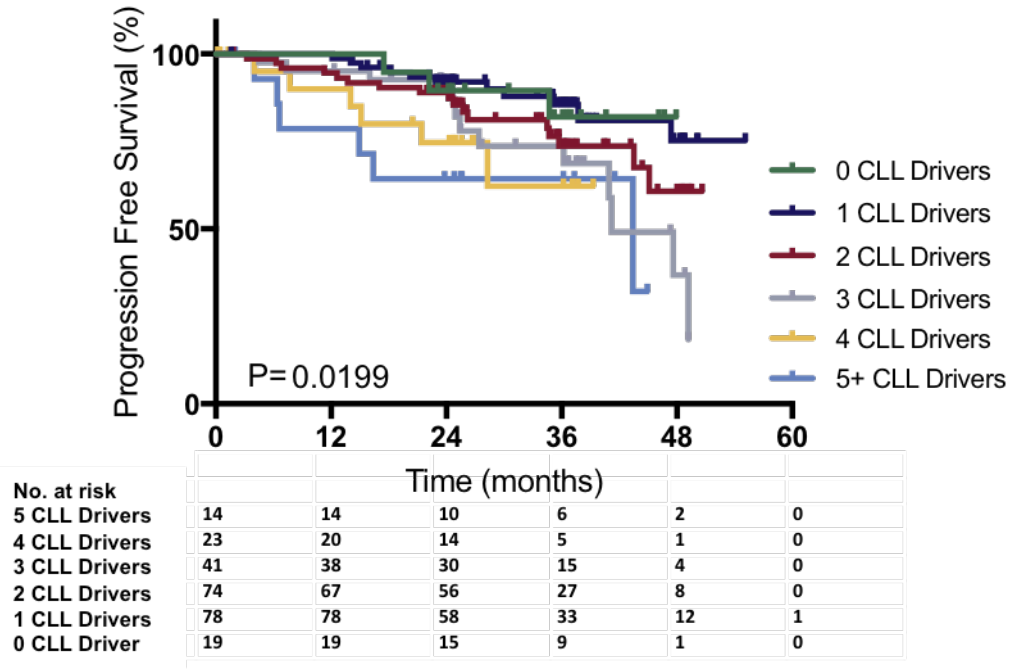


Figure 5.7 Kaplan-Meier estimates of PFS for Chromothripsis

The P value for the overall logrank test is reported. Below the X-axis total numbers of affected cases at each timepoint are shown

5.4.4.3.1 Increasing number of CLL drivers impacts on early progression

A CLL driver lesion was found in 231 of the 250 cases (92.4%) in this series as outlined in Chapter 4 Section 4.4.7. The impact of the number of CLL driver lesions on survival was evaluated.



No. at risk						
5 CLL Drivers	14	14	10	6	2	0
4 CLL Drivers	23	20	14	5	1	0
3 CLL Drivers	41	38	30	15	4	0
2 CLL Drivers	74	67	56	27	8	0
1 CLL Drivers	78	78	58	33	12	1
0 CLL Driver	19	19	15	9	1	0

Figure 5.8 Total number of CLL driver lesions defined by PFS

Kaplan-Meir estimates of survival based on the number of CLL driver events was calaculated per patient. The p-value relates to the overall significance of the findings between multiple curves defined by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown

This revealed a stepwise inverse relationship between the number of CLL driver lesions and PFS (Figure 5.8). A higher number of CLL lesions per sample associated with a shorter PFS. The median progression free survival was not reached for cases without a CLL driver or with one or two CLL drivers present compared to 43 months for cases with five or more CLL drivers. The log rank

test for trend was also highly significant (p=0.0004) which reflects a linear association between each increasing number of CLL drivers and progression.

The total number of CLL drivers were combined into binomial variables and assessed based on PFS. The greatest significance was seen between CLL drivers grouped into 0-3 and ≥4 CLL drivers (HR 7.66 [CI 3.027-19.41], p<0.001) (Figure 5.9 A-D). After correction for multiplicity testing the presence of ≥4 CLL drivers remained significant in survival analysis (**Table 5.5**).

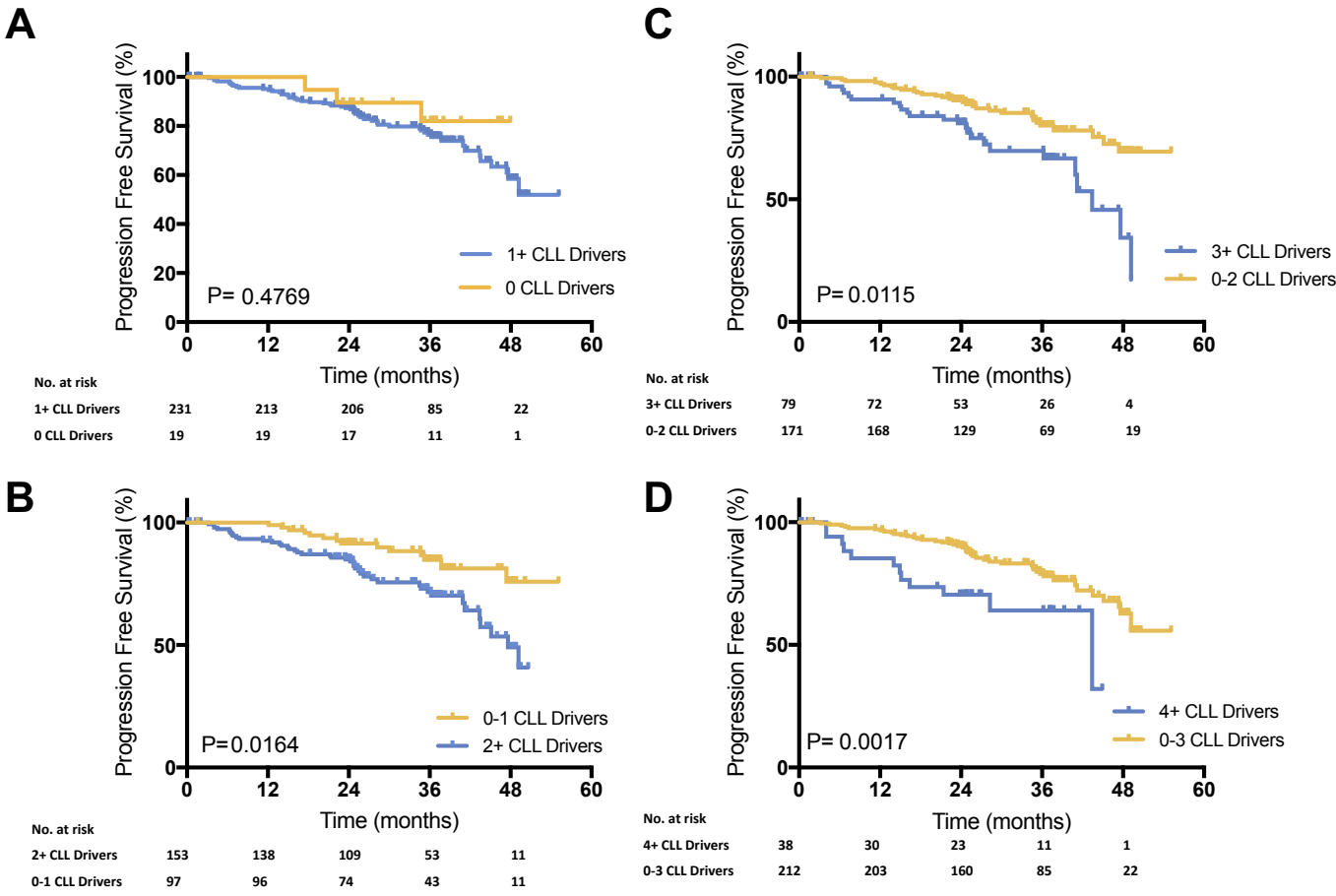


Figure 5.9 Kaplan Meier estimates of PFS based on low and high numbers of CLL driver lesions

A-D: CLL driver lesions grouped with the yellow line representing the lower number of lesions and the blue line the higher number. The P value relates to the overall significance of the findings between multiple curves defined by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown

5.4.4.3.1 Subclonal gene analysis reveals TP53 alone impacts on survival analysis

First, testing the impact on survival of subclonal mutations (VAF>0.05, <0.45) for genes *SF3B1, NOTCH1*, *ATM* and *TP53*. The number of mutations in the remaining genes was too low to further divide into subgroups. For this per gene analysis, with the exception of *TP53* mutations, the presence of a subclonal mutation did not impact significantly on early progression (Figure 5.10 A-D).

While the Kaplan-Meier survival curve highlighted that clonal *NOTCH1* mutations seemed to progress earlier this was not statistically significant (Figure 5.10 B).

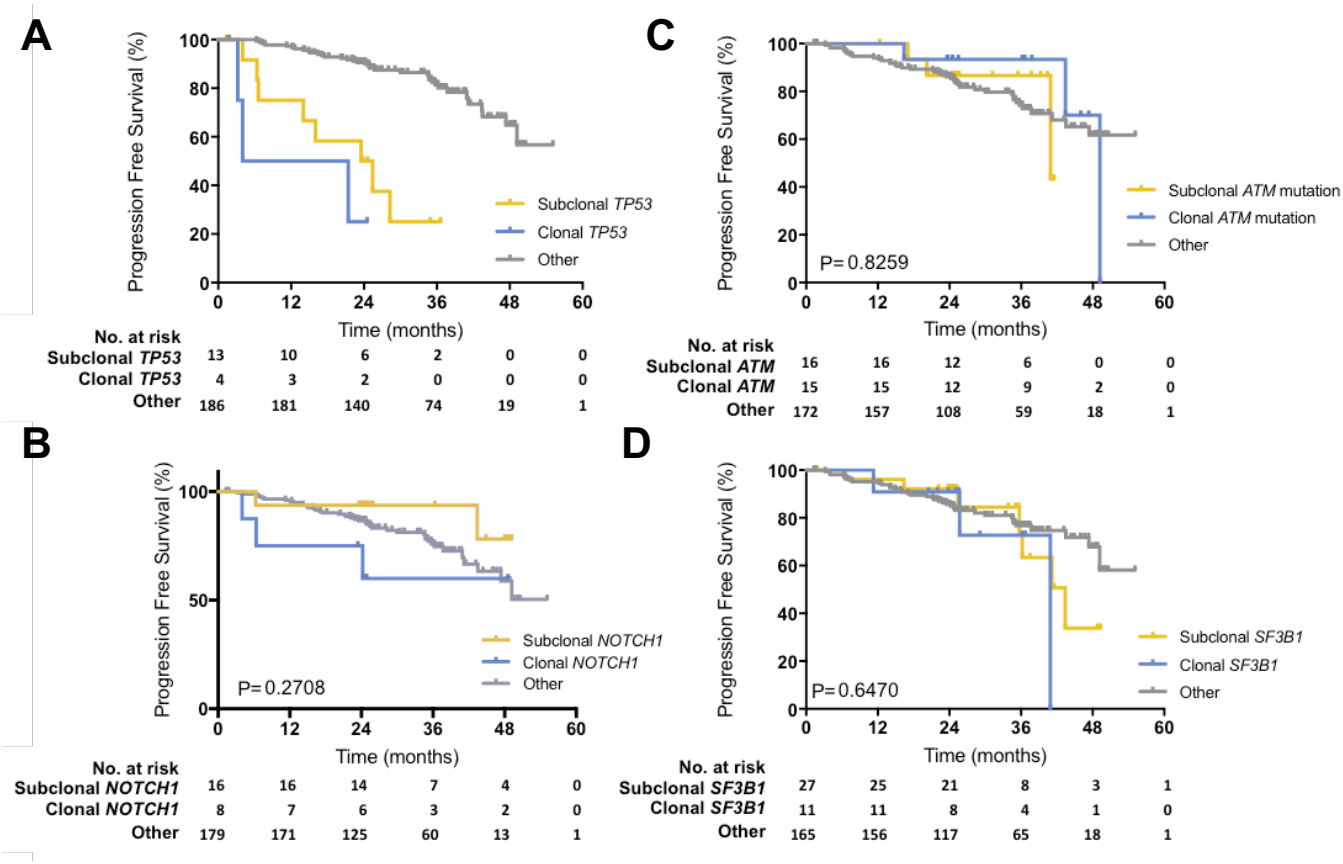


Figure 5.10 Kaplan-Meier estimate of survival based on presence of clonal or subclonal mutations

A-D: This per gene analysis outlines subclonal mutations (VAF>0.05,0.45, yellow line) and clonal mutations (VAF>0.45,blue line) for genes with a mutation frequency of >0.10 in this cohort. The P value relates to the overall significance of the findings between multiple curves defined by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown.

5.4.4.3.1 Minor TP53 subclones and outcome

Taking this one step further, the deeper analysis of *TP53* mutations including VAFs of <0.05 revealed eleven additional *TP53* mutations in eight patients as described in Chapter 4 Section 4.4.5.7. Two of these cases also carried a *TP53* mutation with a VAF>5% leaving six patients with a newly identified low level *TP53* mutation. When this additional group of six cases was included in the survival analysis the statistical power to predict early relapse did not improve for the *TP53* mutation defined group and the hazard ratio decreased (Figure 5.11). None of the six patients had progressed within the follow up period with a median follow up time for these six cases of 35.4 months. As the cut off for our overall TSCA sequencing panel was set at a VAF of 0.05 (Chapter 2 Section 2.16.2.4) and the accepted sensitivity of Sanger sequencing corresponds to a VAF of 0.10 the PFS for all *TP53* mutations of >0.05 and >0.10 was analysed separately. Of the three cases carrying *TP53* mutations with a VAF of 0.05-0.10 two of these cases also had at least one further *TP53* mutation of >0.10 VAF. The one remaining case defined by a *TP53* mutation with a VAF 0.05-0.10, ADM207, had a mutation with a VAF of 0.092 and this patient progressed 25 months after treatment. As a minimal difference in survival curves with comparable hazard rations (7.19 versus 7.0) was seen between VAF cut offs of 0.05 and 0.10 respectively, the data analysed in the multivariate analysis included all mutations from the TSCA panel with a VAF of ≥0.05.

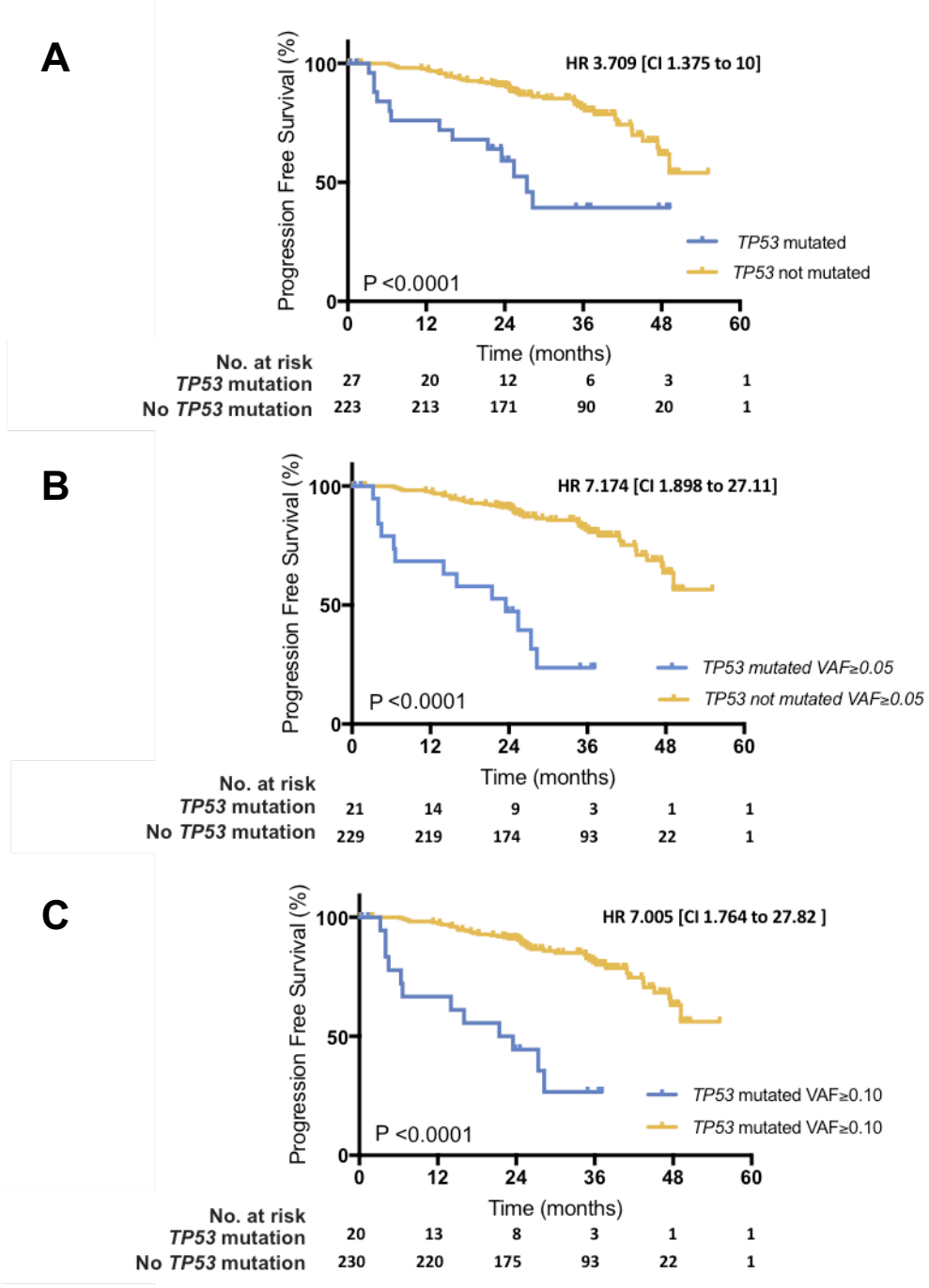


Figure 5.11 Kaplan-Meier estimates of PFS on various VAF subgroups of *TP53* mutations

A-C: depicts PFS on different groups of TP53 mutations with a VAF of A) ≥0.01, B) ≥0.05, C) ≥0.10. The p- value relates to the overall significance of the findings between multiple curves defined by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown. HR: hazard ratio, CI: confidence interval

5.4.4.3.1 Integrated subclonal analysis of CLL drivers

Combining SNV, indel, copy number and cnLOH data in an integrated analysis based on the presence or absence of subclonal lesions was possible in 203 of 250 samples as outlined in Chapter 2 Section 2.21. The presence of subclonal lesions was found in 133 of 203 samples and did not associate with a shorter PFS in the follow-up period of 27 months (HR 0.761 [CI 0.422-1.369, p=0.904) (Figure 5.12).

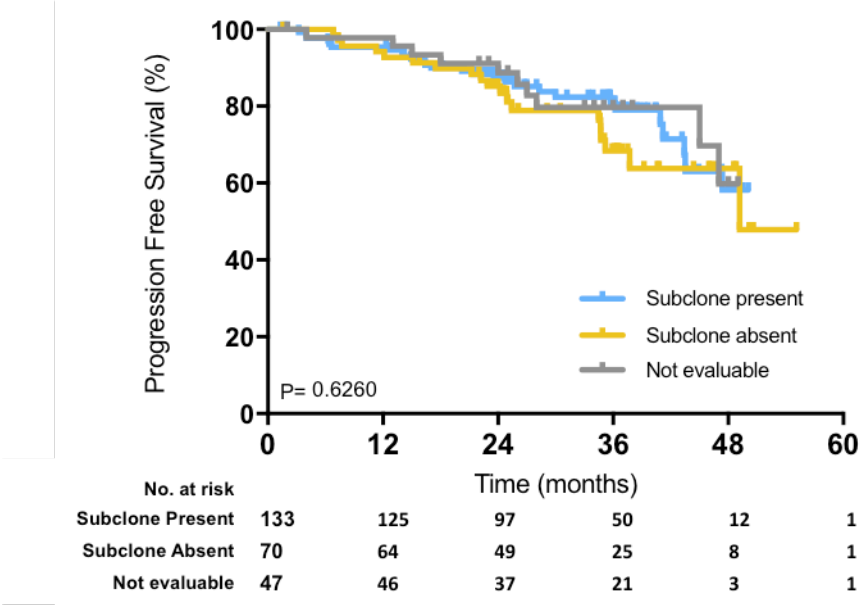


Figure 5.12 Kaplan-Meier estimates of PFS based on subclonal or clonal mutation

Cohort made up of 250 cases: subclone present 133, subclone absent 70, not evaluable 47 described in Methods Chapter 2 Section 2.21. The P value relates to the overall significance of the findings between multiple curves defined by the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown

5.4.5 Mutations alone, not CLL driver burden, independently predict for early progression

TP53 and *SAMHD1* mutations were the strongest independent predictors of shorter PFS in multivariate survival analysis testing all variables (**Table 5.6**). The trisomy 12 without trisomy 19 subgroup remained independently predictive of shorter PFS (p=0.018, HR 2.396 CI 1.162-4.941). In univariate analyses *TP53* and *SAMHD1* mutations, *TP53* deletions, MRD positivity and ≥4CLL drivers were significant variables (**Table 5.5**). Testing only these lesions in a multivariate model revealed *TP53* and *SAMHD1* mutations as the sole independent predictors of early progression (**Table 5.6**, Figure 5.13).

Table 5.6 Multivariate Cox Regression Survival Analysis

	Variable	P value	HR	95% CI	
All CLL drivers	*TP53* mutation	<0.001	10.517	5.291	20.904
	SAMHD1 mutation	<0.001	9.426	3.572	24.870
	Trisomy 12 without trisomy 19	0.018	2.396	1.162	4.941
Significant variables from univariate analysis	*TP53* mutation	<0.001	7.483	3.170	17.665
	TP53 deletion	0.822	1.138	0.369	3.509
	SAMHD1 mutation	<0.001	7.349	2.747	19.660
	4 or more CLL Drivers	0.497	1.286	0.623	2.656

Final Cox Model of progression free survival in multivariate analysis on genomic alterations on all CLL drivers, and on significant variables from the univariate analysis
HR = Hazard Ratio; CI = Confidence Interval.

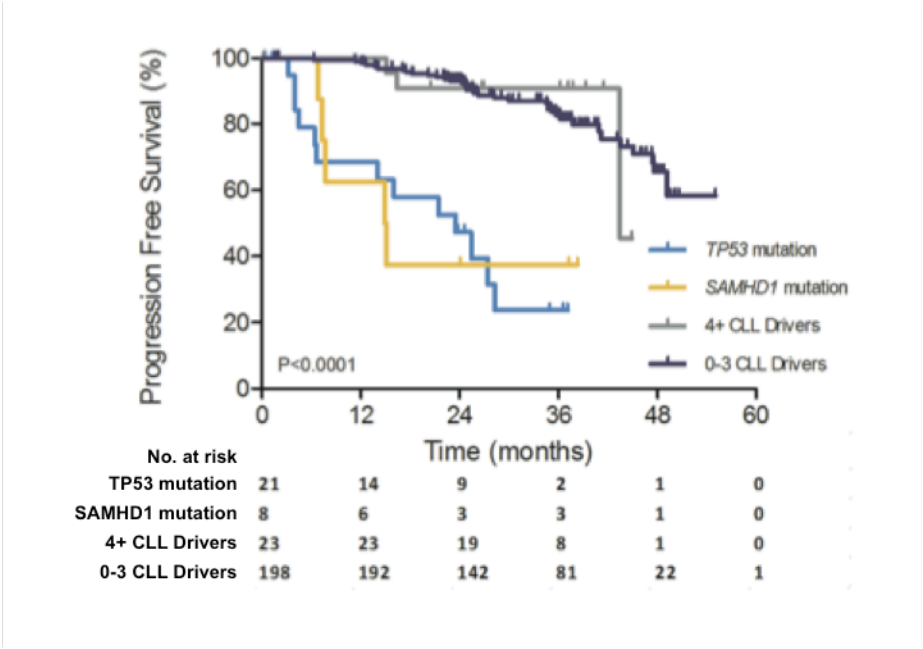


Figure 5.13 Kaplan-Meier estimates of Survival based on high risk genetic variables and CLL drivers

TP53 and *SAMHD1* mutated samples were first selected and analysed for survival with the remaining samples grouped into low and high CLL driver number to reflect the findings from the multivariate analysis. The p-value was calculated using the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown.

5.4.6 Statistical validation of the model

To validate the findings statistically, a penalised Cox proportional hazards regression approach was applied by Dr. Dimitris Vavoulis, MHDS, Oxford, confirming the association of *TP53* and *SAMHD1* mutations with shorter PFS. Features more strongly associated with progression have higher inclusion probability and they are more likely to appear in the fitted model, i.e. they are more stable. Full methodology is outlined in Chapter 2 Section 2.28.2.

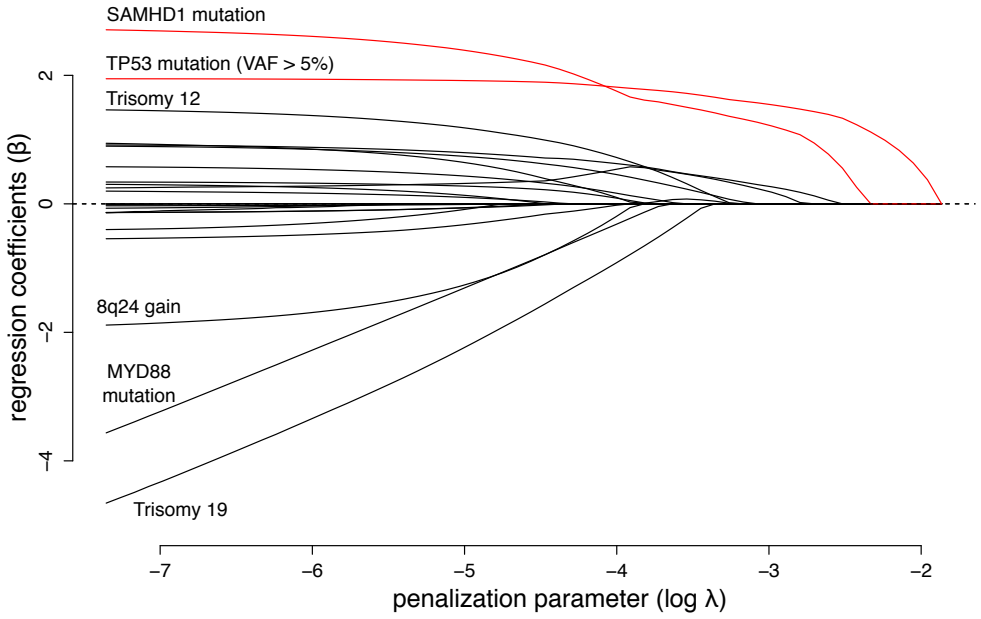
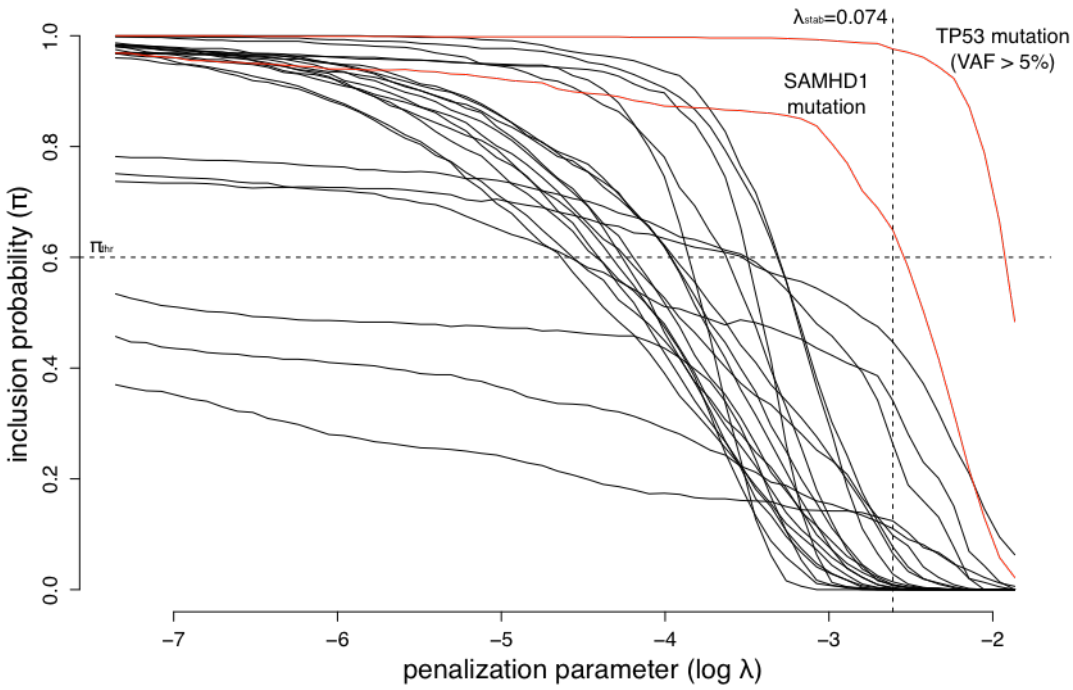


Figure 5.14 LASSO regularization path for the Cox proportional hazards model

Increasing values of the penalty parameter λ, lead to a growing number of genomic features being excluded from the model until the point where no features remain. This figure illustrates the path of each coefficient in the model towards zero as a function of λ. Notice that the coefficients for *SAMHD1*, *TP53* and trisomy 12 are positive, while the coefficients for trisomy 19, *MYD88* and 8q24 gain are negative for a wide range of values of λ. Positive (negative) coefficients indicate a hazard increase (decrease) in the presence of the corresponding genomic feature. *SAMHD1* mutation and *TP53* mutation (VAF > 5%) are indicated in red because they are stable features.

The inclusion probability for each genomic feature, a measure of each association with PFS, was estimated using a sub-sampling methodology. *TP53* and *SAMHD1* mutations were the only stable features in the model above the inclusion probability level (Figure 5.14). Both of these genomic features are associated with a decrease in PFS.

A


λstab=0.074
TP53 mutation
(VAF > 5%)
SAMHD1
mutation
πthr
inclusion probability (π)
0.0
0.2
0.4
0.6
0.8
1.0
−7
−6
−5
−4
−3
−2
penalization parameter (log λ)

B

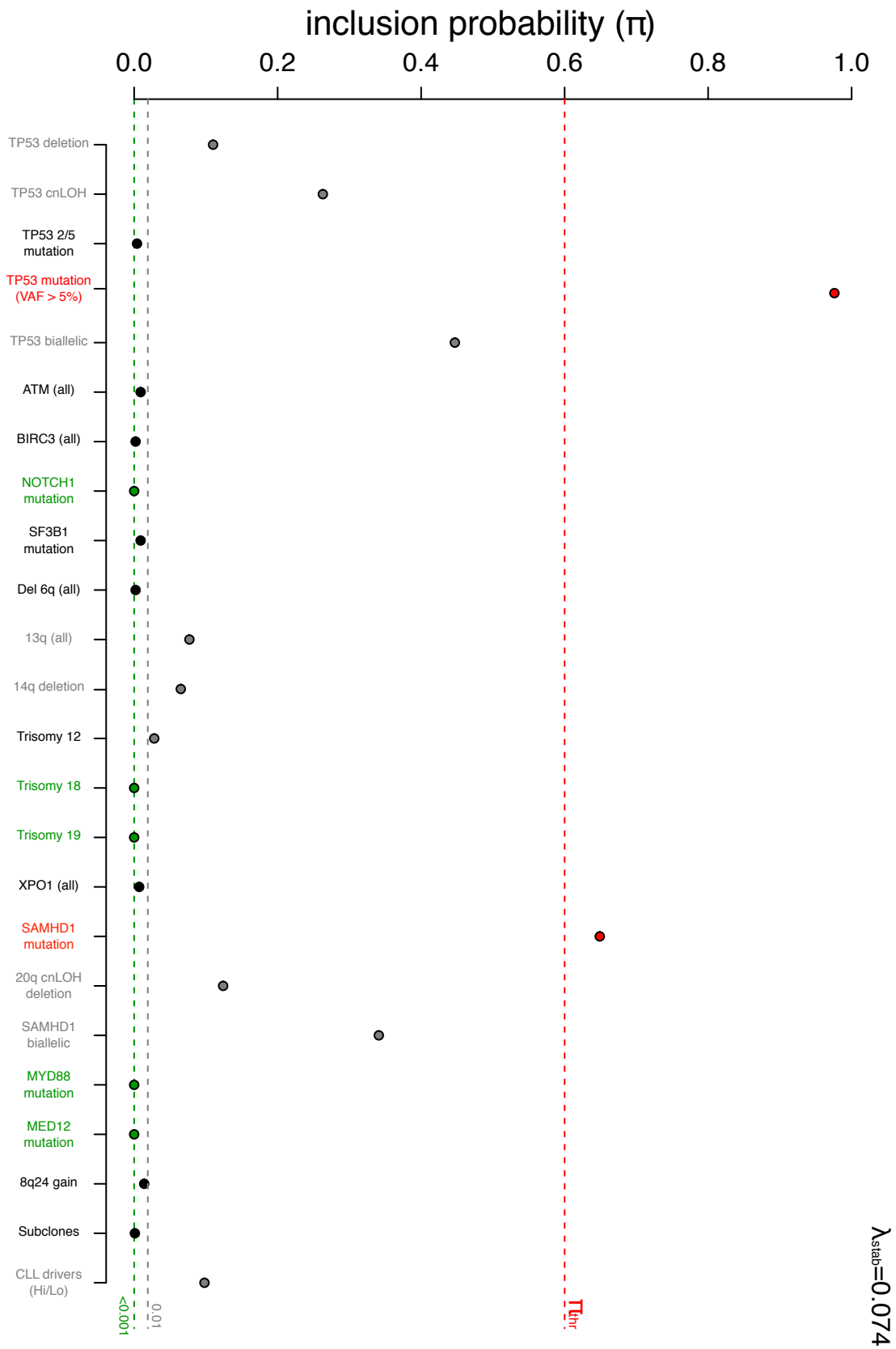

inclusion probability (π)
0.0
0.2
0.4
0.6
0.8
1.0
TP53 deletion
TP53 cnLOH
TP53 2/5 mutation
TP53 mutation (VAF > 5%)
TP53 biallelic
ATM (all)
BIRC3 (all)
NOTCH1 mutation
SF3B1 mutation
Del 6q (all)
13q (all)
14q deletion
Trisomy 12
Trisomy 18
Trisomy 19
XPO1 (all)
SAMHD1 mutation
20q cnLOH deletion
SAMHD1 biallelic
MYD88 mutation
MED12 mutation
8q24 gain
Subclones
CLL drivers (Hi/Lo)
<0.001
0.01
λstab=0.074

Figure 5.15 Stability and Ranking of Genomic Variables from Cox model

(A) Stability path of the Cox proportional hazards model. This figure illustrates the estimated inclusion probabilities as functions of the penalty parameter λ. For a per-comparison-wise error rate PCER=1 (i.e. at most one error) and an inclusion probability threshold π_{thr}=0.6, two stable features were selected: *TP53* and *SAMHD1*. The penalty parameter at the selection boundary is λ_{stab}=0.074.
(B) Ranking of genomic features based on their inclusion probabilities. Inclusion probabilities were calculated at λ_{stab}=0.074 for a per-comparison-wise error rate PCER=1, this threshold of λ_{stab} is illustrated by the vertical dotted line.

In Figure 5.15 A the analysis highlights two features in red; *TP53* and *SAMHD1* mutations, that are stable reflecting a strong association with a short PFS. A number of features below the defined threshold (π_{thr}=0.6, dotted line) have intermediate inclusion probabilities and have a weak association with shorter PFS (Figure 5.15; grey). These include *TP53* deletions, biallelic *ATM* events and 13q deletions. Finally, there are features, such as trisomy 18 and 19 and mutations in *MYD88* and *MED12* that have negligible inclusion probabilities and therefore do not appear associated with PFS (Figure 5.15; green). While trisomy 12 without trisomy 19 was a poor prognostic marker, this finding was not confirmed on bootstrapping analyses.

Furthermore, an optimal bootstrapping methodology was applied to confirm the improved predictive value of a model that includes *TP53* and *SAMHD1* mutations as predictors of PFS (Figure 5.16)

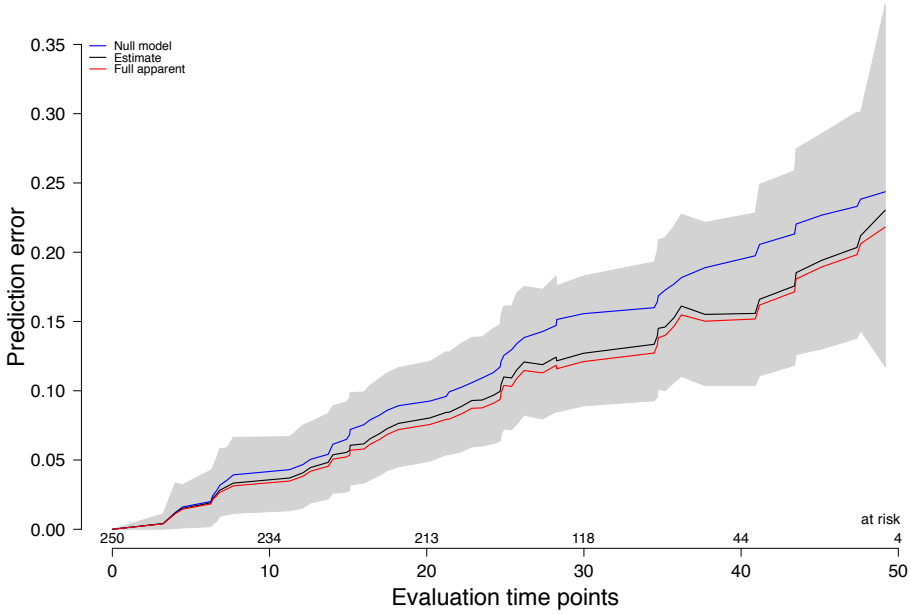


Figure 5.16 Prediction error of the stable Cox proportional hazards model

Prediction error of the stable Cox proportional hazards model. The predictive value of a Cox model including stable genomic features only (*TP53* mutation VAF>5% and *SAMHD1* mutation) was assessed using the time-dependent Brier score and 10-fold cross-validation. The blue line (null model) corresponds to the null model (i.e. the model that includes no genetic markers) tested against the complete dataset. The red line (full apparent) corresponds to the stable model tested against the complete dataset. Since the same data was used to fit the model and assess its performance, the red line represents a lower boundary to the prediction error due to overfitting. The black line is the .632 estimator by Efron and Tibshirani (1997), which is an average between the full apparent model and the stable model tested through cross-validation. The grey region indicates the error associated with the black line. The graph indicates that inclusion of the two stable covariates leads to reduction of the prediction error in relation to the null model.

5.4.7 Validation of high-risk genomics in historical clinical trial

To further validate *SAMHD1* mutations as markers of early progression following chemotherapy, an independent clinical trial cohort was evaluated for *TP53* and *SAMHD1* mutation frequencies (**Table 5.7**). This cohort comprised of 247 patients with a median age of 65 years. This was also a frontline treatment Phase III trial, the LRF CLL4 UK trial, with further details in Chapter 2 Section 2.1. 10-year follow up data was available for this clinical trial with a median time to progression of 24 months. *SAMHD1* mutations were confirmed to be associated significantly with PFS, with a median PFS for *SAMHD1* mutated patients of 5.7 months compared to 18.6 months (p=0.014, HR 2.913; 95% CI 1.234-6.879) (Figure 5.17).

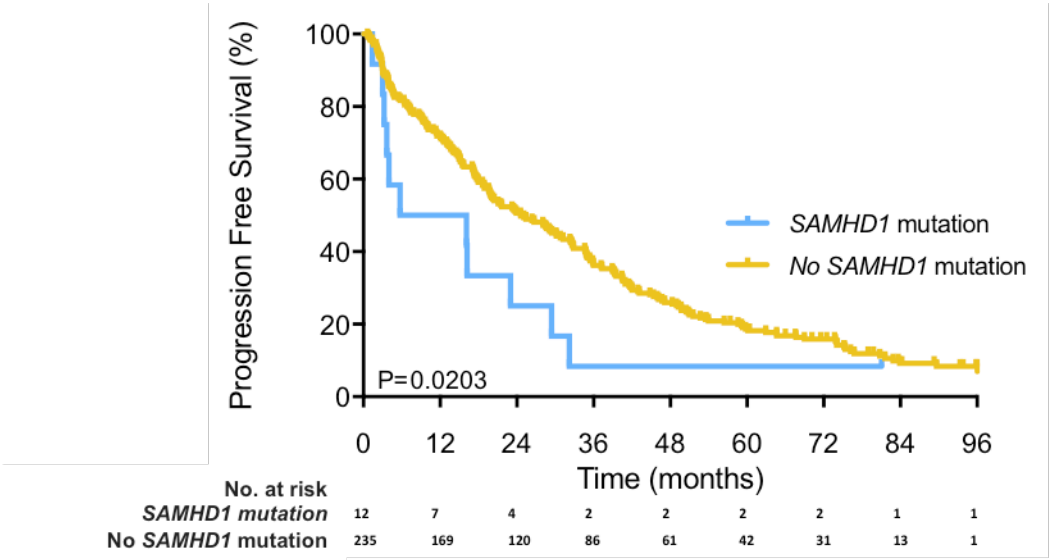


Figure 5.17 Kaplan-Meier estimates of PFS for *SAMHD1* mutation in CLL4 trial cohort

P-value was calculated using the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown.

Table 5.7 Patient characteristics of LRF CLL4 trial

Characteristic	Patients (N=250) No.	%
Age		
Median		65
Range		42-85
WBC		
Median		103
Range		4.9-605
Lymphocyte Count		
Median		94.6
Range		0.5-568
Haemoglobin		
Mean		12.4
Range		3.8-16.2
Platelets		
Median		146
Range		2-684
Sex		
M	189	76.5
F	58	23.5
Binet stage		
A progressive	56	22.6
B or C	191	77.3
Treatment arm		
Chlorambucil	118	47.8
Fludarabine	39	36.4
FC	90	15.8
IGHV mutation status		
Mutated	80	32.4
Umutated	143	57.9
N/A	24	9.7
FISH Del17p		
Del17p	15	6.1
Non del17p	206	83.4
Not available	26	10.5
FISH Del11q		
Del11q	47	19
Non Del11q	174	70.4
Missing	26	10.5
CD38		
Positive	98	39.7
Negative	110	44.5
N/A	39	15.8
Final Response		
CR/PR	185	74.9
PD/SD/Early Death	54	21.9
N/A	8	3.2
Follow up to 1st PFS assessment (months)		
Median		24
Range		1-112
Time to progression (months) N=221		
Median		20
Range		1 - 96

The previously published shorter PFS of *TP53* mutated patients in the setting of chemotherapy was also confirmed by MPS (Figure 5.18). *TP53* and *SAMHD1* mutated samples were mutually exclusive consistent with the earlier findings in the FCR treated patients reported in Chapter 4 Section 4.4.7.1. The persistent negative impact of *SAMHD1* mutations when *TP53* mutations are taken into account is shown in Figure 5.18.

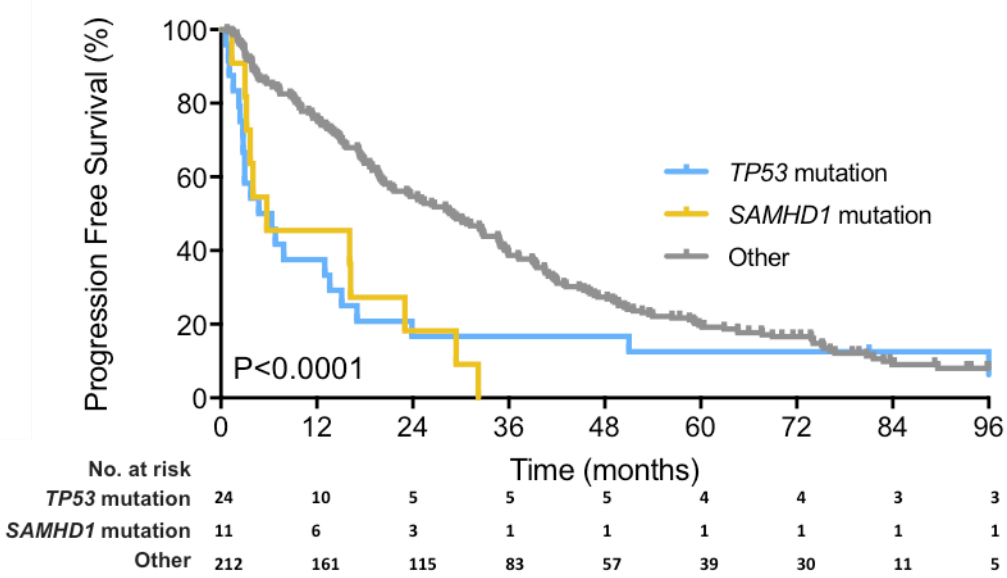
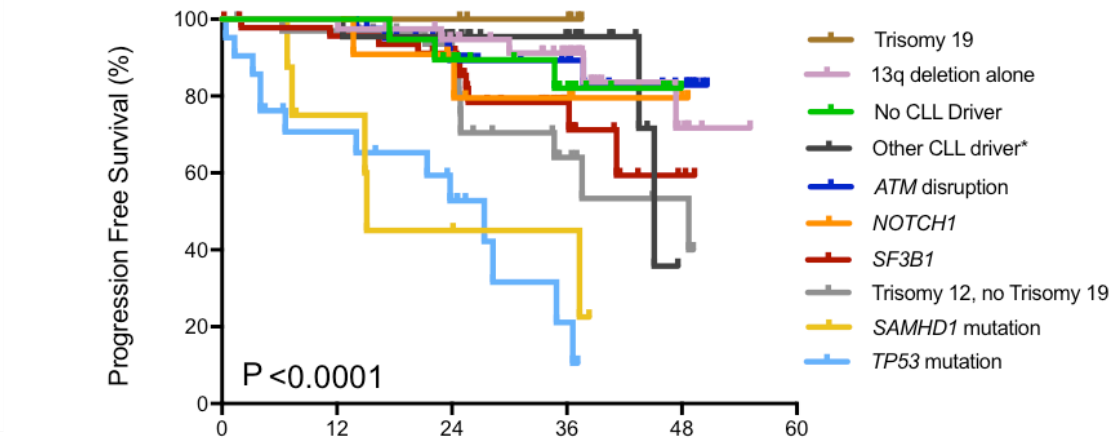


Figure 5.18 Kaplan-Meier estimates of progression free survival in LRF CLL4 clinical trial

CLL4 cohort validation of *SAMHD1* and *TP53* mutations in 247 patients; hiierarchy of firstly *TP53* mutation status, then *SAMHD1* mutation status against wild-type for both *TP53* and *SAMHD1* mutations.

5.4.8 Hierarchical model

Finally, the results of the multivariable analysis on the 250 cases in the test cohort were included in a hierarchical survival curve illustrating genetic lesions with greatest impact on PFS (Figure 5.19).



No. at risk						
TP53 mutation	21	14	9	2	1	0
SAMHD1 mutation	8	6	3	3	1	0
Trisomy 19	7	7	7	5	1	0
Trisomy 12	34	32	24	8	4	0
SF3B1 mutation	48	45	37	13	3	0
ATM disruption	41	41	30	16	8	0
Other CLL Driver	22	22	12	10	1	0
NOTCH1 mutation	11	11	9	6	3	0
13q sole deletion	39	39	31	21	5	1
No CLL Driver	19	19	16	10	1	0

Figure 5.19 Kaplan-Meier estimates of Survival ranked by genomic variable

Variables in the hierarchy were ordered using stepwise backward conditional Cox regression modellin. For example as *TP53* remained in the final multivariate model with the highest hazard ratio this group of patients were first selected from the 250 sample. Of the remaining patients the next genetic lesion remaining in the multivariate model with the highest hazard ratio was *SAMHD1*, then trisomy 12 non-trisomy 19 and so on. Each sample is counted one time only, with no overlap. high CLL driver number to reflect the findings from the multivariate analysis. P value was calculated using the Log rank test. Below the X-axis total numbers of affected cases at each timepoint are shown.

The clear separation of both *TP53* and *SAMHD1* mutated groups from the remaining group is highlighted. In addition the very good risk group defined by the trisomy 19 lesion is seen in brown without a patient progressing. Similarly the 13q deletion alone in pink is clearly visible as a slowly decreasing curve. An intermediate group, comprising the trisomy 12 non-trisomy 19 group (grey curve) and *SF3B1* mutated groups (red curve), may also be seen, but was not statistically proven within the median follow up time available.

5.5 Discussion

A comprehensive study from frontline UK clinical trials has been presented in this Chapter. The relationship between CLL drivers and early progression after CIT is outlined, both regarding individual genomic lesions and global measures. The findings presented are independent of IGHV mutation status and can be determined prior to commencing treatment.

To link genetic variables with clinical outcome measurements of response or MRD measurements poses more difficulty than survival analysis that use continuous variables. A further challenge is the low frequency of many of the individual CLL drivers. As the majority of patients in this study had a clinical response to CIT with an overall response rate of 95%, the only informative genetic marker was *TP53* disruption and this associated with a lack of response to CIT. Patients with a *TP53* deletion or mutation would not currently be offered FCR treatment due to the now accepted chemo-refractory status that *TP53* disruption defines (Zenz et al. 2010; Stilgenbauer et al. 2014; Gonzalez et al. 2011; Malcikova et al. 2009; Pospisilova et al. 2012) and in tandem due to the development of successful *TP53* independent small molecule therapies (Byrd et al. 2014; Furman et al. 2014; Stilgenbauer et al. 2016). During the enrolment period for both ARCTIC and ADMIRE trials, other treatment options were not yet available for this group of patients and *TP53* disruption was not an exclusion criteria for trial entry. Therefore, whilst this data confirms the poor risk profile of this genetic subgroup, in the future CIT would not be appropriate for this subgroup (Hallek 2017a).

In this Chapter, I have conclusively shown that *SAMHD1* mutations are independent strong predictors of shorter PFS in CLL. *SAMHD1* is recurrently mutated in 1-5% of CLL in independent cohorts and at different disease stages (Landau et al. 2013; Landau et al. 2015; Guièze et al. 2015). The main limitation of this work is the due to the infrequent occurrence of SAMHD1 mutations in the frontline treatment setting leading to low numbers available for survival analyses and decreasing the robustness of these tests. Interestingly, in relapsed/refractory CLL the mutation frequency was 13% (Guièze, Robbe, Clifford, De Guibert, Pereira, Timbs, M.-S. S. Dilhuydy, et al. 2015). Further confirmation of the importance of *SAMHD1* in CLL can be taken from two notable international studies that have also found an enrichment of *SAMHD1* mutations at relapse,. Results from the German CLL8 trial revealed that six of 59 (10%) patients harboured *SAMHD1* mutations at relapse (Landau et al. 2015). Additionally, of 61 relapsed CLL cases extensively studied using WES and SNP array, significance for enrichment of only *TP53* and *SAMHD1* mutations was found (Amin et al. 2016). The negative impact of *SAMHD1* mutations on PFS in a frontline clinical trial cohort of CIT treated patients is defined here. This high-risk genetic lesion also revealed early progression in the independent UKCLL4 cohort.

Like *TP53* and *SAMHD1*, *ATM* is involved in DNA damage response and the presence of 11q deletions or bi-allelic *ATM* disruption impacted on survival outcomes in the pre-rituximab era (Thorsten Zenz, Eichhorst, et al. 2010; Zenz

et al. 2008; Jitka Malcikova et al. 2009; Gonzalez, Martinez, Wade, Hockley, Oscier, Matutes, Claire E. Dearden, et al. 2011; Shiloh 2003; Stankovic 2004). Patients with 11q deletions benefit particularly from the addition of rituximab to the chemotherapy backbone (Stephan Stilgenbauer et al. 2014). Consistent with this observation, disruption of *ATM* was not associated with a shortened PFS in our cohort, despite a clear link between 11q deletions and/or *ATM* mutations and MRD positivity seen here (Döhner et al. 2000; Austen et al. 2005; Skowronska, Parker, et al. 2012; Skowronska, Austen, et al. 2012). Longer follow-up data will be needed to confirm the relapse rate of this subgroup of patients. However, we know that MRD positive patients will relapse and require further treatment based on a recent study of 133 patients requiring treatment for CLL where the PFS was 3.3 years for MRD levels >0.01% versus 7.6 years for MRD negative patients (Kwok et al. 2016). Therefore, an intermediate risk, rather than the previously denoted high-risk assignment for del(11q) or biallelic *ATM* disruption seems to be more appropriate in the setting of CIT.

Very low level subclones containing *TP53* mutations were seen in eight patients at a level below our pre-defined 0.05 VAF level for the TSCA panel. Of these, two patients had another *TP53* mutation at >0.05 VAF already defining poor risk. The remaining six patients did not have an aggressive disease course and none of these patients had progressed within the follow up period with a median follow up time of 35.4 months. In comparison, of those cases progressing with *TP53* mutations at a VAF>0.05 the median time to progression was 14 months. In the multivariate analysis this overall *TP53* mutated group (all VAF levels included) weakened the prognostic strength of *TP53* lesions in the model. No

clear prognostic relevance could be attached to these mutations within our follow up period. These findings have since been confirmed in collaboration using data from another frontline clinical trial group (unpublished results). Previous reports of small subclones containing *TP53* mutations conferring poor risk have certainly proven that the outgrowth of these clones at a subsequent time period coinciding with clinical relapse is a feature (Rossi et al. 2014; Malcikova et al. 2014). However the time period described in both publications is in the order of 3-4 years and would not necessarily preclude upfront CIT (M. Hallek et al. 2008). Further study in this area is ongoing and the particular *TP53* mutation, not solely the presence of mutation within the *TP53* gene, may be important. It is worth noting that seven of the cases with a *TP53* mutation of >0.10 VAF did not progress during follow up.

Patients with trisomy 12 are a well-recognised group with a heterogeneous outcome (Döhner et al. 2000; Athanasiadou et al. 2006; Ibbotson et al. 2012). Here, patients with trisomy 12 and trisomy 19 were all MRD negative and no patient has progressed to date. Within the limitation of small sample size, this confirms previous findings and defines trisomy 12 with trisomy 19 as a clinically separate subgroup (Sellmann et al. 2007). This theory is given more weight with the recent discovery that this subgroup contains exclusively isotype-switched sIgD-positive CLL, an otherwise rare finding in CLL (Ibbotson et al. 2012). This observation of an excellent clinical response contrasts with the findings of López *et al.* (2012) where the trisomy 12 with trisomy 18 group had the shortest time to first treatment compared to other trisomy 12 subgroups studied (López et al. 2012). Perhaps the additional trisomy 19 and/or trisomy

18 leads to disease progression; however, in our study this subgroup was very responsive to FCR treatment with a positive predictive value of 100% of reaching an MRD negative remission with CIT. Strong evidence now exists that an MRD negative remission, particularly after FCR treatment, can confer a long PFS and increases overall survival (Tam et al. 2014; Fischer et al. 2016; Kwok et al. 2016).

On the other hand, trisomy 12 without trisomy 19 independently predicts a shorter PFS. There have been reports that trisomy 12 CLL associates significantly with *NOTCH1* mutations and this gene mutation confers a poor prognosis in CLL (Rossi et al. 2012; Del Giudice et al. 2012; Weissmann et al. 2013; Oscier et al. 2013; Baliakas et al. 2014). Early progression was not seen within this subgroup in our data. However, longer follow up time and also a larger data set is required to definitively answer this question. It is clear from the data presented in this Chapter that the trisomy 12 group is heterogenous with varied responses to CIT. The frequent finding of trisomy 12 and *NOTCH1* mutations in Richter's transformation highlights this combination as a potential poor risk feature (G. Fabbri et al. 2011; Fabbri et al. 2013; Strati et al. 2015). This is an area of ongoing research.

In addition to *NOTCH1*, mutations in *SF3B1* are also associated with a shorter time to first treatment, shorter PFS and more recently shorter OS across variously treated groups (Quesada et al. 2011; L. Wang et al. 2011a; D. Rossi et al. 2011; Oscier et al. 2013; Jeromin et al. 2014; Stephan Stilgenbauer et al. 2014; Baliakas et al. 2014). In the data presented here an independent

association between *SF3B1* mutations and MRD results was not seen. This confirms the findings from the CLL8 trial where MRD outcome did not correlate with *SF3B1* mutations (Stilgenbauer et al. 2014). Interestingly, in the CLL8 clinical trial a shorter PFS and OS for cases with *SF3B1* mutations was seen, but this was not the case for our data. While there may be an intermediate risk of progression associated with *SF3B1* mutations defined by the multivariate analysis and depicted in the final Kaplan-Meier hierarchy, longer follow up will be required to confirm this. The goal of this project was to identify genetic markers of early progression. While *SF3B1* mutations clearly define a poor prognostic subtype of CLL, this marker does not yet influence treatment decisions.

Defining significant associations between interacting genomic features and clinical outcome is extremely challenging in CLL with a substantial number of low frequency mutations (Guièze & Wu 2015). To address this we extended previous genomic complexity (GC) scoring approaches (Ouillette et al. 2011; Gunnarsson et al. 2011; Knight et al. 2012). Firstly by including both the total number of CNAs and cnLOH a genomic complexity score was calculated. Measuring progression free survival for each of these groups revealed a shorter PFS with increasing number of lesions. The resulting genomic complexity score was divided into a high and low risk group with the optimal cut off defined by the strongest statistical association and highest hazard ratio defined by PFS. This was defined at a level of four or more lesions. This is comparable to the score devised by Ouillette *et al.* (2011). Rather disappointingly, despite a clear correlation with increasing number of lesions and shorter time to progression,

these findings were not significant in the multivariate analysis. This highlights the association between increasing number of CNAs/cnLOH and high-risk CLL driver mutations. Perhaps with longer follow up prognostic value for this test will be seen.

With a view to include mutations in addition to CNAs and cnLOH a scoring system based on the total number of CLL drivers affected per patient was calculated similarly to the study using MPS in myelodysplastic syndrome (Papaemmanuil et al. 2013). In our study an inverse relationship between the number of CLL drivers and outcome was noted. However, similar to GC defined by number of acquired chromosomal copy number changes, this was not independent of *TP53* mutation status.

A further global method of prognostic relevance, subclonal heterogeneity, can be inferred from variant allele frequencies and genome-wide arrays (Gerlinger et al. 2012; Shah et al. 2012; Ding et al. 2012). Investigating this, our finding of a subclonal driver mutation present in 65% of samples was comparable to previous publications (Landau et al. 2013), suggesting that our targeted deep sequencing method allowed the precise inference of subclonal architecture at least for the target lesions tested. However, differences in outcome between subclonal and clonally defined groups were not defined. A number of differences exist between this study and that of Landau *et al.* (2103), namely in the methodology for defining subclonal populations but also the CLL population tested. As the patients tested in this study were enrolled in frontline treatment clinical trials they were, by definition, already in a higher risk group compared to selecting patients at diagnosis or non-treatment time-points. The methodlogy

applied by Landau *et al.* (2013) is automated using the ABSOLUTE algorithm and while the same principles of purity and ploidy corrections were applied to the analysis described in this thesis, the caveat of a lower number of mutations available for subclonal analyses from a MPS panel per patient compared with increased numbers of mutation data-points from WES, or even WGS, make this analysis more difficult.

Additional clinically relevant driver mutations within the coding regions of the CLL genome have been discovered since commencing the study presented here (Ljungström et al. 2016; Parker et al. 2016). Furthermore, the biological importance of mutations in non-coding genomic regions are emerging (Burns et al. 2014; Puente et al. 2015). Only the analysis of integrated genomic data of up to 2000 uniformly treated CLL patients will refine the current model to further identify patients at risk of early progression who benefit from the established non-CIT targeted treatment approaches (C. Holmes 2014, personal communication). This will require large international collaborative efforts across clinical trial groups.

In conclusion, in this Chapter the value of integrated genomics for refining response prediction to CIT is demonstrated. Patients with a shorter PFS identified by this integrated approach were *TP53* mutated and *SAMHD1* mutated cases. The results presented indicate that patients like these should be considered for alternative frontline non-chemotherapy based treatment. Additionally, good risk patients who are sensitive to CIT may benefit from significantly prolonged PFS and indeed, may even be cured with these agents.

It will be important to fully evaluate these genomic lesions within the setting of novel agents such as Ibrutinib and Venetoclax so that the benefits of integrated genomics can be fully realised in frontline patient care.

CHAPTER 6 DISCUSSION

6.1 Introduction

For many patients with CLL, the outlook has changed significantly compared with just a decade ago. In the past, for the majority of patients requiring treatment, palliative measures alone were available. Contrast this with the current data revealing functional cures for a subset of CLL patients who are disease free for up to 14 years after CIT (Thompson et al. 2016). Additionally, novel targeted treatment options for high-risk patients are leading to significant treatment-free survival in patients refractory to CIT where dismal outcomes were expected previously (Byrd et al. 2014; O'Brien et al. 2015; Stilgenbauer et al. 2016; Fischer et al. 2017).

In parallel, genomic technology has remarkably progressed and facilitated the study of large numbers of cancer genomes including those of CLL to be sequenced at a far more sensitive level at relatively low cost (Rigolin et al. 2016; Sutton et al. 2015; Nadeu et al. 2016; Baliakas et al. 2014; Chalmers et al. 2017). Clinical heterogeneity, which is a feature of CLL, is at least partly explained by the numerous and varied acquired genetic events with at least one CLL driver lesion found in the majority of patients (Landau et al. 2015; Puente et al. 2015). The eruption of genomic knowledge in CLL has led to the description of novel mutations affecting cellular pathways not previously considered important in CLL and has facilitated an improved understanding of

drivers of clonal evolution namely subclonal tumour heterogeneity (Landau et al. 2013; Wang et al. 2017).

6.2 Summary of thesis and key findings

During the course of this thesis, the field of CLL genomics has greatly advanced due to the application of NGS technologies. I sought to address two particular questions in CLL genomics.

Firstly, as outlined in Chapter 3, by applying whole genome sequencing to serial CLL samples, I asked whether novel candidate mutations could be discovered and validated to provide further insights into the underlying disease biology and mechanisms of chemo-resistance. In Chapter 3, recurrent *SAMHD1* mutations increasing from a frequency of 3% pre-treatment to 11% at relapse were found, and associated cnLOH of chromosome 20q was seen typically in patients harbouring one rather than multiple *SAMHD1* mutations. Decreased mRNA and protein expression in *SAMHD1* mutated CLL cells was seen compared to wild type, and further functional validation pointed towards an anti-proliferative function for SAMHD1 along with involvement in the DNA damage response pathway.

Secondly, by applying an integrated approach of MPS and genome-wide array I endeavoured to define the genomic landscape of frontline CLL patients by searching specifically for novel predictors of early progression. In Chapter 4, a description of the genomic landscape of pre-treatment CLL, using next-generation sequencing and array technologies, was undertaken. The frequency and interactions of key CLL drivers at the time, in a clinical trial cohort of 250 patients, were described and measures of inter- and intra-tumour heterogeneity

were evaluated. In Chapter 5, markers of early progression were evaluated from this well-annotated CLL clinical trial cohort treated with CIT; Here, *SAMHD1* was identified as the sole novel independent predictor of shorter PFS and highlighting intermediate and good risk markers. At least within this data set I also ruled out an independent predictive role of global measures of tumour heterogeneity and of very low-level *TP53* subclones (VAF <0.05) for PFS. Furthermore, the finding of *SAMHD1* mutations leading to early progression was confirmed both with a validation series of historical clinical trial samples and with a second statistical method.

6.3 *SAMHD1* as a driver gene in CLL

As outlined in Chapter 3, whole genome sequencing was applied at various disease time-points to samples of three CLL patients with UM-IGHV CLL and without *TP53* disruption in an effort to discover novel candidate mutations driving disease progression. This study was undertaken during a period when publications relating to whole genome sequencing in CLL were not available and very few applications relating to other haematological malignancies or solid tumours were published (Ding et al. 2012; Egan et al. 2012). The finding of varied and dynamic somatic mutation profiles has subsequently informed the field of CLL genomics and facilitated the validation of dedicated bioinformatics pipeline development (Schuh et al. 2012; Landau et al. 2013; Fischer et al. 2014; Jiao et al. 2014).

Specifically for this thesis, the finding in one CLL patient of an acquired clonal *SAMHD1* mutation that was present at all time-points under investigation

became the focus of subsequent work. It was possible to infer the number of CLL cells affected by a mutation based on the VAF. To extrapolate from this, when a somatic mutation is present in all CLL cells, it may have been present early in the disease course or at the time of disease initiation. This was thought to be the case for this *SAMHD1* mutation and it was of interest as a potential early pathogenic event in the CLL disease course.

The intriguing presentation of a new case of CLL in a 24 years old patient with a background history of inherited AGS and a known *SAMHD1* mutation made this WGS finding more relevant. Constitutional genetic diseases have guided the study and understanding of disrupted pathways in haematological malignancies. The alpha-thalassaemia X-linked intellectual disability (ATRX) syndrome that leads to severe developmental delay and a mild to moderate anaemia, is underpinned by the presence of a germline ATRX mutation (Gibbons 2006). Examining the function of the ATRX protein paved the way for a greater understanding of globin gene expression and has facilitated the study of the alpha thalassaemia myelodysplastic syndrome (ATMDS) where acquired mutations of *ATRX* and a severe anaemia phenotype are seen (Gibbons et al. 2003). In relation to hereditary cancer predisposition syndromes, inherited mutations of *ATM* lead to an increased risk of all cancers including lymphoid malignancies (Stankovic et al. 1998; Mavrou et al. 2008). Similarly inherited cancer syndromes such as Li-Fraumeni syndrome are well described in relation to germline *TP53* mutations (Srivastava et al. 1990; Malkin et al. 1990). In CLL both *ATM* and *TP53* disruption are well described in the literature and detailed in Chapters 1, 4 and 5. With this in mind the finding of a germline *SAMHD1*

mutation attributing to CLL development was certainly plausible and worthwhile pursuing further.

Cohort screening of high-risk CLL patients revealed recurrent acquired *SAMHD1* mutations with an increased frequency of *SAMHD1* mutations in patients with relapsed or refractory CLL. Driver mutations in cancer are typically defined initially by an increased frequency of a mutated gene and predicted pathogenicity of the particular mutation – a combined frequency and functional based model (Pon & Marra 2015). The most likely candidates from this method can then be functionally validated, a process that is more time-consuming (Pon & Marra 2015). This enrichment of *SAMHD1* mutations in high-risk populations suggests cases where these mutations might confer or contribute to chemotherapy refractoriness, a finding that has since been confirmed in two international studies (Landau et al. 2015; Amin et al. 2016). Large-scale sequencing efforts have catalogued thousands of somatic mutations in CLL and over 60 CLL driver genes and CNAs (L. Wang et al. 2011a; Puente et al. 2011; Quesada et al. 2011; G. Fabbri et al. 2011; Schuh et al. 2012; Edelmann et al. 2013; Landau et al. 2013; Landau et al. 2015; Puente et al. 2015; Ljungström et al. 2016; Wang et al. 2017). It is now clear that certain mutations are present at a stable level over the course of the CLL disease process and these occur earlier in the course of the disease such as heterozygous del(13q), trisomy 12 and *MYD88* abnormalities (Landau et al. 2015; Guièze & Wu 2015). Other mutations clearly increase over time, such as *TP53* and *ATM* (Guièze et al. 2015; Guièze & Wu 2015; Landau et al. 2015; Amin et al. 2016). For SAMHD1, it has been shown that in some cases, mutations are acquired and increase

over time, whereas in other cases, mutations are more likely early events and may even occur in the germline. Perhaps for selected patients a *SAMHD1* mutation is an early founder event or at least is detected at the first sampling time-point, and for other patients outgrowth of a *SAMHD1* mutation occurs only when the correct selective pressure is applied, such as the administration of CIT.

6.4 SAMHD1 is involved in genomic integrity

It is not surprising that a protein involved in cell cycle and the DNA damage response is involved in CLL pathogenesis as CLL mutations have been consistently shown to involve DNA damage pathways. While the link between *SAMHD1* mutations and AGS was well described at the time of this study, the function of SAMHD1 was only beginning to be revealed. SAMHD1 was known to be an inhibitor of viral replication in the setting of the HIV-1 virus through a potential hydrolase function. Combined with a nuclease function, preliminary studies found that SAMHD1 regulated intracellular dNTPs (Lahouassa et al. 2012; Powell et al. 2011). Work from my collaborators (Dr T Louis, Prof. M Benkirane, Laboratory of Molecular Virology, Institute of Human Genetics, Montpelier, France) revealed increased cell proliferation in HeLa cells lacking a triphosphohydrolase function. Other groups have since confirmed this role and outlined a key function of SAMHD1 in cell cycle regulation through depletion of dNTPs (E. Franzolin et al. 2013). Additionally it has been shown here that transfection studies led to an increase in cell death upon DSB induction in HeLa cells expressing SAMHD1. This led to question the role of SAMHD1 in DNA damage response. Upon treatment with etoposide SAMHD1

interaction with Cyclin A increased and levels of phosphorylated SAMHD1 also increased. Furthermore a protein involved in the DNA damage response; 53BP1, co-localised with SAMHD1 in the setting of DNA breaks.

Together, these data suggest both an anti-proliferative role for SAMHD1 in regulating the cell cycle and also involvement in DNA damage response. These functions are in keeping with a tumour suppressor role. However, conflicting data from studies in acute myeloid leukaemia (AML) where no impact on cell proliferation was seen and increased SAMHD1 expression led to cytarabine resistance (Herold et al. 2017). It is probable that the effects of SAMHD1 may also depend on the overall genomic context, the cell type and the microenvironment (Herold et al. 2017). The dual function of cancer driver genes is not a novel concept (Stepanenko et al. 2013). A simple example in haematology comes from the positive outcome associated with *SF3B1* mutations in myelodysplasia whereas a negative impact on prognosis for the same point mutation is seen CLL (Papaemmanuil et al. 2011; L. Wang et al. 2011a).

Work over recent years has confirmed the dNTPase function of SAMHD1 (Beloglazova et al. 2013; Antonucci et al. 2016; Seamon et al. 2015). It is believed that an increased mutation rate is an early step in cancer evolution (Loeb et al. 1974; Schmitt et al. 2012). DNA replication fidelity is affected by the action of DNA polymerases and the selection of the appropriate nucleotide (Kunz 1982; Watt et al. 2016). The concentration of dNTPs has been shown to affect DNA replication fidelity (Buckland et al. 2014). In colorectal cancer,

hemizygous mutations in *SAMHD1* alter the balance of dNTP concentrations (Rentoft et al. 2016). Small changes in concentrations of dNTPs affect the mutation frequency in mouse embryos. Recently the role of SAMHD1 in DNA repair has been confirmed and a key role in genomic maintenance was postulated (Daddacha et al. 2017). Intriguingly, in this study, SAMHD1 promoted DNA repair and genomic integrity, through direct involvement in homologous recombination maintaining genomic integrity rather than through a non-dNTPase function.

From this, it can postulated that for both subclonal and clonal *SAMHD1* mutations, as outlined in Chapter 3, an increased level of cellular proliferation can be seen and/or dysfunctional DNA damage response mechanisms can be expected ultimately to lead to genomic instability and leukaemogenesis.

6.5 Clinical implications of integrated genomics

The key question for this study was to isolate predictors of early progression that would highlight markers other than the well-established *TP53* response predictor, to potentially avoid CIT in the future. Cataloguing and evaluating the impact of CLL driver lesions at the same treatment time-point in 250 CLL patients has revealed that the majority of CLL drivers tested do not impact on early progression in the CIT era. A novel predictor of early relapse was seen with *SAMDH1* mutations. In the cohort studied in Chapters 4 and 5, neither *ATM* disruption nor mutations of *SF3B1 or NOTCH1* predicted early progression contrary to much of the data published to date (Rossi et al. 2012; Schnaiter et al. 2013; Rossi et al. 2011; Stilgenbauer et al. 2014; Rose-Zerilli

et al. 2014; Oscier et al. 2013; Baliakas et al. 2014). While *ATM* disruption aligned with MRD positive status, this feature was not associated with a shorter PFS. These other reports were on patients treated in the pre-CIT era, or were performed on heterogenous non-clinical trial groups. The findings reported in Chapter 5 are consistent with a recent WES study of relapsed CLL where solely *TP53* and *SAMHD1* mutations were enriched at relapse (Amin et al. 2016). Both *NOTCH1* and *SF3B1* mutations were predominantly clonal prior to treatment in the study of Amin *et al*. (2016), and did not impact on outcome. The German CLL8 trial is a comparable patient cohort treated with CIT at frontline and has more mature follow up data than that presented in this thesis (Stilgenbauer et al. 2014; Landau et al. 2015). While both *TP53* and *SF3B1* mutations predicted shorter PFS and OS in the CLL8 trial, the median time of progression for *SF3B1* mutated CLL would not be defined as refractory to FCR treatment under current guidelines which are relapse within 2-3 years from CIT (Hallek 2017b).

It is now accepted that MRD negativity is a marker of durable PFS and OS (Böttcher et al. 2012; Strati et al. 2014; Ghia 2012; Kwok et al. 2016; Varghese et al. 2016). Defining patients prior to commencing treatment would avoid potentially toxic side effects of ineffective treatments or alternatively highlight patients most likely to benefit from CIT. In this study trisomy 19, occurring with trisomy 12, was a marker of response to CIT. These cases had IGHV-M. In the era of precision medicine testing for these lesions, in addition to IGHV mutation status, could further refine our treatment algorithms in CLL. Should targeted treatments become the standard of care in the near future, with a subset of CLL

patients harbouring favourable genomics, a discussion regarding FCR treatment may remain appropriate in the correct clinical setting.

Since this project was completed, large numbers of genomic studies using NGS to investigate CLL cohorts have revealed newly identified driver lesions. Over 50 CLL driver lesions have now been catalogued (Landau et al. 2015; Puente et al. 2015; Wang et al. 2017). Of particular importance, through the application of WES, a novel CLL driver, RPS15, that aligns with CIT resistant phenotypes has been validated in a large series of CLL patients (Ljungström et al. 2016). This gene encodes a component of the 40S ribosomal subunit and the authors reveal that mutated *RPS15* leads to impaired p53 stability. Interestingly, in addition to the large number of CLL drivers many of which are infrequently mutated at ≤2%, non-coding variants from WGS data have been reported. Three large WGS sequencing studies revealed mutations in non-coding regions incorporating regulatory elements including *NOTCH1*, *PAX-5*, *ATM, IGLL5, BIRC3* and *SAMHD1* (Puente et al. 2015; Burns et al. 2017; Kasar et al. 2015) and continued work by these groups and others will further reveal the impact of non-coding variants in CLL.

6.6 Global heterogeneity measures do not predict resistance to treatment

Measures of tumour heterogeneity, both intra- and inter-tumour, reveal an overall view of a patient's genome and can inform on clonal evolution and

clinical outcome (Ouillette et al. 2011; Knight et al. 2012; Schuh et al. 2012; Landau et al. 2013). In Chapter 5 neither intra-tumour heterogeneity nor genomic complexity independently predicted poor outcome when individual gene mutations were taken into consideration.

Different measures of global inter-tumour heterogeneity were evaluated. Genomic complexity was assessed and included numbers of both CNAs and cnLOH. A second method, the integrated approach, was to calculate the total burden of CLL drivers per patient. Conflicting reports exist regarding the independent power of genomic complexity to predict early progression or reduced overall survival, and this remains important in the era of novel agents (Thompson et al. 2015; Anderson et al. 2017; Ouillette et al. 2011; Malek 2012). In this thesis a shorter PFS with increasing levels of GC was revealed. However, once *TP53* and *SAMHD1* mutations were accounted for in a multivariate analysis, the statistical significance of GC was lost. Additionally, as individual lesions are not taken into consideration in global analyses, a patient with a CNA associated with good prognosis, such as trisomy 19, will potentially be found to have GC but will not have a poor outcome, and these features may hamper the power of GC measurements as response predictors. With targeted treatment, the effect of individual lesions may be less important, as they can function independently of the TP53 pathway. The presence of a complex karyotype was associated with progressive disease independently of *TP53* disruption in the setting of Ibrutinib and Venetoclax (Thompson et al. 2015; Anderson et al. 2017).

In relation to the total number of CLL drivers affected, a similar picture was revealed. In this study, an increasing number of CLL drivers per patient revealed a significantly shorter PFS. The finding of increasing numbers of CLL drivers was inversely related to PFS and the greatest statistical difference was found between 0-3 CLL drivers and ≥4 CLL drivers. Similarly, the largest published project to date on WGS in CLL, reveals the same finding in univariate analysis with a greater number of CLL driver lesions tested (Puente et al. 2015). This finding was also not independently predictive of outcome in multivariate analysis, possibly due to high-risk CLL drivers acting as confounding events.

Considering the ground-breaking paper identifying subclonal mutations and linking this feature with a shorter failure-free survival, a similar analysis was performed in Chapters 4 and 5 (Landau et al. 2013). It has been outlined in this thesis that the presence of subclonal lesions did not associate with shorter PFS compared with patients harbouring solely clonal lesions. While acknowledging that direct comparisons between these studies are not possible, differences in the methodology to calculate subclones and differences in the population tested may explain a potential disparity between results. A limitation of my work is that solely one time-point was analysed and only inference of clonal evolution is made, rather than analysing patients at two time-points when clonal evolution can be confirmed as with Landau *et al*. (2013) and other studies (Ljungström et al. 2016; Nadeu et al. 2016; Amin et al. 2016; Wang et al. 2017). Also, my study focussed on fourteen driver genes rather than genome-wide analysis and although our approach had a higher sequencing depth, it will not have included a number of subclonal drivers that have been more recently reported.

More recently, studies applying single cell DNA and RNA sequencing on solid tumours and CLL, more accurately defined the subclonal hierarchy and identified new CLL driver lesions (Patel et al. 2014; Vickovic et al. 2016; Wang et al. 2017). Though cost and technical feasibility prohibit the widespread use of this technology currently, this may become the standard method of cancer subclonal analyses in the future. There is no doubt that clonal evolution is a marker of disease progression or chemotherapy resistance, and measures of intra-tumour heterogeneity using subclonal analyses are a means of surveying and quantifying this feature. The question outlined in this thesis was to assess if this feature associated with early progression in the setting of homogenously CIT patients and to date no link has been proven.

6.7 Future perspective

The value of integrating genomic features including CNAs, sSNVs and indels is increasingly recognised (Roos-Weil et al. 2013; Landau et al. 2015; Rigolin et al. 2016; Baliakas et al. 2015; Puente et al. 2015). Currently two technical approaches rather than one are required to assess both copy number analysis and mutation burden. Efforts to implement WGS in the clinical laboratory are well underway (Taylor et al. 2015). WGS facilitates the analysis of all genomic lesions including translocations. Furthermore non-coding regions, which are increasingly relevant in CLL are evaluated (Puente et al. 2015; Burns et al. 2017). As costs decrease and analysis pipelines for cancer samples are validated, in the future WGS may replace separate targeted NGS panels and copy number analyses (J. Klintman *et al*. 2017 submitted).

While the CLL community currently defines a reasonable response from CIT as non-progression within three years of treatment, this may change in the future as less toxic and potentially more efficacious treatments that achieve MRD negative remissions become available (Hallek 2017b; Fischer et al. 2017). In that case, the definitions of treatment response predictors will also need to be updated and this may include markers such as trisomy 12 without trisomy 19, *ATM* disruption, *NOTCH1*, *SF3B1*, and *BIRC3* mutations based on the findings I have outlined in Chapters 4 and 5. In addition, as novel CLL drivers which influence outcome have been discovered since this study, these will need to be incorporated into any novel response predictive algorithm.

Importantly these data suggest that low frequency mutations can be candidate drivers of disease and that large collaborative efforts are required to gain insight into the biological and clinical consequences of these low-frequency mutations, also referred to as the 'long-tail' as many low-frequency variants exist across cancer types. Secondly, mutations in *SAMHD1* can occur as early as in the germline DNA. While differing VAFs were seen in our cohort ruling SAMHD1 out as a founder mutation in all cases, consideration for the impact of germline mutations must be made. This raises issues regarding our currents methods for analysing whole genome sequencing data and the importance of considering germline variants as pathogenic drivers of disease in certain cases, not solely for *ATM* mutations as is now accepted, but also in the cases of *SAMHD1* and *TP53* (Skowronska et al. 2012; Nadeu et al. 2016). Finally, now that the role of SAMHD1 in DNA repair is becoming clearer, possible novel

therapeutics options can be considered combining DSB-inducing agents with novel small molecules that could disrupt the interaction between SAMHD1 and cellular molecules involved in the DNA damage response.

6.8 Conclusion

CLL is an exciting disease to explore and as increasing numbers of treatment options with improved outcomes become available for our patients the need to accurately define treatment response predictors is even more imperative. As many genomic changes are found at a low frequency in CLL this requires a truly personalised approach. International efforts to confirm prospectively in clinical trials the true effects of individual and global genomic alterations are required along with increasing collaborations and systematic storage of genomic and clinical data. Several national, government funded projects are underway such as the 100,000 Genomes Project in the UK. As increasing numbers of cases undergo MPS more germline variants of clinical relevance may be uncovered.

In this work I have outlined a novel recurrently mutated gene in CLL that is involved in cell cycle regulation and maintenance of genomic integrity. All other CLL driver lesions tested were of intermediate or good risk, and did not reflect a CIT resistant profile. As *SAMHD1* mutations align with CIT resistance, independently of *TP53* disruption, the main implication of this finding would be to avoid the use of DNA damaging agents in patients with mutations in *SAMHD1*.

REFERENCES

1000 Genomes Project Consortium, T. 1000 G.P. et al., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), pp.56–65.

Abou-Nassar, K. & Brown, J.R., 2010. Novel agents for the treatment of chronic lymphocytic leukemia. *Clinical advances in hematology & oncology : H&O*, 8(12), pp.886–895.

Adams, J.M. et al., 1985. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature*, 318(6046), pp.533–538.

Adolfsson, J. et al., 2005. Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential. *Cell*, 121(2), pp.295–306.

Adzhubei, I.A. et al., 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), pp.248–249.

Ahn, I.E. et al., 2017. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. *Blood*, p.blood-2016-06-719294.

Alsolami, R., Knight, S.J. & Schuh, a, 2013. Clinical application of

targeted and genome-wide technologies: can we predict treatment responses in chronic lymphocytic leukemia? *Per Med*, 10(4), pp.361–376.

Amin, N. et al., 2016. A Quantitative Analysis of Subclonal and Clonal Gene Mutations Pre- and Post-therapy in Chronic Lymphocytic Leukemia. *Clinical Cancer Research*.

Anderson, K. et al., 2011. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, 469(7330), pp.356–61.

Anderson, M., 2013. Selective Bcl-2 inhibition with ABT-199 is highly active against chronic lymphocytic leukemia (CLL) irrespective of TP53 mutation or dysfunction. *Blood*, 122, p.1304.

Anderson, M.A. et al., 2017. Clinico-pathological features and outcomes of progression of CLL on the BCL2 inhibitor venetoclax. *Blood*, 129(25), p.blood-2017-01-763003.

Andreeff, M. et al., 1980. Discrimination of human leukemia subtypes by flow cytometric analysis of cellular DNA and RNA. *Blood*, 55(2), pp.282–93.

Antonucci, J.M. et al., 2016. SAMHD1-mediated HIV-1 restriction in cells does not involve ribonuclease activity. *Nature medicine*, 22(10), pp.1072–1074.

Artavanis-Tsakonas, S., Rand, M.D. & Lake, R.J., 1999. Notch

signaling: cell fate control and signal integration in development. *Science (New York, N.Y.)*, 284(5415), pp.770–6.

Athanasiadou, A. et al., 2006. Clinical, immunophenotypic, and molecular profiling of trisomy 12 in chronic lymphocytic leukemia and comparison with other karyotypic subgroups defined by cytogenetic analysis. *Cancer genetics and cytogenetics*, 168(2), pp.109–19.

Austen, B. et al., 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. *Journal of Clinical Oncology*, 25(34), pp.5448–5457.

Austen, B. et al., 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(9), pp.3175–3182.

Badoux, X.C. et al., 2009. Patients with Relapsed CLL and 17p Deletion by FISH Have Very Poor Survival Outcomes. *Blood*, 114(22), pp.1248.

Balatti, V. et al., 2012. NOTCH1 mutations in CLL associated with trisomy 12. *Blood*, 119(2), pp.329-331.

Baldauf, H.-M. et al., 2012. SAMHD1 restricts HIV-1 infection in resting CD4+ T cells. *Nature Medicine*, 18(11), pp.1682–1689.

Baliakas, P. et al., 2015. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*, 29(2), pp.329–336.

Baliakas, P. et al., 2014. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*, (April), pp.1–8.

Beloglazova, N. et al., 2013. Nuclease activity of the human SAMHD1 protein implicated in the Aicardi-Goutieres syndrome and HIV-1 restriction. *The Journal of biological chemistry*, 288(12), pp.8101–10.

Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of th e Royal Statistical Society, Series B (Methodological)*, 57(1), pp.289–300.

Bentley, D.R. et al., 2009. Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, 456(7218), pp.53–59.

Bester, A.C. et al., 2011. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell*, 145(3), pp.435–46.

Böttcher, S. et al., 2012. Minimal residual disease quantification is an

independent predictor of progression-free and overall survival in chronic lymphocytic leukemia: A Multivariate analysis from the randomized GCLLSG CLL8 trial. *Journal of Clinical Oncology*, 30(9), pp.980–988.

Bouaoun L, 2016. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, Olivier M. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. Hum Mutat. 2016 Sep;37(9):865-76.

DE Braekeleer, M. et al., 2016. Immunoglobulin gene translocations in chronic lymphocytic leukemia: A report of 35 patients and review of the literature. *Molecular and clinical oncology*, 4(5), pp.682–694.

Braggio, E. et al., 2012. Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. *Leukemia*, 26(February), pp.1698–1701.

Brenner, H., Gondos, A. & Pulte, D., 2008. Trends in long-term survival of patients with chronic lymphocytic leukemia from the 1980s to the early 21st century. *Blood*, 111(10), pp.4916–4921.

Brion, A. et al., 2012. Autologous transplantation in CLL patients with B and C Binet stages: final results of the prospective randomized GOELAMS LLC 98 trial. *Bone Marrow Transplantation*, 47(4),

pp.542–548.

Brown, J.A.L. et al., 2016. Big Data-Led Cancer Research, Application, and Insights. *Cancer research*, 76(21), pp.6167–6170.

Brown, J.R. et al., 2012. Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia. *Clinical Cancer Research*, 18(14), pp.3791–3802.

Brown, J.R. et al., 2012. Integrative Genomic Analysis Implicates Gain of PIK3CA at 3q26 and MYC at 8q24 in Chronic Lymphocytic Leukemia. *Clinical Cancer Research*, 18(14), pp.3791–3802.

Buckland, R.J. et al., 2014. Increased and Imbalanced dNTP Pools Symmetrically Promote Both Leading and Lagging Strand Replication Infidelity G. P. Copenhaver, ed. *PLoS Genetics*, 10(12), p.e1004846.

Bullrich, F. et al., 1996. Minimal region of loss at 13q14 in B-cell chronic lymphocytic leukemia. *Blood*, 88(8), p.3109–15.

Burger, J.A. et al., 2016. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nature communications*, 7, p.11589.

Burger, J.A., 2011. Nurture versus nature: the microenvironment in chronic lymphocytic leukemia. *Hematology. American Society of Hematology. Education Program*, 2011(1), pp.96–103.

Burger, J.A. et al., 2014. Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: a single-arm, phase 2 study. *The Lancet Oncology*, 15(10), pp.1090–1099.

Burns, A. et al., 2017. Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups. *Leukemia*, 32, pp.332-342.

Burns, A. et al., 2014. Whole Genome Sequencing, Targeted Deep Sequencing and Copy-Number Analysis Provide a Comprehensive Picture of the Mutational Landscape of 41 Clinically Annotated CLL Cases. *Blood*, 124(21), p.1958.

Byrd, J.C. et al., 2014. Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. *The New England journal of medicine*, 371(3), pp.213–23.

Byrd, J.C. et al., 2013. Targeting BTK with Ibrutinib in Relapsed Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 369(1), pp.32–42.

Caligaris-Cappio, F., 1996. PERSPECTIVE B-Chronic Lymphocytic Leukemia: A Malignancy of Anti-Self B Cells THE QUEST FOR THE NORMAL COUNTERPART OF B-CLL. *The Journal of The American Society of Hematology*, 87(7), pp.2615-2620.

Caligaris-Cappio, F. & Ghia, P., 2008. Novel Insights in Chronic

Lymphocytic Leukemia: Are We Getting Closer to Understanding the Pathogenesis of the Disease? *Journal of Clinical Oncology*, 26(27), pp.4497–4503.

Calin, G.A. et al., 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *The New England journal of medicine*, 353(17), pp.1793–1801.

Calin, G.A. et al., 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), pp.15524–15529.

Cambier, J.C. et al., 2007. B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nature Reviews Immunology*, 7(8), pp.633–643.

Campbell, P.J. et al., 2008. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature genetics*, 40(6), pp.722–9.

Carter, S.L. et al., 2012. Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology*, 30(5), pp.413–421.

Catovsky, D., Richards, S., Matutes, E., Oscier, D., Dyer, M.J.S., et al., 2007. Assessment of fludarabine plus cyclophosphamide for

patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet (London, England)*, 370(9583), pp.230–9.

Catovsky, D., Richards, S., Matutes, E., Oscier, D., Dyer, M., et al., 2007. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet*, 370(9583), pp.230–239.

Chabosseau, P. et al., 2011. Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. *Nature communications*, 2(May), p.368.

Chalmers, Z.R. et al., 2017. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine*, 9(1), p.34.

Chapiro, E. et al., 2010. Gain of the short arm of chromosome 2 (2p) is a frequent recurring chromosome aberration in untreated chronic lymphocytic leukemia (CLL) at advanced stages. *Leukemia Research*, 34(1), pp.63–68.

Chapman, M.A. et al., 2011. Initial genome sequencing and analysis of multiple myeloma. *Nature*, 471(7339), pp.467–472.

Cheson, B.D. et al., 1996. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood*, 87(12),

pp.4990–7.

Chevallier, P. et al., 2002. CD38 expression and secondary 17p deletion are important prognostic factors in chronic lymphocytic leukaemia. *British journal of haematology*, 116(1), pp.142–50.

Chigrinova, E. et al., 2013. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood*, 122, pp.2673–82.

Chiorazzi, N., Rai, K.R. & Ferrarini, M., 2005. Chronic lymphocytic leukemia. *The New England journal of medicine*, 352(8), pp.804–815.

Cimmino, A. et al., 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proceedings of the National Academy of Sciences*, 102(39), pp.13944–13949.

Clifford, R. et al., 2014. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood*, 123(7), pp.1021–1031.

Clifford, R. & Schuh, A., 2012. *State-of-the-art management of patients suffering from chronic lymphocytic leukemia*, SAGE Publications.

Coon, S., Wang, D. & Wu, L., 2012. Polymorphisms of the SAMHD1 gene are not associated with the infection and natural control of HIV type 1 in Europeans and African-Americans. *AIDS research*

and human retroviruses, 28(12), pp.1565–73.

Corcoran, M.M. et al., 1998. Detailed molecular delineation of 13q14.3 loss in B-cell chronic lymphocytic leukemia. *Blood*, 91(4), pp.1382–90.

Cortelezzi, A. et al., 2009. Low-dose subcutaneous alemtuzumab in refractory chronic lymphocytic leukaemia (CLL): results of a prospective, single-arm multicentre study. *Leukemia*, 23, pp.2027–2033.

Coutre, S.E., 2013. Clinical activity of idelalisib (GS-1101), a selective inhibitor of pi3kd,in phase 1 and 2 trials in chronic lymphocytic leukemia (CLL): effect of Del(17p)/TP53 mutation, Del(11q), IGHV mutation, and NOTCH1 mutation. *Blood (ASH Annual Meeting Abstracts)*, p.122: abs no. 1632.

Cramer, P. & Hallek, M., 2011. Prognostic factors in chronic lymphocytic leukemia-what do we need to know? *Nature reviews. Clinical oncology*, 8(1), pp.38–47.

Crespo, M. et al., 2003. ZAP-70 Expression as a Surrogate for Immunoglobulin-Variable-Region Mutations in Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 348(18), pp.1764–1775.

Cribier, A. et al., 2013. *Phosphorylation of SAMHD1 by Cyclin A2/CDK1 Regulates Its Restriction Activity toward HIV-1*. *Cell*

Reports, 3(4), pp.1036-1043.

Crow, Y.J. & Rehwinkel, J., 2009. Aicardi-Goutie'res syndrome and related phenotypes: Linking nucleic acid metabolism with autoimmunity. *Human Molecular Genetics*, 18(2), pp.130–136.

Daddacha, W. et al., 2017. SAMHD1 Promotes DNA End Resection to Facilitate DNA Repair by Homologous Recombination. *Cell Reports*, 20(8), pp.1921–1935.

Dal Bo, M. et al., 2011. 13q14 Deletion size and number of deleted cells both influence prognosis in chronic lymphocytic leukemia. *Genes Chromosomes and Cancer*, 50(8), pp.633–643.

Damle, R.N. et al., 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*, 94(6), pp.1840–7.

Damm, F. et al., 2014. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discovery,* 4(9), pp.1088-101

Delgado, J. et al., 2012. Chronic lymphocytic leukaemia with 17p deletion: a retrospective analysis of prognostic factors and therapy results. *British journal of haematology*, 157(1), pp.67–74.

Delogu, A. et al., 2006. Gene Repression by Pax5 in B Cells Is Essential for Blood Cell Homeostasis and Is Reversed in Plasma Cells. *Immunity*, 24(3), pp.269–281.

Descours, B. et al., 2012. SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4+ T-cells. *Retrovirology*, 9, p.87.

Desmedt, C. et al., 2012. Next-generation sequencing in breast cancer: first take home messages. *Current opinion in oncology*, 24(6), pp.597–604.

Dighiero, G. et al., 1998. Chlorambucil in Indolent Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 338(21), pp.1506–1514.

Ding, L. et al., 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, 481(7382), pp.506–10.

Döhner, H. et al., 1997. 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. *Blood*, 89(7), pp.2516–22.

Döhner, H. et al., 1999. Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. *Journal of molecular medicine (Berlin, Germany)*, 77(2), pp.266–81.

Döhner, H. et al., 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *The New England journal of medicine*, 343(26), pp.1910–6.

Döhner, H. et al., 1995. p53 gene deletion predicts for poor survival

and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood*, 85(6), pp.1580–9.

Dores, G.M. et al., 2007. Chronic lymphocytic leukaemia and small lymphocytic lymphoma: overview of the descriptive epidemiology. *British Journal of Haematology*, 139(5), pp.809–819.

Dreger, P. et al., 2007. Indications for allogeneic stem cell transplantation in chronic lymphocytic leukemia: the EBMT transplant consensus. *Leukemia*, 21(1), pp.12–17.

Van Dyke, D.L. et al., 2010. A comprehensive evaluation of the prognostic significance of 13q deletions in patients with B-chronic lymphocytic leukaemia. *British Journal of Haematology*, 148(4), pp.544–550.

Edelmann, J. et al., 2013. reveals new recurrent genomic alterations High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood*, 120(24), pp.4783–4794.

Egan, J.B. et al., 2012. Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood*, 120(5), pp.1060–6.

Eichhorst, B.F. et al., 2009. First-line therapy with fludarabine compared with chlorambucil does not result in a major benefit for

elderly patients with advanced chronic lymphocytic leukemia. *Blood*, 114(16), pp.3382–3391.

Eichorst, B., 2009. FCR or BR in Patients With Previously Untreated B-Cell Chronic Lymphocytic Leukemia. *https://clinicaltrials.gov/ct2/show/NCT00769522*.

Elhanati, Y. et al., 2015. Inferring processes underlying B-cell repertoire diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1676).

Eyre, T.A. et al., 2016. NCRI phase II study of CHOP in combination with ofatumumab in induction and maintenance in newly diagnosed Richter syndrome. *British Journal of Haematology*, 175(1), pp.43–54.

Fabbri, G. et al., 2011. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *The Journal of experimental medicine*, 208(7), pp.1389–1401.

Fabbri, G. et al., 2013. Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. , 210(11), pp.2273–2288.

Fabbri, G. & Dalla-Favera, R., 2016. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nature Reviews Cancer*, 16(3).

Fabbri, M. et al., 2011. Association of a MicroRNA/TP53 Feedback Circuitry With Pathogenesis and Outcome of B-Cell Chronic

Lymphocytic Leukemia. *JAMA*, 305(1), p.59.

Fais, F. et al., 1998. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *Journal of Clinical Investigation*, 102(8), pp.1515–1525.

Fazi, C. et al., 2011. General population low-count CLL-like MBL persists over time without clinical progression, although carrying the same cytogenetic abnormalities of CLL. *Blood*, 118(25), pp.6618-6625.

Fernandez-Mercado, M. et al., 2013. Targeted re-sequencing analysis of 25 genes commonly mutated in myeloid disorders in del(5q) myelodysplastic syndromes. *Haematologica*, 98(12), pp.1856–1864.

Ferraro, F., Celso, C. Lo & Scadden, D., 2010. Adult stem cels and their niches. *Advances in experimental medicine and biology*, 695, pp.155–68.

Fink, A.M. et al., 2013. Prediction of poor outcome in CLL patients following first-line treatment with fludarabine, cyclophosphamide and rituximab. *Leukemia*, 27(9), pp.1949–1952.

Finn, W.G. et al., 1998. Secondary abnormalities of chromosome 6q in B-cell chronic lymphocytic leukemia: a sequential study of karyotypic instability in 51 patients. *American journal of hematology*, 59(3), pp.223–9.

Fischer, A. et al., 2014. High-definition reconstruction of clonal composition in cancer. *Cell reports*, 7(5), pp.1740–1752.

Fischer, K. et al., 2011. Bendamustine Combined With Rituximab in Patients With Relapsed and/or Refractory Chronic Lymphocytic Leukemia: A Multicenter Phase II Trial of the German Chronic Lymphocytic Leukemia Study Group. *Journal of Clinical Oncology*, 29(26), pp.3559–3566.

Fischer, K. et al., 2016. Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood*, 127(2), pp.208–15.

Fischer, K. et al., 2017. Venetoclax and obinutuzumab in chronic lymphocytic leukemia. *Blood*, 129(19), pp.2702–2705.

Flicek, P. et al., 2011. Ensembl 2011. *Nucleic Acids Research*, 39(Database), pp.D800–D806.

Forbes, S.A. et al., 2009. COSMIC (the Catalogue of Somatic Mutations In Cancer): A resource to investigate acquired mutations in human cancer. *Nucleic Acids Research*, 38(SUPPL.1).

Forconi, F. et al., 2008. Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. *Br J Haematol*, 143(4), pp.532–536.

Franzolin, E. et al., 2013. The deoxynucleotide triphosphohydrolase

SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proceedings of the National Academy of Sciences*, 110(35), pp.14272–14277.

Franzolin, E. et al., 2013. The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proceedings of the National Academy of Sciences*, 110(35), pp.14272–14277.

Freise, K.J. et al., 2016. Impact of Venetoclax Exposure on Clinical Efficacy and Safety in Patients with Relapsed or Refractory Chronic Lymphocytic Leukemia. *Clinical pharmacokinetics*.

Furman, R.R. et al., 2014. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *The New England journal of medicine*, 370(11), pp.997–1007.

Galton, D.A.G. et al., 1961. The Use of Chlorambucil and Steroids in the Treatment of Chronic Lymphocytic Leukaemia. *British Journal of Haematology*, 7(1), pp.73–98.

Gerlinger, M. et al., 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine*, 366(10), pp.883–92.

Ghia, P., 2012. A look into the future: can minimal residual disease guide therapy and predict prognosis in chronic lymphocytic leukemia? *Hematology / the Education Program of the American*

Society of Hematology. American Society of Hematology. Education Program, 2012, pp.97–104.

Ghia, P. et al., 2003. The pattern of CD38 expression defines a distinct subset of chronic lymphocytic leukemia (CLL) patients at risk of disease progression. *Blood*, 101(4), pp.1262–1269.

Ghia, P. & Hallek, M., 2014. Management of chronic lymphocytic leukemia. *Haematologica*, 99(6), pp.965–72.

Gibbons, R., 2006. Alpha thalassaemia-mental retardation, X linked. *Orphanet Journal of Rare Diseases*, 1(1), p.15.

Gibbons, R.J. et al., 2003. Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the α-thalassemia myelodysplasia syndrome (ATMDS). *Nature Genetics*, 34(4), pp.446–449.

Del Giudice, I. et al., 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(CII), pp.437–441.

Gladstone, D.E. et al., 2012. The importance of IGHV mutational status in del(11q) and del(17p) chronic lymphocytic leukemia. *Clinical lymphoma, myeloma & leukemia*, 12(2), pp.132–7.

Goldstone, D.C. et al., 2011. HIV-1 restriction factor SAMHD1 is a

deoxynucleoside triphosphate triphosphohydrolase. *Nature*, 480, pp.379–382.

Goncalves, A. et al., 2012. SAMHD1 is a nucleic-acid binding protein that is mislocalized due to aicardi-goutières syndrome-associated mutations. *Human Mutation*, 33, pp.1116–1122.

Gondek, L.P. et al., 2007. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood*, 111(3), pp.1534–1542.

Gonzalez, D., Martinez, P., Wade, R., Hockley, S., Oscier, D., Matutes, E., Dearden, C.E., et al., 2011. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: Results from the LRF CLL4 trial. *Journal of Clinical Oncology*, 29(16), pp.2223–2229.

Gonzalez, D., Martinez, P., Wade, R., Hockley, S., Oscier, D., Matutes, E., Dearden, C.E., et al., 2011. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 29(16), pp.2223–9.

Good, R.A. & Zak, S.J., 1956. DISTURBANCES IN GAMMA GLOBULIN SYNTHESIS AS "EXPERIMENTS OF NATURE" *Pediatrics*, 18(1).

Greaves, M. & Maley, C.C., 2012. *Clonal evolution in cancer*, Nature Research, 481(7381), pp.306-313.

Grubor, V. et al., 2009. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood*, 113(6), pp.1294–1303.

Guarini, A. et al., 2008. BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgVH unmutated chronic lymphocytic leukemia (CLL) cells. *Blood*, 112(3), pp.782–792.

de Guibert, S. et al., 2014. Bendamustine, Ofatumumab and High-Dose Methylprednisolone (BOMP) in Relapsed/Refractory CLL: Results of a Planned Interim Analysis of the French CLL Intergroup ICLL01 Phase II Trial. *Blood*, 124(21).

Guièze, R., Robbe, P., Clifford, R., De Guibert, S., Pereira, B., Timbs, A., Dilhuydy, M.-S.S., et al., 2015. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood*, 126(18), pp.2110–7.

Guièze, R. & Wu, C.J., 2015. Genomic and epigenomic heterogeneity in chronic lymphocytic leukemia. *Blood*, 126(4)., pp.445-453.

Gunnarsson, R. et al., 2011. Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia.

Haematologica, 96(8), pp.1161–1169.

Gunnarsson, R. et al., 2010. Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24, pp.211–215.

Haferlach, C. et al., 2007. Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. *Leukemia*, 21(12), pp.2442–2451.

Hallek, M. et al., 2010. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet,* 376(9747), pp.1164-1174.

Hallek, M., 2017a. Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology*, 92(9), pp.946–965.

Hallek, M. et al., 2008. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*,

111(12), pp.5446–5456.

Hallek, M., 2017b. Revised iWCLL guidelines 2017 XVII International Workshop on Chronic Lymphocytic Leukemia guidelines 2017.

Hallek, M. et al., 1996. Serum β_2-Microglobulin and Serum Thymidine Kinase are Independent Predictors of Progression-Free Survival in Chronic Lymphocytic Leukemia and Immunocytoma. *Leukemia & Lymphoma*, 22(5–6), pp.439–447.

Hamblin, T.J. et al., 2000. Immunoglobulin V genes and CD38 expression in CLL. *Blood*, 95(7), pp.2455-2457.

Hamblin, T.J. et al., 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*, 94(6), pp.1848–54.

Hanada, M. et al., 1993. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood*, 82(6), pp.1820–8.

Haupt, Y. et al., 1997. Mdm2 promotes the rapid degradation of p53. *Nature*, 387(6630), pp.296–299.

Heintel, D. et al., 2001. Association of CD38 Antigen Expression with Other Prognostic Parameters in Early Stages of Chronic Lymphocytic Leukemia. *Leukemia and Lymphoma*, 42(6), pp.1315–1321.

Henriques, A. et al., 2013. Combined Patterns of IGHV Repertoire and

Cytogenetic/Molecular Alterations in Monoclonal B Lymphocytosis versus Chronic Lymphocytic Leukemia P. G. Thomas, ed. *PLoS ONE*, 8(7), p.e67751.

Herishanu, Y. et al., 2011. The lymph node microenvironment promotes B-cell receptor signaling, NF- B activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood*, 117(2), pp.563–574.

Herman, S.E.M. et al., 2011. Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood*, 117(23), pp.6287–6296.

Herman, S.E.M. et al., 2010. Phosphatidylinositol 3-kinase- inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. *Blood*, 116(12), pp.2078–2088.

Hernández, J.A. et al., 2009. A high number of losses in 13q14 chromosome band is associated with a worse outcome and biological differences in patients with B-cell chronic lymphoid leukemia. *Haematologica*, 94(3), pp.364–71.

Herold, N. et al., 2017. Targeting SAMHD1 with the Vpx protein to improve cytarabine therapy for hematological malignancies. *Nature Medicine*, 23(2), pp.256–263.

Herrmann, A., Happel, A.-U. & Gramberg, T., 2016. SAMHD1 in Retroviral Restriction and Innate Immune Sensing--Should We Leash the Hound? *Current HIV research*, 14(3), pp.225–34.

Hillmen, P., A trial looking at treatment for people with newly diagnosed chronic lymphocytic leukaemia (ADMIRE) | Cancer Research UK. Available at: http://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-trial-looking-at-treatment-for-people-with-newly-diagnosed-chronic-lymphocytic-leukaemia.

Hillmen, P. et al., 2007. NCRI CLL201 Trial: A Randomized Phase II Trial of Fludarabine, Cyclophosphamide and Mitoxantrone (FCM) with or without Rituximab in Previously Treated CLL. *Blood*, 110(11), pp.752.

Hillmen, P., 2010. Rituximab Plus Chlorambucil In Patients with CD20-Positive B-Cell Chronic Lymphocytic Leukemia (CLL): Final Response Analysis of An Open-Label Phase II Study. *Blood (ASH Annual Meeting Abstracts)*, 116(21), p.697.

Hillmen, P., 2011. Using the Biology of Chronic Lymphocytic Leukemia to Choose Treatment. *Hematology*, 2011(Figure 1), pp.104–109.

Hjalmar, V. et al., 2001. Sequential fluorescence in situ hybridization analyses for trisomy 12 in chronic leukemic B-cell disorders. *Haematologica*, 86(2), pp.174–80.

Hoehn, K.B. et al., 2016. The Diversity and Molecular Evolution of B-Cell Receptors during Infection. *Molecular biology and evolution*, 33(5), pp.1147–57.

Hoellenriegel, J. et al., 2011. The phosphoinositide 3'-kinase delta inhibitor, CAL-101, inhibits B-cell receptor signaling and chemokine networks in chronic lymphocytic leukemia. *Blood*, 118(13), pp.3603–3612.

Hohenstein, P., 2004. Tumour Suppressor Genes—One Hit Can Be Enough. *PLoS Biology*, 2(2), p.e40.

Howard, D.R. et al., 2017. Results of the randomized phase IIB ARCTIC trial of low-dose rituximab in previously untreated CLL. *Leukemia*, 31(11), pp.2416–2425.

Howlader, 2016. No Title. *SEER Cancer Statistics Factsheets: Chronic Lymphocytic Leukemia. National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/statfacts/html/clyl.html*.

Hrecka, K. et al., 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, 474(7353), pp.658–661.

http://dgv.tcag.ca/dgv/app/home, Database of Genomic Variants.

https://www.broadinstitute.org/igv/, Integrated Genomics Viewer.

https://www.r-project.org, The R project for Statistical Computing.

Huang, W. et al., 2009. Prognostic value of CRM1 in pancreas cancer.

Clinical and investigative medicine. Medecine clinique et experimentale, 32(6), p.E315.

Ian Mockridge, C. et al., 2007. Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by VH-gene mutational status. *Blood*, 109(10), pp.4424–4431.

Ibbotson, R. et al., 2012. Coexistence of trisomies of chromosomes 12 and 19 in chronic lymphocytic leukemia occurs exclusively in the rare IgG-positive variant. *Leukemia*, 26(1), pp.170–172.

Jackson, S.P. & Bartek, J., 2009. The DNA-damage response in human biology and disease. *Nature*, 461(7267), pp.1071–8.

Jasek, M. et al., 2010. TP53 mutations in myeloid malignancies are either homozygous or hemizygous due to copy number-neutral loss of heterozygosity or deletion of 17p. *Leukemia*, 24(1), pp.216–219.

Jeromin, S. et al., 2014. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia*, 28, pp.108–17.

Jiao, W. et al., 2014. Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC bioinformatics*, 15, p.35.

Jou, Y.-S. et al., 2004. Clustering of Minimal Deleted Regions Reveals Distinct Genetic Pathways of Human Hepatocellular Carcinoma. *Cancer Research*, 64(9), pp.3030-3036.

Juliusson, G. et al., 1990. Prognostic Subgroups in B-Cell Chronic Lymphocytic Leukemia Defined by Specific Chromosomal Abnormalities. *New England Journal of Medicine*, 323(11), pp.720–724.

Juliusson, G. & Gahrton, G., 1990. Chromosome aberrations in B-cell chronic lymphocytic leukemia. Pathogenetic and clinical implications. *Cancer genetics and cytogenetics*, 45(2), pp.143–60.

Juliusson, G. & Merup, M., 1998. Cytogenetics in chronic lymphocytic leukemia. *Seminars in oncology*, 25(1), pp.19–26.

Kalachikov, S. et al., 1997. Cloning and gene mapping of the chromosome 13q14 region deleted in chronic lymphocytic leukemia. *Genomics*, 42(3), pp.369–77.

Kasar, S. et al., 2015. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nature Communications*, 6, p.8866.

Kienle, D.L. et al., 2005. Evidence for distinct pathomechanisms in genetic subgroups of chronic lymphocytic leukemia revealed by quantitative expression analysis of cell cycle, activation, and apoptosis-associated genes. *Journal of Clinical Oncology*, 23(16), pp.3780–3792.

Kikushige, Y. et al., 2011. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell,* 16;20(2), pp.246-59.

Kipps, T.J. et al., 1989. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia (CD5 B celi/autoantibody/immunglobulin genes/fetal immune repertoire). *Immunology*, 86, pp.5913–5917.

Klein, U. et al., 2001. Gene Expression Profiling of B Cell Chronic Lymphocytic Leukemia Reveals a Homogeneous Phenotype Related to Memory B Cells ✪. *J. Exp. Med*, 121400(11), pp.1625–1638.

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

Knauf, W., 2009. Bendamustine in the treatment of chronic lymphocytic leukemia. *Expert Review of Anticancer Therapy*, 9(2), pp.165–174.

Knauf, W.U. et al., 1997. Serum Levels of Soluble CD23, but not Soluble CD25, Predict Disease Progression in Early Stage B-Cell Chronic Lymphocytic Leukemia. *Leukemia & Lymphoma*, 27(5–6), pp.523–532.

Knight, S.J.L. et al., 2012. Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. *Leukemia*, 26(February), pp.1564–1575.

Koníková, E. & Kusenda, J., 2003. Altered expression of p53 and MDM2 proteins in hematological malignancies. *Neoplasma*, 50(1), pp.31–40.

Kruiswijk, F., Labuschagne, C.F. & Vousden, K.H., 2015. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nature reviews. Molecular cell biology*, 16(7), pp.393–405.

Kujawski, L. et al., 2008. Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. *Blood*, 112, pp.1993–2003.

Kumar, S.K. et al., 2008. Improved survival in multiple myeloma and the impact of novel therapies. *Blood*, 111(5), pp.2516–20.

Kunz, B. a, 1982. Genetic effects of deoxyribonucleotide pool imbalances. *Environmental mutagenesis*, 4, pp.695–725.

Kwok, M. et al., 2016. Minimal residual disease is an independent predictor for 10-year survival in CLL. *Blood*, 128(24), pp.2770–2773.

Laguette, N. et al., 2012. Evolutionary and Functional Analyses of the Interaction between the Myeloid Restriction Factor SAMHD1 and

the Lentiviral Vpx Protein. *Cell Host & Microbe*, 11(2), pp.205–217.

Laguette, N. et al., 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*, 474, pp.654–657.

Lahouassa, H. et al., 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nature Immunology*, 13(3), pp.621–621.

Lakin, N.D. & Jackson, S.P., 1999. Regulation of p53 in response to DNA damage. *Oncogene*, 18(53), pp.7644–7655.

Lamanna, N. et al., 2009. Sequential therapy with fludarabine, high-dose cyclophosphamide, and rituximab in previously untreated patients with chronic lymphocytic leukemia produces high-quality responses: Molecular remissions predict for durable complete responses. *Journal of Clinical Oncology*, 27(4), pp.491–497.

Landau, D.A. et al., 2015. Mutations driving CLL and their evolution in progression and relapse. *Nature*, 526(7574), pp.525–30.

Landau, D. a. et al., 2013. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, 152(4), pp.714–726.

Landau, D. a & Wu, C.J., 2013. Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics.

Genome medicine, 5, p.47.

Lane, D.P., 1992. Cancer. p53, guardian of the genome., *Nature,* 358(6381), pp.15-16.

Lau, R., Niu, M.Y. & Pratt, M.A.C., 2012. cIAP2 represses IKKα/β-mediated activation of MDM2 to prevent p53 degradation. *Cell Cycle*, 11(21), pp.4009–4019.

Lazarian, G., Guièze, R. & Wu, C.J., 2017. Clinical Implications of Novel Genomic Discoveries in Chronic Lymphocytic Leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 35(9), pp.984–993.

LeBien, T.W. & Tedder, T.F., 2008. B lymphocytes: how they develop and function. *Blood*, 112(5), pp.1570-1580.

Lefranc, M.-P. & Lefranc, G., 2001. *The immunoglobulin factsbook*,

Lehmann, S. et al., 2008. Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer*, 112(February), pp.1296–1305.

Levine, A.J. & Oren, M., 2009. The first 30 years of p53: growing ever more complex. *Nature Reviews Cancer*, 9(10), pp.749–758.

Ley, T.J. et al., 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, 456(7218), pp.66–72.

Ling, F. et al., 2016. SAMHD1 polymorphisms were significantly correlated with HIV/SIV virus load in PBMC from Chinese rhesus

macaques and cynomolgus macaques. *Current HIV research*, 15(1), pp.15-22.

Linn, B.S., Linn, M.W. & Gurel, L., 1968. Cumulative illness rating scale. *Journal of the American Geriatrics Society*, 16(5), pp.622–6.

Liu, Y. et al., 1995. 13q deletions in lymphoid malignancies. *Blood*, 86(5), pp.1911-1915.

Ljungström, V., Cortese, D., Young, E., Pandzic, T., Mansouri, L., et al., 2016. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: Clinical impact of recurrent RPS15 mutations. *Blood*, 127(8), pp.1007–1016.

Ljungström, V., Cortese, D., Young, E., Pandzic, T., Mansouri, L., et al., 2016. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: Clinical impact of recurrent RPS15 mutations. *Blood*, 127(8), pp.1007–1016.

Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), pp.129–137.

Loeb, L.A., Springgate, C.F. & Battala4, N., 1974. Errors in DNA Replication as a Basis of Malignant Changes1. *CANCER RESEARCH*, 34, pp.2311–2321.

López, C. et al., 2012. Different distribution of NOTCH1 mutations in chronic lymphocytic leukemia with isolated trisomy 12 or

associated with other chromosomal alterations. *Genes, Chromosomes and Cancer*, 51(9), pp.881–889.

Lunter, G. & Goodson, M., 2010. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, 21(6), pp.936–939.

Mackenbach, J.P., Karanikolos, M. & McKee, M., 2013. The unequal health of Europeans: successes and failures of policies. *The Lancet*, 381(9872), pp.1125–1134.

Maddocks-Christianson, K. et al., 2007. Risk factors for development of a second lymphoid malignancy in patients with chronic lymphocytic leukaemia. *British Journal of Haematology*, 139(3), pp.398–404.

Malcikova, J. et al., 2014. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. *Leukemia*, 29(October), pp.1–9.

Malcikova, J. et al., 2009. Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage. *Blood*, 114(26), pp.5307–5314.

Malcikova, J. et al., 2009. Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: Selection, impact on survival, and response to DNA damage. *Blood*, 114(26), pp.5307–

5314.

Malek, S.N., 2012. The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene*, 32(23), pp.2805–2817.

Malkin, D. et al., 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*, 250(4985), pp.1233–1238.

Mantovani, A. et al., 2008. Tumour immunity: effector response to tumour and role of the microenvironment. *The Lancet*, 371(9614), pp.771–783.

Mao, X., Young, B.D. & Lu, Y.-J., 2007. The application of single nucleotide polymorphism microarrays in cancer research. *Current genomics*, 8(4), pp.219–28.

Mardis, E.R. et al., 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine*, 361, pp.1058–1066.

Marti, G., Orfao, A. & Goolsby, C., 2006. ZAP-70 in CLL: Towards standardization of a biomarker for patient management: History of clinical cytometry special issue. *Cytometry Part B: Clinical Cytometry*, 70B(4), pp.197–200.

Marti, G.E. et al., 2005. Diagnostic criteria for monoclonal B-cell

lymphocytosis. *British Journal of Haematology*, 130(3), pp.325–332.

Matutes, E. et al., 1994. The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia*, 8(10), pp.1640–5.

Matutes, E. et al., 1996. trisomy 12 defines a group of CLL with atypical morphology: correlation between cytogenetic, clinical and laboratory features in 544 patients. *British journal of haematology*, 92(2), pp.382–8.

Mavrou, A. et al., 2008. The ATM gene and ataxia telangiectasia. *Anticancer Research*, 28(1 B), pp.401–405.

McCurdy, R.D., McGrath, J.J. & Mackay-Sim, A., 2008. Validation of the comparative quantification method of real-time PCR analysis and a cautionary tale of housekeeping gene selection. *Gene Therapy and Molecular Biology*, 12(1), pp.15–24.

McLaren, W. et al., 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26(16), pp.2069–2070.

McVean, G.A. et al., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), pp.56–65.

Meinshausen, N. & Bühlmann, P., 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*,

72(4), pp.417–473.

Messmer, B.T. et al., 2005. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. *Journal of Clinical Investigation*, 115(3), pp.755–764.

Messmer, B.T. et al., 2004. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *The Journal of experimental medicine*, 200(4), pp.519–25.

Metzker, M.L., 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), pp.31–46.

Mian, M. et al., 2013. Large genomic aberrations detected by SNP array are independent prognosticators of a shorter time to first treatment in chronic lymphocytic leukemia patients with normal FISH. *Annals of Oncology*, 24, pp.1378–1384.

Migliazza, A. et al., 2001. Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood*, 97(7), pp.2098–104.

Mitelman, F., Mertens, F. & Johansson, B., 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genetics*, 15(4s), pp.417–474.

Molica, S. et al., 2016. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis

of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma*, pp.1–4.

Molica, S. et al., 1999. Clinico-prognostic implications of simultaneous increased serum levels of soluble CD23 and beta2-microglobulin in B-cell chronic lymphocytic leukemia. *European journal of haematology*, 62(2), pp.117–22.

Morton, L.M. et al., 2006. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood*, 107(1), pp.265–276.

Munir, T. et al., 2017. Results of the randomized phase IIB ADMIRE trial of FCR with or without mitoxantrone in previously untreated CLL. *Leukemia*, 31(10), pp.2085–2093.

Nadeu, F. et al., 2016. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood*, 127(17), pp.2122–2130.

Nakatani, Y. & Ogryzko, V., 2003. Immunoaffinity Purification of Mammalian Protein Complexes. In *Methods in enzymology*. pp. 430–444.

Neilson, J.R. et al., 1997. Deletions at 11q identify a subset of patients with typical CLL who show consistent disease progression and reduced survival. *Leukemia*, 11(11), pp.1929–32.

Neuland, C.Y. et al., 1983. Familial chronic lymphocytic leukemia.

Journal of the National Cancer Institute, 71(6), pp.1143–50.

Ngo, V.N. et al., 2011. Oncogenically active MYD88 mutations in human lymphoma. *Nature*, 470(7332), pp.115–119.

Nicholson, I.C., Brisco, M.J. & Zola, H., 1995. Memory B lymphocytes in human tonsil do not express surface IgD. *Journal of immunology (Baltimore, Md. : 1950)*, 154(3), pp.1105–13.

Nieto, W.G. et al., 2009. Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood*, 114(1), pp.33–37.

Noske, A. et al., 2008. Expression of the nuclear export protein chromosomal region maintenance/exportin 1/Xpo1 is a prognostic factor in human ovarian cancer. *Cancer*, 112(8), pp.1733–1743.

Nutt, S.L. & Kee, B.L., 2007. The Transcriptional Regulation of B Cell Lineage Commitment. *Immunity*, 26(6), pp.715–725.

O'Brien, S. et al., 2009. 5-year survival in patients with relapsed or refractory chronic lymphocytic leukemia in a randomized, phase III trial of fludarabine plus cyclophosphamide with or without oblimersen. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(31), pp.5208–12.

O'Brien, S., Jones, J.A., et al., 2014. Efficacy and Safety of Ibrutinib in Patients with Relapsed or Refractory Chronic Lymphocytic

Leukemia or Small Lymphocytic Leukemia with 17p Deletion: Results from the Phase II RESONATE™-17 Trial. *Blood*, 124(21), pp.327.

O'Brien, S., Furman, R.R., et al., 2014. Ibrutinib as initial therapy for elderly patients with chronic lymphocytic leukaemia or small lymphocytic lymphoma: an open-label, multicentre, phase 1b/2 trial. *The Lancet Oncology*, 15(1), pp.48–58.

O'Brien, S.M. et al., 2015. A phase 2 study of idelalisib plus rituximab in treatment-naïve older patients with chronic lymphocytic leukemia. *Blood*, 126(25), pp.2686–94.

O'Brien, S.M. et al., 2008. Phase I study of obatoclax mesylate (GX15-070), a small molecule pan-Bcl-2 family antagonist, in patients with advanced chronic lymphocytic leukemia. *Blood*, 113(2), pp.299–305.

Oeckinghaus, A. & Ghosh, S., 2009. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harbor perspectives in biology*, 1(4), p.a000034.

Orchard, J.A. et al., 2004. ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *The Lancet*, 363(9403), pp.105–111.

Orkin, S.H. & Zon, L.I., 2008. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*, 132(4), pp.631–644.

Ortíz-Maldonado, V., Mozas, P. & Delgado, J., 2016. The biology

behind B-cell lymphoma 2 as a target in chronic lymphocytic leukemia. *Therapeutic Advances in Hematology*, 7(6), pp.321–329.

Oscier, D. et al., 2012. Guidelines on the diagnosis, investigation and management of chronic lymphocytic leukaemia. *British journal of haematology*, 159(5), pp.541–64.

Oscier, D. et al., 2010. Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica*, 95, pp.1705–1712.

Oscier, D.G. et al., 1990. Correlation of chromosome abnormalities with laboratory features and clinical course in B-cell chronic lymphocytic leukaemia. *British Journal of Haematology*, 76(3), pp.352–358.

Oscier, D.G. et al., 1997. Differential Rates of Somatic Hypermutation in VH Genes Among Subsets of Chronic Lymphocytic Leukemia Defined by Chromosomal Abnormalities. *Blood*, 89(11).

Oscier, D.G. et al., 2013. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood*, 121(3), pp.468–75.

Osterborg, A. et al., 1997. Phase II multicenter study of human CD52 antibody in previously treated chronic lymphocytic leukemia. European Study Group of CAMPATH-1H Treatment in Chronic

Lymphocytic Leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 15(4), pp.1567–74.

Ouillette, P. et al., 2011. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood*, 118, pp.3051–3061.

Ouillette, P. et al., 2010. Aggressive chronic lymphocytic leukemia with elevated genomic complexity is associated with multiple gene defects in the response to DNA double-strand breaks. *Clinical Cancer Research*, 16(3), pp.835–847.

Ouillette, P. et al., 2008. Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. *Cancer Research*, 68, pp.1012–1021.

Packham, G. & Stevenson, F., 2010. The role of the B-cell receptor in the pathogenesis of chronic lymphocytic leukaemia. *Seminars in Cancer Biology*, 20(6), pp.391–399.

Papaemmanuil, E. et al., 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*, 122, pp.3616–3627.

Papaemmanuil, E. et al., 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts.

Parikh, S.A. et al., 2013. Diffuse large B-cell lymphoma (Richter

syndrome) in patients with chronic lymphocytic leukaemia (CLL): a cohort study of newly diagnosed patients. *British Journal of Haematology*, 162(6), pp.774–782.

Parikh, S.A., Kay, N.E. & Shanafelt, T.D., 2014. How we treat Richter syndrome. *Blood*, 123(11).

Parker, H. et al., 2010. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia, Published online: 10 December 2010; | doi:10.1038/leu.2010.288*, 25(3), p.489.

Parker, H. et al., 2011. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia*, 25(3), pp.489–497.

Parker, H. et al., 2016. Genomic disruption of the histone methyltransferase SETD2 in chronic lymphocytic leukemia.

Patel, A.P. et al., 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190), pp.1396–1401.

Peled, J.U. et al., 2008. The Biochemistry of Somatic Hypermutation. *Annual Review of Immunology*, 26(1), pp.481–511.

Pettitt, A.R., 2003. Mechanism of action of purine analogues in chronic lymphocytic leukaemia. *British Journal of Haematology*, 121(5), pp.692–702.

Pfeifer, D. et al., 2007. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. *Blood*, 109(3), pp.1202–1210.

Pleasance, E.D. et al., 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, 463(7278), pp.191–6.

Pon, J.R. & Marra, M.A., 2015. Driver and Passenger Mutations in Cancer. *Annual Review of Pathology: Mechanisms of Disease*, 10(1), pp.25–50.

Ponader, S. et al., 2012. The Bruton tyrosine kinase inhibitor PCI-32765 thwarts chronic lymphocytic leukemia cell survival and tissue homing in vitro and in vivo. *Blood*, 119(5), pp.1182–1189.

Porpaczy, E. et al., 2009. Gene expression signature of chronic lymphocytic leukaemia with trisomy 12. *European Journal of Clinical Investigation*, 39(7), pp.568–575.

Porzelius, C., Schumacher, M. & Binder, H., 2010. A general, prediction error-based criterion for selecting model complexity for high-dimensional survival models. *Statistics in medicine*, 29(7–8), pp.830–8.

Pospisilova, S. et al., 2012. ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. *Leukemia*, 26(7), pp.1458–1461.

Powell, R.D. et al., 2011. Aicardi-Gouti??res syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *Journal of Biological Chemistry*, 286(51), pp.43596–43600.

Puente, X.S. et al., 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 526(7574), pp.519–24.

Puente, X.S. et al., 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 475, pp.101–105.

Quesada, V. et al., 2011. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature genetics*, 44(1), pp.47–52.

Quesada, V. et al., 2013. The genomic landscape of chronic lymphocytic leukemia: clinical implications. *BMC medicine*, 11, p.124.

Quesada, V., Ramsay, A.J. & Lopez-Otin, C., 2012. Chronic lymphocytic leukemia with SF3B1 mutation. *The New England journal of medicine*, 366, p.2530.

Rai, K.R. et al., 2000. Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia. *N Engl J Med*, 343(24), pp.1750–1757.

Rassenti, L.Z. et al., 2004. ZAP-70 Compared with Immunoglobulin

Heavy-Chain Gene Mutation Status as a Predictor of Disease Progression in Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 351(9), pp.893–901.

Rausch, T. et al., 2012. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell*, 148(1–2), pp.59–71.

Rawstron, A.C. et al., 2010. Different biology and clinical outcome according to the absolute numbers of clonal B-cells in monoclonal B-cell lymphocytosis (MBL). *Cytometry Part B: Clinical Cytometry*, 78B(S1), pp.S19–S23.

Rawstron, A.C. et al., 2008. Monoclonal B-Cell Lymphocytosis and Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 359(6), pp.575–583.

Rawstron, A.C. et al., 2001. Quantitation of minimal disease levels in chronic lymphocytic leukemia using a sensitive flow cytometric assay improves the prediction of outcome and can be used to optimize therapy. *Blood*, 98(1), pp.29–35.

Rawstron, A.C., Bennett, F. & Hillmen, P., 2007. The biological and clinical relationship between CD5 $^{+}$ 23 $^{+}$ monoclonal B-cell lymphocytosis and chronic lymphocytic leukaemia. *British Journal of Haematology*, 139(5), pp.724–729.

Reindl, L. et al., 2010. Biological and clinical characterization of

recurrent 14q deletions in CLL and other mature B-cell neoplasms. *British Journal of Haematology*, 151(1), pp.25–36.

Rentoft, M. et al., 2016. Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proceedings of the National Academy of Sciences of the United States of America*, 113(17), pp.4723–8.

Reya, T. et al., 2001. Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859), pp.105–111.

Rice, G. et al., 2007. Clinical and Molecular Phenotype of Aicardi-Goutières Syndrome. *The American Journal of Human Genetics*, 81(4), pp.713–725.

Rice, G.I. et al., 2009. Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature genetics*, 41(7), pp.829–832.

Richards, S. et al., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.

Rigolin, G.M. et al., 2016. Extensive next-generation sequencing analysis in chronic lymphocytic leukemia at diagnosis: clinical and biological correlations. *Journal of hematology & oncology*, 9(1), p.88.

Rimmer, A. et al., 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics*, 46(8), pp.912–8.

Rinaldi, A. et al., 2011. Genome-wide DNA profiling better defines the prognosis of chronic lymphocytic leukaemia. *British Journal of Haematology*, 154, pp.590–599.

Rizzo, J.M. & Buck, M.J., 2012. Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing. *Cancer Prevention Research*, 5(7), pp.887–900.

Robak, T. et al., 2010. Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia. *Journal of Clinical Oncology*, 28(10), pp.1756–1765.

Roberts, A., 2009. An ongoing phase 1/2a study of ABT-263; pharmacokinetics (PK), safety and anti-tumor activity in patients (pts) with relapsed or refractory chronic lymphocytic leukemia (CLL). *Blood (ASH Annual Meeting Abstracts)*, 114(883).

Roberts, K.G. et al., 2012. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer cell*, 22(2), pp.153–66.

Robertson, L.E. et al., 1996. Bcl-2 expression in chronic lymphocytic

leukemia and its correlation with the induction of apoptosis and clinical outcome. *Leukemia*, 10(3), pp.456–9.

Roos-Weil, D. et al., 2013. Mutational and Cytogenetic Analyses Of 177 CLL Patients With trisomy 12: A Retrospective Study Of The CLL French Intergroup. *Blood*, 122(21), p.4144-.

Rose-Zerilli, M.J.J. et al., 2014. ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia: Data from the UK LRF CLL4 trial. *Haematologica*, 99, pp.736–742.

Rossi, D. et al., 2008. Biological and clinical risk factors of chronic lymphocytic leukaemia transformation to Richter syndrome. *British Journal of Haematology*, 142(2), pp.202–215.

Rossi, D. et al., 2014. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*, 123(14), pp.2139–47.

Rossi, D., Fangazio, M., et al., 2012. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood*, 119, pp.2854–2862.

Rossi, D. et al., 2013. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood*, 121, pp.1403–1412.

Rossi, D. et al., 2012. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*,

119(2), pp.521–529.

Rossi, D., Rasi, S., et al., 2012. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*, 119(2), pp.521–529.

Rossi, D. et al., 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*, 118(26), pp.6904–6908.

Rossi, D. et al., 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: Association with progression and fludarabine-refractoriness. *Blood*, 118(26), pp.6904–6908.

Rossi, D. et al., 2009. The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: Implications for overall survival and chemorefractoriness. *Clinical Cancer Research*, 15(3), pp.995–1004.

Rossi, D. & Gaidano, G., 2009. Richter syndrome: molecular insights and clinical perspectives. *Hematological Oncology*, 27(1), pp.1–10.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp.5463–7.

Santos, F.P.S. & O'Brien, S., 2012. Small Lymphocytic Lymphoma

and Chronic Lymphocytic Leukemia. *The Cancer Journal*, 18(5), pp.396–403.

Saußele, S. et al., 2016. The concept of treatment-free remission in chronic myeloid leukemia. *Leukemia*, 30(8), pp.1638–1647.

Schetelig, J. et al., 2008. Allogeneic hematopoietic stem-cell transplantation for chronic lymphocytic leukemia with 17p deletion: a retrospective European Group for Blood and Marrow Transplantation analysis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 26(31), pp.5094–100.

Schmitt, M.W., Prindle, M.J. & Loeb, L.A., 2012. Implications of genetic heterogeneity in cancer. *Annals of the New York Academy of Sciences*, 1267(1), pp.110–116.

Schnaiter, A. et al., 2013. NOTCH1, SF3B1, and TP53 mutations in fludarabine-refractory CLL patients treated with alemtuzumab: results from the CLL2H trial of the GCLLSG. *Blood*, 122(7), pp.1266–70.

Schroeder, H.W., Cavacini, L. & Cavacini, L., 2010. Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology*, 125(2 Suppl 2), pp.S41-52.

Schroeder, H.W. & Dighiero, G., 1994. The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire.

Immunology today, 15(6), pp.288–94.

Schroers, R. et al., 2005. Combined analysis of ZAP-70 and CD38 expression as a predictor of disease progression in B-cell chronic lymphocytic leukemia. *Leukemia*, 19(5), pp.750–758.

Schuh, A., Becq, J., Humphray, S., et al., 2012. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood*, 120, pp.4191–4196.

Seamon, K.J. et al., 2015. SAMHD1 is a single-stranded nucleic acid binding protein with no active site-associated nuclease activity. *Nucleic acids research*, 43(13), pp.6486–99.

Seiffert, M.et al., 2012. Cellular origin and pathophysiology of chronic lymphocytic leukemia. *Journal of Experimental Medicine* 19;209(12), pp.2183-98

Sellmann, L. et al., 2007. Trisomy 19 is associated with trisomy 12 and mutated IGHV genes in B-chronic lymphocytic leukaemia. *British journal of haematology*, 138(2), pp.217–20.

Seymour, J.F., 2013. BCL-2 inhibitor ABT-199 (GDC-0199) monotherapy shows anti-tumor activity including complete remissions in high-risk relapsed/refractory (R/R) chronic lymphocytic leukemia

(CLL) and small lymphocytic lymphoma (SLL). *Blood (ASH Annual Meeting Abstracts)*. 2013 122:872

Shah, S.P. et al., 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, 486(7403), pp.395–9.

Shanafelt, T.D. et al., 2009. B-cell count and survival: differentiating chronic lymphocytic leukemia from monoclonal B-cell lymphocytosis based on clinical outcome. *Blood*, 113(18).

Shanafelt, T.D. et al., 2006. Prospective evaluation of clonal evolution during long-term follow-up of patients with untreated early-stage chronic lymphocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 24(28), pp.4634–41.

Shihab, H.A. et al., 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation*, 34(1), pp.57–65.

Shiloh, Y., 2003. ATM and related protein kinases: safeguarding genome integrity. *Nature Reviews Cancer*, 3(3), pp.155–168.

Šidák, Z., 1967. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318), pp.626–633.

Sill, M. et al., 2014. c060 : Extended Inference with Lasso and Elastic-Net Regularized Cox and Generalized Linear Models. *Journal of Statistical Software*, 62(5).

Simon, N. et al., 2011. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *JSS Journal of Statistical Software*, 39(5).

Sims, D. et al., 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), pp.121–132.

Skowronska, A., Austen, B., et al., 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(1), pp.142–146.

Skowronska, A., Parker, A., et al., 2012. Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom leukemia research fund chronic lymphocytic leukemia 4 trial. *Journal of Clinical Oncology*, 30(36), pp.4524–4532.

Sorror, M.L. et al., 2008. Five-year follow-up of patients with advanced chronic lymphocytic leukemia treated with allogeneic hematopoietic cell transplantation after nonmyeloablative conditioning. *Journal of Clinical Oncology*, 26(30), pp.4912–4920.

Souers, A.J. et al., 2013. ABT-199, a potent and selective BCL-2

inhibitor, achieves antitumor activity while sparing platelets. *Nature Medicine*, 19(2), pp.202–208.

Srivastava, S. et al., 1990. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome. *Nature*, 348(6303), pp.747–749.

Stankovic, T. et al., 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(1).

Stankovic, T. et al., 1998. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. *American journal of human genetics*, 62, pp.334–345.

Stankovic, T. et al., 1999. Early reports Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. *The Lancet*, 353, pp.26–29.

Stankovic, T., 2004. Microarray analysis reveals that TP53- and ATM-mutant B-CLLs share a defect in activating proapoptotic responses after DNA damage but are distinguished by major differences in activating prosurvival responses. *Blood*, 103(1), pp.291–300.

Stepanenko, A.A., Vassetzky, Y.S. & Kavsan, V.M., 2013.

Antagonistic functional duality of cancer genes. *Gene*, 529, pp.199–207.

Stephens, P.J. et al., 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1), pp.27–40.

Stevenson, F.K. et al., 2001. The occurrence and significance of V gene mutations in B cell-derived human malignancy. *Advances in cancer research*, 83, pp.81–116.

Stevenson, F.K. & Caligaris-Cappio, F., 2004. Chronic lymphocytic leukemia: Revelations from the B-cell receptor. *Blood*, 103(12), pp.4389–4395.

Stilgenbauer, S., 2008. Biologic and clinical markers for outcome after fludarabine (F) or F plus cyclophosphamide (FC) - comprehensive analysis of the CLL4 trial of the GCLLSG. In *Blood (ASH Annual Meeting Abstracts)*. p. 112:2089.

Stilgenbauer, S. et al., 2007. Clonal evolution in chronic lymphocytic leukemia: Acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. *Haematologica*, 92(9), pp.1242–1245.

Stilgenbauer, S. et al., 2014. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood*, 123, pp.3247–3254.

Stilgenbauer, S. et al., 2014. Gene mutations and treatment outcome in chronic lymphocytic leukemia: Results from the CLL8 trial. *Blood*, 123, pp.3247–3254.

Stilgenbauer, S. et al., 2009. Subcutaneous alemtuzumab in fludarabine-refractory chronic lymphocytic leukemia: Clinical results and prognostic marker analyses from the CLL2H study of the German chronic lymphocytic leukemia study group. *Journal of Clinical Oncology*, 27(24), pp.3994–4001.

Stilgenbauer, S. et al., 2016. Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. *The Lancet Oncology*, 17(6), pp.768–778.

Stilgenbauer, S. & Zenz, T., 2010. Understanding and Managing Ultra High-Risk Chronic Lymphocytic Leukemia. *Hematology*, 2010(1), pp.481–488.

Strati, P. et al., 2014. Eradication of bone marrow minimal residual disease may prompt early treatment discontinuation in CLL. *Blood*, 123(24), pp.3727–32.

Strati, P. et al., 2015. Second Cancers and Richter Transformation Are the Leading Causes of Death in Patients With trisomy 12 Chronic Lymphocytic Leukemia. *Clinical Lymphoma Myeloma and Leukemia*, 15(7), pp.420–427.

Strati, P. & Shanafelt, T.D., 2015. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*, 126(4), pp.454–462.

Stratton, M.R., Campbell, P.J. & Andrew F, P., 2009. The cancer genome. *nature*, 458, pp.719–724.

Sutton, L. et al., 2011. Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC. *Blood*, 117(23), pp.6109–6119.

Sutton, L.-A. et al., 2015. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica*, 100(3), pp.370–376.

Swerdlow, S H. et al, The 2016 Reviision of the World Health Organisation classification of Lymphoid Neoplams, *Blood,* 127(20), 2375-2390.

Swerdlow SH, Campo E, H.N., 2008. *World Health Organization classification of tumours: pathology and genetics of tumours of haematopoietic and lymphoid tissues.*,

Tachibana, A., 2009. *Mutations in the Ataxia Telangiectasia*,

Tam, C.S. et al., 2009. De novo deletion 17p13.1 chronic lymphocytic leukemia shows significant clinical heterogeneity: The M. D.

Anderson and Mayo Clinic experience. *Blood*, 114(5), pp.957–964.

Tam, C.S. et al., 2014. Long-term results of first salvage treatment in CLL patients treated initially with FCR (fludarabine, cyclophosphamide, rituximab). *Blood*, 124(20), pp.3059–64.

Tam, C.S. et al., 2008. Long-term results of the fludarabine, cyclophosphamide, and rituximab regimen as initial therapy of chronic lymphocytic leukemia. *Blood*, 112(4), pp.975–980.

Taylor, J.C. et al., 2015. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature genetics*, 47(7), pp.717–726.

Teng, G. & Papavasiliou, F.N., 2007. Immunoglobulin Somatic Hypermutation. *Annual Review of Genetics*, 41(1), pp.107–120.

Thompson, P.A. et al., 2015. Complex karyotype is a stronger predictor than del(17p) for an inferior outcome in relapsed or refractory chronic lymphocytic leukemia patients treated with ibrutinib-based regimens. *Cancer*, 121(20), pp.3612–3621.

Thompson, P.A. et al., 2016. Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. *Blood*, 127(3).

Thornton, P.D. et al., 2005. Richter's transformation of chronic lymphocytic leukemia. *Leukemia Research*, 29(4), pp.389–395.

Thorselius, M. et al., 2006. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. *Blood*, 107(7), pp.2889–2894.

Thunberg, U., 2001. CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. *Blood*, 97(6), pp.1892–1894.

Tibshirani, R., 1997. THE LASSO METHOD FOR VARIABLE SELECTION IN THE COX MODEL. *STATISTICS IN MEDICINE*, 16, pp.385–395.

Tobin, G. et al., 2003. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood*, 101(12), pp.4952–7.

Tobin, G. et al., 2004. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*, 104(9), pp.2879–2885.

Traver, D. & Akashi, K., 2004. *Lineage Commitment and Developmental Plasticity in Early Lymphoid Progenitor Subsets*, Academic Press.

Treon, S.P. et al., 2012. MYD88 L265P somatic mutation in

Waldenstrom's macroglobulinemia. *N Engl J Med*, 367(9), pp.826–833.

Tse, C. et al., 2008. ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. *Cancer Research*, 68(9), pp.3421–3428.

Tsimberidou, A.-M. et al., 2009. Chemoimmunotherapy may overcome the adverse prognostic significance of 11q deletion in previously untreated patients with chronic lymphocytic leukemia. *Cancer*, 115(2), pp.373–80.

Tsukada, S. et al., 1993. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell*, 72(2), pp.279–90.

Varghese, A.M. et al., 2016. Eradication of minimal residual disease improves overall and progression-free survival in patients with chronic lymphocytic leukaemia, evidence from NCRN CLL207: a phase II trial assessing alemtuzumab consolidation. *British Journal of Haematology*.

Venkatachalam, S. et al., 1998. Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *The EMBO journal*, 17(16), pp.4657–67.

Vetrie, D. et al., 1993. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-

tyrosine kinases. *Nature*, 361(6409), pp.226–233.

Vickovic, S. et al., 2016. Massive and parallel expression profiling using microarrayed single-cell sequencing. *Nature Communications*, 7, p.13182.

Visconte, V. et al., 2012. Emerging roles of the spliceosomal machinery in myelodysplastic syndromes and other hematological disorders. *Leukemia*, 26(12), pp.2447–2454.

Wang, D.-M. et al., 2011. Intermediate prognosis of 6q deletion in chronic lymphocytic leukemia. *Leukemia & Lymphoma*, 52(2), pp.230–237.

Wang, K., Li, M. & Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), p.e164.

Wang, L. et al., 2011a. and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *New England Journal of Medicine*, 365(26), pp.2497–2506.

Wang, L. et al., 2017. Integrated single-cell genetic and transcriptional analysis suggests novel drivers of chronic lymphocytic leukemia. *Genome Research*, 27(8), pp.1300–1311.

Wang, L. et al., 2011b. SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *New England Journal of Medicine*.

Watson, L., Wyld, P. & Catovsky, D., 2008. Disease burden of chronic lymphocytic leukaemia within the European Union. *European Journal of Haematology*, 81(4), pp.253–258.

Watt, D.L. et al., 2016. Genome-wide analysis of the specificity and mechanisms of replication infidelity driven by imbalanced dNTP pools. *Nucleic acids research*, 44(4), pp.1669–80.

Weissmann, S. et al., 2013. Prognostic impact and landscape of NOTCH1 mutations in chronic lymphocytic leukemia (CLL): a study on 852 patients. *Leukemia*, 27(July), pp.2393–6.

Weng, A.P. et al., 2004. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*, 306(5694).

White, T.E. et al., 2013. The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation. *Cell host & microbe*, 13(4), pp.441–51.

Widhalm, S., 2004. Maligne Lymphome?The medically non fit patient. *Der Onkologe*, 10(8), pp.883–889.

Wierda, W.G. et al., 2009. Characteristics associated with important clinical end points in patients with chronic lymphocytic leukemia at initial treatment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(10), pp.1637–43.

Wierda, W.G. et al., 2011. Ofatumumab is active in patients with

fludarabine-refractory CLL irrespective of prior rituximab: Results from the phase 2 international study. *Blood*, 118(19), pp.5126–5129.

Xu, D., Farmer, A. & Chook, Y.M., 2010. Recognition of nuclear targeting signals by Karyopherin-β proteins. *Current opinion in structural biology*, 20(6), pp.782–90.

Yancik, R., 1997. Cancer burden in the aged: an epidemiologic and demographic overview. *Cancer*, 80(7), pp.1273–83.

Yao, Y. et al., 2009. The expression of CRM1 is associated with prognosis in human osteosarcoma. *Oncology reports*, 21(1), pp.229–35.

Yau, C. et al., 2010. A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biology*, 11(9), p.R92.

Yoshida, K. et al., 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, 478, pp.64–69.

Yuille, M.R. et al., 2000. Familial chronic lymphocytic leukaemia: a survey and review of published studies. *British Journal of Haematology*, 109(4), pp.794–799.

Zenz, T., Mertens, D., et al., 2010. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature reviews. Cancer*, 10(1),

pp.37–50.

Zenz, T. et al., 2011. Importance of genetics in chronic lymphocytic leukemia. *Blood Reviews*, 25(3), pp.131–137.

Zenz, T. et al., 2008. Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia : results from a detailed genetic characterization with long-term follow-up. *October*, 112(8), pp.3322–3329.

Zenz, T., Fröhling, S., et al., 2010. Moving from prognostic to predictive factors in chronic lymphocytic leukaemia (CLL). *Best Practice & Research Clinical Haematology*, 23(1), pp.71–84.

Zenz, T., Eichhorst, B., et al., 2010. TP53 mutation and survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology*, 28(29), pp.4473–4479.

Zenz, T. et al., 2010. TP53 mutation profile in chronic lymphocytic leukemia: evidence for a disease specific profile from a comprehensive analysis of 268 mutations. *Leukemia*, 24(12), pp.2072–2079.

Zhang, S. et al., 2012. Identification and Function of Myeloid Differentiation Factor 88 (MyD88) in Litopenaeus vannamei S. D. Fugmann, ed. *PLoS ONE*, 7(10), p.e47038.

Zhu, J. & Emerson, S.G., 2002. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene*, 21(21), pp.3295–

3313.

Zomas, A.P. et al., 1996. Expression of the immunoglobulin-associated protein B29 in B cell disorders with the monoclonal antibody SN8 (CD79b). *Leukemia*, 10(12), pp.1966–70.

APPENDIX 1 LIST OF SEQUENCING PRIMER FOR SANGER SEQUENCING VALIDATION

Gene	Exon	Direction name	Name	Sequence (5'-3')
ATM	3	M13F	ATMexon3M13F	GTAAAACGACGGCCAGTTTTAAGCGCCTGATTCGAGA
ATM	3	M13R	ATMexon3M13R	CAGGAAACAGCTATGACTGCCAAATTCATATGCAAGGC
ATM	7	M13R	ATMexon7M13R	CAGGAAACAGCTATGACTTGGGTTTTGGCTCCTTTCG
ATM	7	M13F	ATMexon7M13F	GTAAAACGACGGCCAGTACCCAGTTGAGCTTGTTTGT
ATM	8	M13F	ATMexon8M13F	GTAAAACGACGGCCAGTTTTGTGGGAGCTAGCAGTGT
ATM	8	M13R	ATMexon8M13R	CAGGAAACAGCTATGACAAGCCCAAAATGCCCAGTTT
ATM	10	M13F	ATMexon10M13F	GTAAAACGACGGCCAGTTTCTACCCCAACAGCGACAT
ATM	10	M13R	ATMexon10M13R	CAGGAAACAGCTATGACTCGGCCAAACAAGAAAAGCA
ATM	11	M13F	ATMexon11M13F	GTAAAACGACGGCCAGTTTGACTTTGGCACTGACCAC
ATM	11	M13R	ATMexon11M13R	CAGGAAACAGCTATGACGCAATCAGGCATAAAGACACAG
ATM	12	M13F	ATMexon12M13F	GTAAAACGACGGCCAGTTCTTTACATGGCTTTTGGTCTTC
ATM	12	M13R	ATMexon12M13R	CAGGAAACAGCTATGACTACCATTCTGGCACGCTTTG
ATM	18	M13F	ATMexon18M13F	GTAAAACGACGGCCAGTGCTGTTGTGCCCTTCTCTTAG
ATM	18	M13R	ATMexon18M13R	CAGGAAACAGCTATGACGGTGGAGGGATTTGGTAGGT
ATM	20	M13R	ATMexon20M13R	CAGGAAACAGCTATGACCCTACCAAAATGCTCCAATTAC
ATM	20	M13F	ATMexon20M13F	GTAAAACGACGGCCAGTTGTTCTTTGTATCGTCGTGACC
ATM	22	M13F	ATMexon22M13F	GTAAAACGACGGCCAGTCTTTTCCCGTAGGCTGATCC
ATM	22	M13R	ATMexon22M13R	CAGGAAACAGCTATGACTACTGCCATCTGCAGCATTC

ATM	23	M13R	ATMexon23M13R	CAGGAAACAGCTATGACTCATACAGTTGTTTTAGAGCAGC
ATM	23	M13F	ATMexon23M13F	GTAAAACGACGGCCAGTCCAGGACACGAAGGGAGATT
ATM	24	M13R	ATMexon24M13R	CAGGAAACAGCTATGACCCTTTTTCACAAGGTGAGGTTC
ATM	24	M13F	ATMexon24M13F	GTAAAACGACGGCCAGTGAGAACCCTGAAACTTTGGATG
ATM	27	M13R	ATMexon27M13R	CAGGAAACAGCTATGACACTGTGGGGAGACTATGGTAA
ATM	27	M13F	ATMexon27M13F	GTAAAACGACGGCCAGTGACGTTACATGAGCCAGCAA
ATM	29	M13R	ATMexon29M13R	CAGGAAACAGCTATGACAAATAACGTCTCGAAGAACAAAGG
ATM	29	M13F	ATMexon29M13F	GTAAAACGACGGCCAGTTTGCCATATGTGAGCAAGCA
ATM	30	M13R	ATMexon30M13R	CAGGAAACAGCTATGACACCTCCACCTGCTCATACAC
ATM	30	M13F	ATMexon30M13F	GTAAAACGACGGCCAGTTGCTGAACAAAAGGACTTCTGA
ATM	31	M13R	ATMexon31M13R	CAGGAAACAGCTATGACCCAAGAGTGAAAAGGGTCCTC
ATM	31	M13F	ATMexon31M13F	GTAAAACGACGGCCAGTTGGCTTACTTTAAAATTATTTCTCTCC
ATM	35	M13R	ATMexon35M13R	CAGGAAACAGCTATGACATAGGCCAGCATTGGATCTG
ATM	35	M13F	ATMexon35M13F	GTAAAACGACGGCCAGTCAGTGGAGGTTAACATTCATCAAG
ATM	37	M13R	ATMexon37M13R	CAGGAAACAGCTATGACATCCAAGTTTGCAGGGGTTG
ATM	37	M13F	ATMexon37M13F	GTAAAACGACGGCCAGTTGTCAACGGGGCATGAAAAT
ATM	41	M13R	ATMexon41M13R	CAGGAAACAGCTATGACTGGGTTTTACACACACATAACTCC
ATM	41	M13F	ATMexon41M13F	GTAAAACGACGGCCAGTGAGTTGGGAGTTACATATTGGTAATG
ATM	46	M13F	ATMexon46M13F	GTAAAACGACGGCCAGTTCTGACCGCATAGCATTTTG
ATM	46	M13R	ATMexon46M13R	CAGGAAACAGCTATGACTTTTTCCCTCAGGCTTTCTG
ATM	47	M13R	ATMexon47M13R	CAGGAAACAGCTATGACAGCTGGCATCCAACTTCTTG
ATM	47	M13_F	ATMexon47M13F	GTAAAACGACGGCCAGTTCCCATATGTCATTTTCATTTCAG
ATM	50	M13R	ATMexon50M13R	CAGGAAACAGCTATGACCACTTGCCTTCATCATGCCA
ATM	50	M13F	ATMexon50M13F	GTAAAACGACGGCCAGTTCATGGCTTTTGTGTTTTACCTT
ATM	53	M13R	ATMexon53M13R	CAGGAAACAGCTATGACAGTAACCAGGGAATGCTGAAG
ATM	53	M13F	ATMexon53M13F	GTAAAACGACGGCCAGTTGGTCAGAAGTGTTGAGGCA
ATM	56	M13R	ATMexon56M13R	CAGGAAACAGCTATGACGCCACTTCACCCAACCAAAT

ATM	56	M13F	ATMexon56M13F	GTAAAACGACGGCCAGTAGACTGTCAAGAGGTGCACA
ATM	57	M13R	ATMexon57M13R	CAGGAAACAGCTATGACTTTGGGTGTCACTCACCATC
ATM	57	M13F	ATMexon57M13F	GTAAAACGACGGCCAGTCAGGTGGTTCCCCTCTCTC
ATM	58	M13F	ATMexon58M13F	GTAAAACGACGGCCAGTTTCTATTGCAACCCAATGCTG
ATM	58	M13R	ATMexon58M13R	CAGGAAACAGCTATGACTGCCAAACAACAAAGTGCT
ATM	59	M13F	ATMexon59M13F	GTAAAACGACGGCCAGTGACCCAGGTAAATAATGGGTCTCA
ATM	59	M13R	ATMexon59M13R	CAGGAAACAGCTATGACTTGGTAGGCAAACAACATTCC
ATM	60	M13F	ATMexon60M13F	GTAAAACGACGGCCAGTCATGTGGTTTCTTGCCTTTG
ATM	60	M13R	ATMexon60M13R	CAGGAAACAGCTATGACCACTTCAGCCTCTTGAGCAG
ATM	61	M13R	ATMexon61M13R	CAGGAAACAGCTATGACTGCACAATCTCCTCCTTTCTG
ATM	61	M13F	ATMexon61M13F	GTAAAACGACGGCCAGTTCTAATGAAAGCCCACTCTGC
BIRC3	7	M13F	BIRC3exon7M13F	GTAAAACGACGGCCAGTTGGAAGGAAGTTTGTGAGCAG
BIRC3	7	M13R	BIRC3exon7M13R	CAGGAAACAGCTATGACTCAGCCTCCCAAGGTGTTAG
SF3B1	14	M13F	SF3B1exon14M13F	GTAAAACGACGGCCAGTATTTTAGGCTGCTGGTCTGG
SF3B1	14	M13R	SF3B1exon14M13R	CAGGAAACAGCTATGACAAACATGGACAGGCTGTGTG
SF3B1	15	M13R	SF3B1exon15M13R	CAGGAAACAGCTATGACTCCTCTGTGTTGGCGGATAC
SF3B1	15	M13F	SF3B1exon15M13F	GTAAAACGACGGCCAGTTGCAGTTTGGCTGAATAGTTG
TP53	2&3	M13F	TP53exon2_3M13F	GTAAAACGACGGCCAGTTCAGACACTGGCATGGTGTT
TP53	2&3	M13R	TP53exon2_3M13R	CAGGAAACAGCTATGACTCATCTGGACCTGGGTCTTC
TP53	4	M13F	TP53exon4M13F	GTAAAACGACGGCCAGTTCCCAAGCAATGGATGATTT
TP53	4	M13R	TP53exon4M13R	CAGGAAACAGCTATGACCAGGCATTGAAGTCTCATGG
TP53	5&6	M13F	TP53exon5&6M13F	GTAAAACGACGGCCAGTTATCTGTTCACTTGTGCCCTG
TP53	5&6	M13R	TP53exon5&6M13R	CAGGAAACAGCTATGACTTGCACATCTCATGGGGTTA
TP53	7	M13F	TP53exon7M13F	GTAAAACGACGGCCAGTCCTGCTTGCCACAGGTCT
TP53	7	M13R	TP53exon7M13R	CAGGAAACAGCTATGACTGATGAGAGGTGGATGGGTAG
TP53	8&9	M13F	TP53exon8&9M13F	GTAAAACGACGGCCAGTCAAGGGTGGTTGGGAGTAGA
TP53	8&9	M13R	TP53exon8&9M13R	CAGGAAACAGCTATGACCCCCAATTGCAGGTAAAACA

TP53	10	M13F	TP53exon10M13F	GTAAAACGACGGCCAGTGTCAGCTGTATAGGTACTTGAAGTGC
TP53	10	M13R	TP53exon10M13R	CAGGAAACAGCTATGACGAAGGCAGGATGAGAATGGA
TP53	11	M13F	TP53exon10M13F	GTAAAACGACGGCCAGTCCTTAGGCCCTTCAAAGCAT
TP53	11	M13R	TP53exon11M13R	CAGGAAACAGCTATGACAGACCCAAAACCCAAAATGG

F= forward, R=reverse

APPENDIX 2 LIST OF VARIANTS REPORTED BY TARGETED SEQUENCING

Sample ID	Gene	Transcript	Exon	Exonic function	Nucleotide	Amino Acid Change	Chr	Start (hg19)	End (hg19)	VAF	Depth (X)
ADM003	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.464	1188
ADM005	MED12	NM_005120	43	NS	C6386T	P2129L	X	70361198	70361198	0.984	62
ADM005	NOTCH1	NM_017617	34	Sgain	C7459T	Q2487X	9	139390732	139390732	0.438	870
ADM007	SAMHD1	NM_015474	5	NS	A617G	H206R	20	35559171	35559171	0.176	267
ADM007	SAMHD1	NM_015474	5	NS	T602A	I201N	20	35559186	35559186	0.492	267
ADM007	SAMHD1	NM_015474	11	NS	G1165A	A389T	20	35539726	35539726	0.081	885
ADM012	TP53	NM_001126115	3	NS	G317C	C106S	17	7577568	7577568	0.022	2912
ADM012	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.412	2764
ADM013	SF3B1	NM_012433	14	NS	G1998T	K666N	2	198267359	198267359	0.212	1007
ADM013	TP53	NM_001126115	6	NS	C613T	R205C	17	7574018	7574018	0.247	1999
ADM013	TP53	NM_001126115	2	Fdel	230_231del	R77del	17	7578222	7578223	0.334	2992
ADM015	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.139	5000
ADM020	NOTCH1	NM_017617	34	Sgain	C7318T	Q2440X	9	139390873	139390873	0.462	108
ADM020	TP53	NM_001276761	8	NA	556-2A>G	NA	17	7577610	7577610	0.315	1533
ADM020	TP53	NM_001126115	3	NS	C346T	R116W	17	7577539	7577539	0.444	3006
ADM027	TP53	NM_001126115	3	NS	G338T	G113V	17	7577547	7577547	0.666	1313
ADM028	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.528	1542
ADM028	TP53	NM_001126115	1	NS	G128A	R43H	17	7578406	7578406	0.458	845
ADM029	BIRC3	NM_001165	7	Fins	1539dupT	S513fs	11	102206910	102206910	0.354	2012
ADM030	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.174	3331
ADM041	ATM	NM_000051	59	NS	A8662C	I2888L	11	108218083	108218083	0.534	4006
ADM045	MED12	NM_005120	2	NS	G130T	G44C	X	70339253	70339253	0.492	1268
ADM068	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.454	163
ADM068	BIRC3	NM_001165	9	Fdel	1639delC	Q547fs	11	102207657	102207657	0.253	715
ADM069	SAMHD1	NM_015474	4	NS	G434A	R145Q	20	35563507	35563507	0.958	3162
ADM078	SF3B1	NM_012433	16	NS	A2328T	E776D	2	198266508	198266508	0.452	739
ADM080	ATM	NM_000051	29	NS	A4424G	Y1475C	11	108160516	108160516	0.474	587
ADM081	SF3B1	NM_012433	15	NS	A2110T	I704F	2	198266822	198266822	0.094	2822
ADM081	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.481	3642

ADM106	ATM	NM_000051	31	Fdel	4667delA	Y1556fs	11	108164095	108164095	0.302	1838
ADM107	ATM	NM_000051	20	NS	C2944T	R982C	11	108142000	108142000	0.088	692
ADM117	ATM	NM_000051	63	NS	C9022T	R3008C	11	108236086	108236086	0.722	870
ADM127	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.378	2564
ADM129	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.495	1827
ADM141	BIRC3	NM_001165	9	Fdel	1639delC	Q547fs	11	102207657	102207657	0.227	916
ADM141	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.115	1337
ADM148	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.384	801
ADM155	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.161	2345
ADM169	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.474	1020
ADM169	ATM	NM_000051	53	Sgain	G7913A	W2638X	11	108203613	108203613	0.95	1041
ADM181	TP53	NM_001126115	3	NS	C346T	R116W	17	7577539	7577539	0.933	1123
ADM207	TP53	NM_001126115	3	NS	G337A	G113S	17	7577548	7577548	0.092	3003
ADM207	SF3B1	NM_012433	14	NS	A1868G	Y623C	2	198267489	198267489	0.371	1913
ADM208	XPO1	NM_003400	15	NS	A1712G	E571G	2	61719471	61719471	0.333	1835
ADM210	SAMHD1	NM_015474	15	nFdel	1694_1705del	565_569del	20	35526266	35526277	0.935	1697
ADM216	TP53	NM_001276761	9	Splicing	802+1G>A	NA	17	7577018	7577018	0.087	270
ADM216	TP53	NM_001126115	6	nFdel	599_601del	R200_201del	17	7574030	7574032	0.263	369
ADM216	TP53	NM_001126115	5	Sgain	C595T	Q199X	17	7576855	7576855	0.32	636
ADM224	ATM	NM_000051	35	NS	T5291C	L1764P	11	108172488	108172488	0.857	252
ADM226	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.397	5000
ADM228	SAMHD1	NM_015474	15	Sgain	G1715A	W572X	20	35526256	35526256	0.982	2327
ADM242	TP53	NM_001276761	8	Fins	ND	ND	17	7577027	7577098	0.68	432
ADM247	BIRC3	NM_001165	9	Fdel	1663_1664del	555_555del	11	102207681	102207682	0.086	244
ADM247	BIRC3	NM_001165	9	nFdel	1706_1708del	569_570del	11	102207724	102207726	0.066	258
ADM257	SF3B1	NM_012433	14	NS	A1868G	Y623C	2	198267489	198267489	0.306	736
ADM269	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.478	1319
ADM272	ATM	NM_000051	41	NS	A6056G	Y2019C	11	108186599	108186599	0.12	937
ADM272	MED12	NM_005120	2	NS	T107G	L36R	X	70339230	70339230	0.491	751
ADM275	NOTCH1	NM_017617	34	NS	G6556A	G2186S	9	139391635	139391635	0.532	1299
ADM279	BIRC3	NM_001165	9	Fdel	1639delC	Q547fs	11	102207657	102207657	0.098	245
ADM279	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.118	1337
ADM287	SF3B1	NM_012433	14	NS	A2017G	I673V	2	198267340	198267340	0.224	519
ADM287	ATM	NM_000051	39	NS	A5882G	Y1961C	11	108181006	108181006	0.189	4361
ADM288	MYD88	NM_002468	5	NS	T478C	L265P	3	38182641	38182641	0.422	659
ADM293	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.196	1031

ADM293	TP53	NM_001126115	5	NS	G578T	G193V	17	7576872	7576872	0.502	960
ADM295	ATM	NM_000051	50	NS	T7469A	L2490H	11	108201102	108201102	0.928	1765
ADM321	SF3B1	NM_012433	16	NS	C2339G	P780R	2	198266497	198266497	0.095	539
ADM321	XPO1	NM_003400	15	NS	A1712C	E571A	2	61719471	61719471	0.394	1904
ADM321	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.07	2543
ADM324	ATM	NM_000051	41	NS	G6062A	C2021Y	11	108186605	108186605	0.493	796
ADM326	ATM	NM_000051	55	NS	T8058A	F2686L	11	108205743	108205743	0.185	271
ADM326	ATM	NM_000051	3	Fdel	73_74del	K25fs	11	108098501	108098504	0.394	1055
ADM326	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.471	1258
ADM332	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.401	461
ADM333	ATM	NM_000051	50	Fdel	7388-7393del	L2463fs	11	108201021	108201026	0.099	1542
ADM333	ATM	NM_000051	7	Fdel	702delA	A234fs	11	108115554	108115554	0.322	1685
ADM333	ATM	NM_000051	7	NS	T707G	L236R	11	108115559	108115559	0.324	1683
ADM333	ATM	NM_000051	56	NS	C8188A	Q2730K	11	108206608	108206608	0.359	468
ADM333	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.469	1997
ADM337	SF3B1	NM_012433	15	NS	G2219A	G740E	2	198266713	198266713	0.125	2088
ADM338	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.164	2156
ADM340	ATM	NM_000051	10	Fdel	1561_1562del	R521fs	11	108121753	108121754	0.511	1767
ADM354	SAMHD1	NM_015474	13	NS	T1478G	L493R	20	35532585	35532585	0.511	1137
ADM364	SF3B1	NM_012433	14	NS	G1998C	K666N	2	198267359	198267359	0.271	507
ADM364	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.384	2589
ADM366	ATM	NM_000051	50	nFdel	7406_7408del	N2469fs	11	108201039	108201041	0.59	1642
ADM373	BIRC3	NM_001165	9	NS	G1718A	G573D	11	102207736	102207736	0.476	679
ADM373	BIRC3	NM_001165	7	Fins	1492dupA	D497fs	11	102206865	102206865	0.09	2981
ADM384	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.064	2336
ADM399	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.272	2383
ADM401	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.131	4857
ADM407	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.476	4361
ADM427	ATM	NM_000051	37	Fdel	5599delC	Q1867fs	11	108175504	108175504	0.716	443
ADM441	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.44	2456
ADM446	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.466	4453
ADM446	MED12	NM_005120	2	NS	G131A	G44D	X	70339254	70339254	0.215	5000
ADM455	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.266	5000
ADM456	SF3B1	NM_012433	14	NS	C1873T	R625C	2	198267484	198267484	0.292	718
ADM456	BIRC3	NM_001165	9	Fdel	1667delC	T556fs	11	102207685	102207685	0.148	1046
ADM477	SF3B1	NM_012433	14	NS	G1866C	E622D	2	198267491	198267491	0.498	1263

ADM477	TP53	NM_001126115	4	NS	G457A	E153K	17	7577085	7577085	0.641	933
ARC017	TP53	NM_001126115	3	NS	C346T	R116W	17	7577539	7577539	0.023	787
ARC017	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.073	1596
ARC026	ATM	NM_000051	29	Fdel	A4292fs	N1431fs	11	108160384	108160384	0.098	2766
ARC026	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.489	3013
ARC033	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.478	1051
ARC037	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.287	1471
ARC038	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.486	3260
ARC040	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.075	777
ARC042	SF3B1	NM_012433	14	NS	A1996G	K666E	2	198267361	198267361	0.429	641
ARC042	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.42	1317
ARC042	ATM	NM_000051	57	NS	G8293A	G2765S	11	108213973	108213973	0.973	667
ARC043	NOTCH1	NM_017617	34	Sgain	7444delC	L2482X	9	139390747	139390747	0.144	721
ARC043	BIRC3	NM_001165	9	Fdel	1639delC	Q547fs	11	102207657	102207657	0.067	1216
ARC043	BIRC3	NM_001165	6	Fdel	1298_1299del	433_433del	11	102201946	102201947	0.122	1744
ARC047	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.343	2800
ARC079	TP53	NM_001126115	2	Fdel	230_231del	R77del	17	7578222	7578223	0.024	1819
ARC079	TP53	NM_001126115	1	NS	G121T	V41L	17	7578413	7578413	0.024	822
ARC079	TP53	NM_001126115	3	NS	C346G	R116G	17	7577539	7577539	0.029	787
ARC079	TP53	NM_001126115	3	NS	G337A	G113S	17	7577548	7577548	0.071	3465
ARC079	TP53	NM_001126115	3	NS	G347A	R116Q	17	7577538	7577538	0.101	3464
ARC079	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.387	2880
ARC082	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.089	853
ARC085	TP53	NM_001126115	2	NS	A218G	Y73C	17	7578235	7578235	0.576	507
ARC087	TP53	NM_001126115	3	NS	G347A	R116Q	17	7577538	7577538	0.803	5000
ARC091	ATM	NM_000051	47	NS	C6919T	L2307F	11	108196896	108196896	0.418	409
ARC091	TP53	NM_001126115	2	NS	A182T	H61L	17	7578271	7578271	0.045	815
ARC091	ATM	NM_000051	10	NS	T1450C	W484R	11	108121642	108121642	0.5	1407
ARC095	BIRC3	NM_001165	6	Fdel	1281_1285del	427_429del	11	102201929	102201933	0.071	1780
ARC098	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.312	820
ARC104	ATM	NM_000051	18	Fdel	2717_2718del	L906fs	11	108139215	108139216	0.216	706
ARC104	ATM	NM_000051	39	NS	C5858A	T1953K	11	108180982	108180982	0.448	1852
ARC108	SF3B1	NM_012433	14	NS	A1997C	K666T	2	198267360	198267360	0.215	719
ARC108	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.473	1772
ARC120	SF3B1	NM_012433	14	NS	A1996G	K666E	2	198267361	198267361	0.465	810
ARC125	ATM	NM_000051	62	Sgain	C8977T	R2993X	11	108235935	108235935	0.502	4212

ARC125	XPO1	NM_003400	15	NS	A1712T	E571V	2	61719471	61719471	0.522	2970
ARC137	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.341	5000
ARC138	ATM	NM_000051	10	NS	T1484A	I495K	11	108121676	108121676	0.892	557
ARC143	ATM	NM_000051	60	NS	G8672A	G2891D	11	108224493	108224493	0.076	1838
ARC143	SF3B1	NM_012433	14	NS	A1876C	N626H	2	198267481	198267481	0.202	1193
ARC143	SF3B1	NM_012433	15	nFdel	2094_2099del	698_700del	2	198266833	198266838	0.062	1763
ARC154	SF3B1	NM_012433	16	NS	G2302A	E768K	2	198266534	198266534	0.325	197
ARC157	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.908	1899
ARC159	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.381	2302
ARC160	SF3B1	NM_012433	15	NS	G2219A	G740E	2	198266713	198266713	0.16	238
ARC171	MYD88	NM_002468	5	NS	T478C	L265P	3	38182641	38182641	0.514	1252
ARC176	ATM	NM_000051	34	NS	A5071C	S1691R	11	108170506	108170506	0.529	276
ARC176	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.47	2597
ARC178	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.082	2991
ARC185	ATM	NM_000051	56	NS	G8161A	D2721N	11	108206581	108206581	0.941	665
ARC201	ATM	NM_000051	24	Sgain	G3505T	E1169X	11	108151824	108151824	0.763	457
ARC203	MYD88	NM_002468	5	NS	T478C	L265P	3	38182641	38182641	0.369	939
ARC221	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.096	1214
ARC227	SF3B1	NM_012433	14	NS	A1876G	N626D	2	198267481	198267481	0.461	684
ARC240	XPO1	NM_003400	15	NS	A1712G	E571G	2	61719471	61719471	0.127	535
ARC241	ATM	NM_000051	56	NS	A8189T	Q2730L	11	108206609	108206609	0.308	474
ARC241	SAMHD1	NM_015474	10	NS	G1112A	R371H	20	35540906	35540906	0.464	3633
ARC245	ATM	NM_000051	29	NS	A4424G	Y1475C	11	108160516	108160516	0.509	277
ARC245	TP53	NM_001126115	3	NS	T325C	S109P	17	7577560	7577560	0.43	4015
ARC252	NOTCH1	NM_017617	34	nFdel	7244_7246del	2415_2416del	9	139390945	139390947	0.07	185
ARC254	ATM	NM_000051	50	NS	G7342A	D2448N	11	108200975	108200975	0.961	1027
ARC274	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.161	2951
ARC280	TP53	NM_001276761	5	Splice site	258+1G>A	NA	17	7579311	7579311	0.546	1125
ARC286	MYD88	NM_002468	5	NS	T478C	L265P	3	38182641	38182641	0.455	1251
ARC292	ATM	NM_000051	27	Fdel	4096_4097del	C1366fs	11	108158429	108158430	0.138	551
ARC292	SF3B1	NM_012433	16	NS	C2323G	R775G	2	198266513	198266513	0.103	688
ARC299	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.078	553
ARC301	SF3B1	NM_012433	15	NS	A2110T	I704F	2	198266822	198266822	0.246	2440
ARC304	TP53	NM_001126115	1	NS	G64A	G22S	17	7578470	7578470	0.036	1057
ARC304	TP53	NM_001126115	2	nFdel	200_214del	E67_72del	17	7578239	7578253	0.844	902

ARC334	TP53	NM_001126115	4	NS	G422A	R141H	17	7577120	7577120	0.434	412
ARC344	TP53	NM_001276761	7	NA	555+1G>T	NA	17	7578176	7578176	0.659	3735
ARC393	TP53	NM_001126115	3	NS	G347A	R116Q	17	7577538	7577538	0.026	4835
ARC393	NOTCH1	NM_017617	34	NS	G7606A	V2536I	9	139390585	139390585	0.513	847
ARC417	TP53	NM_001126115	2	NS	T249G	S83R	17	7578204	7578204	0.42	997
ARC484	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.513	2778
ARC484	TP53	NM_001276761	9	NA	666-2A>T	NA	17	7577157	7577157	0.87	624
ARC502	BIRC3	NM_001165	9	Fdel	1622-9_1639del	541_547del	11	102207612	102207657	0.24	560
ARC528	ATM	NM_000051	61	NS	C8840A	T2947N	11	108225591	108225591	0.432	1134
ARC528	ATM	NM_000051	3	Fdel	73_74del	K25fs	11	108098501	108098504	0.455	2099
ARC554	ATM	NM_000051	61	Fdel	8831_8832del	T2944fs	11	108225582	108225583	0.078	204
ARC554	ATM	NM_000051	46	NS	A6787G	R2263G	11	108196251	108196251	0.814	553
ARC566	ATM	NM_000051	10	Fdel	1561_1562del	R521fs	11	108121753	108121754	0.49	923
ARC574	ATM	NM_000051	19	Splice site	2921+1G>A	NA	11	108141874	108141874	0.805	1241
ARC576	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.494	3351
ARC579	SF3B1	NM_012433	14	NS	C1984G	H662D	2	198267373	198267373	0.088	113
ARC582	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.097	2730
ARC592	TP53	NM_001126115	5	Sgain	C553T	Q185X	17	7576897	7576897	0.015	6708
ARC592	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.382	4996
ARC601	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.468	2352
ARC602	MED12	NM_005120	2	NS	G130A	G44S	X	70339253	70339253	0.39	5000
ARC603	TP53	NM_001126115	2	Sgain	C241T	R81X	17	7578212	7578212	0.823	1839
ARC604	ATM	NM_000051	63	Sgain	C9139T	R3047X	11	108236203	108236203	0.442	481
ARC604	ATM	NM_000051	11	NA	1802+1G>A	NA	11	108122759	108122759	0.461	2500
ARC604	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.462	2596
ARC609	NOTCH1	NM_017617	34	Sgain	G7516T	E2506X	9	139390675	139390675	0.366	3018
ARC621	SAMHD1	NM_015474	8	Fdel	884delG	S295fs	20	35545421	35545421	0.448	2010
ARC627	BIRC3	NM_001165	6	Fdel	1281_1285del	427_429del	11	102201929	102201933	0.351	3269
ARC631	BIRC3	NM_001165	9	NS	A1774T	R592W	11	102207792	102207792	0.267	327
ARC631	BIRC3	NM_001165	6	Fdel	1319_1322del	440_441del	11	102201967	102201970	0.117	435
ARC634	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.463	1912
ARC639	BIRC3	NM_001165	9	nFdel	1796_1807del	600_603del	11	102207817	102207825	0.075	1315
ARC639	BIRC3	NM_001165	6	Fdel	1298_1299del	433_433del	11	102201946	102201947	0.075	1879
ARC639	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.463	2914
ARC643	BIRC3	NM_001165	9	Fdel	1654delC	Q552fs	11	102207672	102207672	0.279	936

ARC643	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.28	2054
ARC646	ATM	NM_000051	56	NS	A8264T	Y2755F	11	108206684	108206684	0.092	685
ARC646	ATM	NM_000051	59	NS	T8627A	I2876N	11	108218048	108218048	0.119	1381
ARC646	ATM	NM_000051	58	Fdel	8478delT	N2826fs	11	108216529	108216529	0.188	2208
ARC649	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.439	1998
ARC663	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.432	2042
ARC663	SAMHD1	NM_015474	16	NS	G1753A	D585N	20	35521463	35521463	0.513	4998
ARC679	BIRC3	NM_001165	9	NS	A1672G	K558E	11	102207690	102207690	0.404	312
ARC679	ATM	NM_000051	7	NS	G743A	R248Q	11	108115595	108115595	0.921	1463
ARC679	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.207	2632
ARC682	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.076	2157
ARC683	MED12	NM_005120	23	NS	T3341C	L1114P	X	70348277	70348277	0.095	63
ARC683	SF3B1	NM_012433	14	NS	A1997C	K666T	2	198267360	198267360	0.498	424
ARC683	TP53	NM_001126115	4	NS	T433G	C145G	17	7577109	7577109	0.03	1757
ARC683	TP53	NM_001126115	3	NS	G317T	C106F	17	7577568	7577568	0.034	5609
ARC689	ATM	NM_000051	23	Fdel	3334delC	P1112fs	11	108150267	108150267	0.2	2474
ARC689	ATM	NM_000051	20	Splice site	3077_3077+18del	NA	11	108142134	108142151	0.207	972
ARC689	SF3B1	NM_012433	16	NS	G2225A	G742D	2	198266611	198266611	0.11	789
ARC689	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.484	2782
ARC714	BIRC3	NM_001165	9	Fdel	1639_1642del	547_548del	11	102207657	102207660	0.253	392
ARC716	SF3B1	NM_012433	14	NS	A1996G	K666E	2	198267361	198267361	0.295	259
ARC716	XPO1	NM_003400	15	NS	G1711A	E571K	2	61719472	61719472	0.077	1263
ARC716	NOTCH1	NM_017617	34	Fdel	7541_7542del	2514_2514del	9	139390649	139390650	0.31	1562
ARC724	SF3B1	NM_012433	15	NS	A2098G	K700E	2	198266834	198266834	0.123	2345

Fdel= frameshift deletion, NS= non-synonymous variant, Sgain= stop gain variant, NA= not applicable, Fins= frameshift insertion, nFdel : non-frameshift deletion, Splicing= variant in splicing site.

APPENDIX 3 LIST OF COPY NUMBER VARIANTS REPORTED BY SNP ARRAY

Sample ID	Chr	Start (hg19)	End (hg19)	Event	Length (b)
ADM003	2	77488458	77885374	Heterozygous gain	396,916
ADM005	12	0	133851895	Heterozygous gain	133,851,895
ADM006	4	117379988	117507824	Heterozygous loss	127,836
ADM006	12	11717490	13294352	Heterozygous loss	1,576,862
ADM006	13	46921166	51475046	Heterozygous loss	4,553,880
ADM006	16	3567098	3903064	Heterozygous loss	335,966
ADM007	13	48319303	64453778	Heterozygous loss	16,134,475
ADM007	20	29790091	62942535	cnLOH	33,152,444
ADM009	13	34353059	115169878	cnLOH	80,816,819
ADM009	13	50588647	51354718	Heterozygous loss	766,071
ADM009	13	50588647	51354718	Heterozygous loss	766,071
ADM009	13	50588507	50997647	Homozygous loss	409,140
ADM010	11	94518394	114510681	Heterozygous loss	19,992,287
ADM011	1	26904304	28380290	Heterozygous loss	1,475,986
ADM011	9	139624304	140109546	Heterozygous gain	485,242
ADM011	13	50600738	51324476	Homozygous loss	4,609,713
ADM011	13	47215456	51825169	Heterozygous loss	4,609,713
ADM013	4	96962004	97078945	Heterozygous loss	116,941
ADM013	8	32961398	33709007	Heterozygous gain	747,609
ADM013	8	100274743	146364022	Heterozygous gain	46,089,279
ADM013	13	37093572	115169878	cnLOH	78,076,306
ADM013	13	50484922	51629365	Heterozygous loss	1,144,443
ADM013	18	58956080	66404208	Heterozygous gain	7,448,128
ADM013	18	66404208	77790774	Heterozygous loss	11,386,566
ADM014	11	84068686	116211943	Heterozygous loss	32,143,257
ADM014	12	0	133851895	Heterozygous gain	133,851,895

ADM015	11	106555865	119424481	Heterozygous loss	12,868,616
ADM015	13	48517331	51640030	Heterozygous loss	3,122,699
ADM020	1	226842479	227326877	Heterozygous loss	484,398
ADM020	2	231174025	231440365	Heterozygous loss	266,340
ADM020	7	138623740	140166136	Heterozygous loss	1,542,396
ADM020	10	63665741	63991293	Heterozygous loss	325,552
ADM020	10	104161348	105488031	Heterozygous loss	1,326,683
ADM020	13	20798053	21120811	Heterozygous loss	322,758
ADM020	14	69256151	106186158	Heterozygous loss	36,930,007
ADM020	17	35643513	36018200	Heterozygous loss	374,687
ADM020	17	55929739	57107422	Heterozygous loss	1,177,683
ADM027	9	14179148	29984310	Heterozygous loss	15,805,162
ADM027	10	99740245	119320412	Heterozygous loss	19,580,167
ADM027	10	120276099	135534747	Heterozygous loss	15,258,648
ADM027	13	19097765	115169878	cnLOH	96,072,113
ADM027	13	50702562	51480715	Heterozygous loss	778,153
ADM027	17	0	19014529	Heterozygous loss	19,014,529
ADM027	17	19136888	20228737	Heterozygous loss	1,091,849
ADM027	17	21685023	25648982	Heterozygous loss	3,963,959
ADM027	21	29609532	48113476	Heterozygous gain	18,503,944
ADM028	2	0	89250931	Heterozygous gain	89,250,931
ADM028	2	63437646	64225667	Heterozygous loss	788,021
ADM028	2	108240807	108371033	Heterozygous loss	130,226
ADM028	2	136540546	138354229	Heterozygous loss	1,813,683
ADM028	2	138543155	145507150	Heterozygous loss	6,963,995
ADM028	2	168440734	169545106	Heterozygous loss	1,104,372
ADM028	2	194085775	197078057	Heterozygous loss	2,992,282
ADM028	2	198025919	198146102	Heterozygous loss	120,183
ADM028	2	223552672	228514661	Heterozygous loss	4,961,989
ADM028	2	232802655	233184867	Heterozygous loss	382,212
ADM028	2	233397450	235059128	Heterozygous loss	1,661,678
ADM028	2	235139592	235248991	Heterozygous loss	109,399
ADM028	2	237633549	237811040	Heterozygous loss	177,491
ADM028	2	237887460	238657659	Heterozygous loss	770,199
ADM028	3	47127640	50092979	Heterozygous loss	2,965,339
ADM028	6	122037144	122582515	Heterozygous gain	545,371

ADM028	8	0	43573131	Heterozygous gain	43,573,131
ADM028	13	49731455	51677296	Heterozygous loss	1,945,841
ADM028	18	62268893	62535524	Heterozygous gain	266,631
ADM029	11	87074878	116177100	Heterozygous loss	29,102,222
ADM030	18	70291527	71460615	Heterozygous loss	1,169,088
ADM032	2	111392259	113195775	Heterozygous loss	1,803,516
ADM032	13	50400102	51608024	Heterozygous loss	1,207,922
ADM032	16	4104130	4257929	Heterozygous gain	153,799
ADM034	3	163234753	163933648	Heterozygous loss	698,895
ADM034	11	101857674	115731562	Heterozygous loss	13,873,888
ADM034	13	48650246	48768133	Heterozygous loss	117,887
ADM034	13	48865610	49190737	Heterozygous loss	325,127
ADM034	13	49824565	52362223	Heterozygous loss	2,537,658
ADM041	2	0	65145551	Heterozygous gain	65,145,551
ADM041	12	14262398	19392422	Heterozygous loss	5,130,024
ADM041	12	19379465	26817368	Heterozygous gain	7,437,903
ADM041	13	50601924	51388481	Homozygous loss	786,557
ADM041	13	49967327	115169878	Heterozygous loss	65,202,551
ADM041	15	20005287	42260090	Heterozygous loss	22,254,803
ADM045	12	0	133851895	Heterozygous gain	133,851,895
ADM058	4	39286732	39438706	Heterozygous loss	151,974
ADM058	11	78159475	117282724	Heterozygous loss	39,123,249
ADM059	5	86205934	86417000	Heterozygous loss	211,066
ADM059	9	5096592	5229166	Heterozygous gain	132,574
ADM059	13	50599872	51477169	Homozygous loss	877,297
ADM059	13	46816955	55344415	Heterozygous loss	8,527,460
ADM059	18	21424795	21555903	Heterozygous loss	131,108
ADM068	9	23927587	24051248	Heterozygous loss	123,661
ADM068	15	42751496	102531392	cnLOH	59,779,896
ADM069	11	92678700	128133510	Heterozygous loss	35,454,810
ADM069	11	128738462	128990487	Heterozygous loss	252,025
ADM069	13	49858362	52458431	Heterozygous loss	2,600,069
ADM069	20	34013689	63025520	cnLOH	29,011,831
ADM069	X	42261597	49412390	Heterozygous gain	7,150,793
ADM070	2	86086000	89131265	Heterozygous loss	3,045,265
ADM070	9	70961350	76162559	Heterozygous loss	5,201,209

ADM070	14	69262813	107073785	Heterozygous loss	37,810,972
ADM078	13	50425097	51657349	Heterozygous loss	1,232,252
ADM080	8	11191170	11609324	Heterozygous loss	418,154
ADM081	8	59807358	61936502	Heterozygous loss	2,129,144
ADM081	13	49568025	51587721	Homozygous loss	2,019,696
ADM081	13	41250464	49568025	Heterozygous loss	8,317,561
ADM081	16	10336824	12093445	Heterozygous loss	1,756,621
ADM096	3	126927688	197891568	Heterozygous gain	70,963,880
ADM096	15	91134224	91386042	Heterozygous loss	251,818
ADM096	15	91466596	102531392	Heterozygous loss	11,064,796
ADM106	13	48898646	52454960	Heterozygous loss	3,556,314
ADM106	X	1720919	2068907	Heterozygous gain	347,988
ADM107	2	0	40349459	Heterozygous gain	40,349,459
ADM107	11	82770109	135006516	Heterozygous loss	52,236,407
ADM107	13	49120578	51481460	Heterozygous loss	2,360,882
ADM107	16	15485286	18001331	Heterozygous gain	2,516,045
ADM113	13	47268255	51479537	Heterozygous loss	4,211,282
ADM117	2	230712748	230903085	Heterozygous gain	190,337
ADM117	8	103042314	146364022	Heterozygous gain	43,321,708
ADM117	11	66361322	134661415	cnLOH	68,300,093
ADM117	13	48670735	51623232	Heterozygous loss	2,952,497
ADM117	16	67319165	90354753	cnLOH	23,035,588
ADM119	1	151173090	156916109	Heterozygous gain	5,743,019
ADM119	13	48695280	52338614	Heterozygous loss	3,643,334
ADM127	13	48557371	51665376	Heterozygous loss	3,108,005
ADM129	13	50508342	51989359	Heterozygous loss	1,481,017
ADM129	15	42437765	102531392	cnLOH	60,093,627
ADM131	13	49508724	51613855	Heterozygous loss	2,105,131
ADM140	X	488478	770003	Heterozygous gain	281,525
ADM141	7	139030982	140046075	Heterozygous loss	1,015,093
ADM141	7	140042642	159131596	Heterozygous gain	19,088,954
ADM141	11	92998352	114449700	Heterozygous loss	21,451,348
ADM141	13	25088843	49194471	cnLOH	24,105,628
ADM141	13	48683748	115169878	Heterozygous loss	66,486,130
ADM148	13	50574980	52988581	Heterozygous loss	2,413,601
ADM148	14	19000480	107349540	cnLOH	88,349,060

ADM148	X	117627605	118169870	Heterozygous loss	542,265
ADM151	13	49951112	51598443	Heterozygous loss	1,647,331
ADM153	13	47409041	51501814	Heterozygous loss	4,092,773
ADM169	11	77498953	135006516	cnLOH	57,507,563
ADM169	13	50304747	51630902	Heterozygous loss	1,326,155
ADM169	13	50540730	50700986	Homozygous loss	160,256
ADM181	4	0	18767213	Heterozygous loss	18,767,213
ADM181	5	149265604	150843326	Heterozygous loss	1,577,722
ADM181	5	150857642	180853771	Heterozygous gain	29,996,129
ADM181	16	46412605	65705806	Heterozygous loss	19,293,201
ADM181	16	87888588	90339646	Heterozygous gain	2,451,058
ADM181	17	0	16368436	Heterozygous loss	16,368,436
ADM181	21	14338959	19602076	Heterozygous loss	5,263,117
ADM181	21	20215461	27532240	Heterozygous loss	7,316,779
ADM182	1	111593142	111966488	Heterozygous loss	373,346
ADM182	13	50464191	51632119	Heterozygous loss	1,167,928
ADM182	13	50601924	51388481	Homozygous loss	786,557
ADM204	2	0	90275455	Heterozygous gain	90,275,455
ADM204	20	0	26316921	Heterozygous loss	26,316,921
ADM207	13	50579475	51482740	Heterozygous loss	903,265
ADM208	13	50510351	51542115	Heterozygous loss	1,031,764
ADM210	4	127851514	130778196	Heterozygous gain	2,926,682
ADM210	13	46703229	51565628	Heterozygous loss	4,862,399
ADM210	20	29804016	63025520	cnLOH	33,221,504
ADM211	13	19295745	115169878	cnLOH	95,874,133
ADM211	13	50503797	51496705	Homozygous loss	992,908
ADM215	13	50071154	52476282	Heterozygous loss	2,405,128
ADM215	17	6748275	6862345	Heterozygous loss	114,070
ADM216	17	0	20515810	Heterozygous loss	20,515,810
ADM216	20	0	22886483	Heterozygous loss	22,886,483
ADM217	10	6681460	6841868	Heterozygous gain	160,408
ADM217	13	48669758	48907632	Heterozygous loss	237,874
ADM217	13	50508342	51641773	Heterozygous loss	1,133,431
ADM219	12	0	133851895	Heterozygous gain	133,851,895
ADM219	14	69258450	105811340	Heterozygous loss	36,552,890
ADM223	1	73793122	73895793	Heterozygous loss	102,671

ADM223	4	147421971	147653064	Heterozygous loss	231,093
ADM223	6	149636	771446	Heterozygous gain	621,810
ADM223	8	20481559	20835160	Heterozygous loss	353,601
ADM223	15	90898304	102531392	Heterozygous loss	11,633,088
ADM224	8	115641560	117316016	Heterozygous loss	1,674,456
ADM224	11	84890771	112990425	Heterozygous loss	28,099,654
ADM224	15	78677044	80864260	Heterozygous loss	2,187,216
ADM224	18	47269590	49839910	Heterozygous loss	2,570,320
ADM224	18	50746306	52059065	Heterozygous loss	1,312,759
ADM226	13	49414237	51709242	Heterozygous loss	2,295,005
ADM228	11	102190527	114051661	Heterozygous loss	11,861,134
ADM228	15	30917982	59379646	Heterozygous loss	28,461,664
ADM228	20	29423716	43306458	cnLOH	13,882,742
ADM229	5	116126056	116302947	Heterozygous gain	176,891
ADM229	9	15718207	15852435	Heterozygous loss	134,228
ADM229	13	46927506	51886770	Heterozygous loss	4,959,264
ADM238	12	0	133851895	Heterozygous gain	133,851,895
ADM238	19	0	59128983	Heterozygous gain	59,128,983
ADM242	2	0	14109321	Heterozygous gain	14,109,321
ADM242	2	15815293	26372607	Heterozygous gain	10,557,314
ADM242	2	31215979	48202191	Heterozygous gain	16,986,212
ADM242	2	53831486	59208932	Heterozygous gain	5,377,446
ADM242	2	59184385	72552246	Heterozygous gain	13,367,861
ADM242	2	85597893	87123114	Heterozygous gain	1,525,221
ADM242	2	216979765	219234964	Heterozygous loss	2,255,199
ADM242	2	220037270	228247659	Heterozygous gain	8,210,389
ADM242	8	0	41559699	Heterozygous loss	41,559,699
ADM242	8	49233618	49388771	Heterozygous loss	155,153
ADM242	10	0	5487351	Heterozygous loss	5,487,351
ADM242	11	56175463	91250773	Heterozygous gain	35,075,310
ADM242	11	91250773	131857154	Heterozygous loss	40,606,381
ADM242	11	131857154	135000002	Heterozygous gain	3,142,848
ADM242	13	41267301	51524894	Heterozygous loss	10,257,593
ADM242	17	63924952	64314690	Heterozygous loss	389,738
ADM247	12	0	133851895	Heterozygous gain	133,851,895
ADM247	18	0	78077248	Heterozygous gain	78,077,248

ADM247	19	0	59128983	Heterozygous gain	59,128,983
ADM249	21	44795444	48129895	Heterozygous gain	3,334,451
ADM257	13	48810641	51521850	Heterozygous loss	2,711,209
ADM257	15	73150942	73546163	Heterozygous gain	395,221
ADM259	13	50440726	72006388	Heterozygous loss	21,565,662
ADM269	11	83631914	123862929	Heterozygous loss	40,231,015
ADM271	12	0	133851895	Heterozygous gain	133,851,895
ADM271	13	19979462	115169878	cnLOH	95,190,416
ADM272	12	0	133851895	Heterozygous gain	133,851,895
ADM275	13	50212411	51712610	Heterozygous loss	1,500,199
ADM279	12	0	133851895	Heterozygous gain	133,851,895
ADM279	13	50467529	51528518	Heterozygous loss	1,060,989
ADM279	18	0	78077248	Heterozygous gain	78,077,248
ADM279	19	0	59075809	Heterozygous gain	59,075,809
ADM283	13	50540730	52406197	Heterozygous loss	1,865,467
ADM283	17	6756523	6869318	Heterozygous loss	112,795
ADM284	13	49012672	51474262	Homozygous loss	2,461,590
ADM284	13	19971213	115051024	Heterozygous loss	95,079,811
ADM284	X	23498576	23913715	Heterozygous gain	415,139
ADM287	6	109333063	109501920	Heterozygous loss	168,857
ADM287	11	94516526	116511101	Heterozygous loss	21,994,575
ADM287	18	12248937	14411093	Heterozygous loss	2,162,156
ADM288	11	4401788	4590103	Heterozygous loss	188,315
ADM288	11	116803917	117029397	Heterozygous loss	225,480
ADM288	12	46735486	50911022	Heterozygous loss	4,175,536
ADM288	13	50450955	50701435	Heterozygous loss	250,480
ADM288	14	63578712	66754867	Heterozygous loss	3,176,155
ADM288	14	75540929	76165874	Heterozygous loss	624,945
ADM288	14	91246360	93693995	Heterozygous loss	2,447,635
ADM288	18	0	78077248	Heterozygous gain	78,077,248
ADM293	4	30083700	30313010	Heterozygous loss	229,310
ADM293	7	22645977	23109216	Heterozygous loss	463,239
ADM293	13	50020862	51699432	Heterozygous loss	1,678,570
ADM295	11	95428103	115222932	Heterozygous loss	19,794,829
ADM295	13	19167959	21561299	Heterozygous loss	2,393,340
ADM295	13	46927506	50679794	Heterozygous loss	3,752,288

ADM296	4	86400114	86649706	Heterozygous loss	249,592
ADM296	13	50255235	51572757	Heterozygous loss	1,317,522
ADM318	1	81982516	84159597	Heterozygous loss	2,177,081
ADM318	3	46406718	48227837	Heterozygous loss	1,821,119
ADM318	6	0	44831407	Heterozygous gain	44,831,407
ADM318	6	44831407	45359839	Heterozygous loss	528,432
ADM318	6	46040182	51724224	Heterozygous gain	5,684,042
ADM318	6	57193830	62882536	Heterozygous loss	5,688,706
ADM318	6	62882536	63103238	Heterozygous gain	220,702
ADM318	6	63103238	72354323	Heterozygous loss	9,251,085
ADM318	6	72354323	85081974	Heterozygous gain	12,727,651
ADM318	6	91977739	92099166	Heterozygous loss	121,427
ADM318	6	92381047	156287941	Heterozygous loss	63,906,894
ADM318	6	157013130	157159560	Heterozygous loss	146,430
ADM318	6	163819745	171115067	Heterozygous loss	7,295,322
ADM318	7	63416982	64660976	Heterozygous loss	1,243,994
ADM318	7	64660976	64840456	Heterozygous gain	179,480
ADM318	7	64840456	65092316	Heterozygous loss	251,860
ADM318	7	65092316	67702765	Heterozygous gain	2,610,449
ADM318	7	67702765	68190947	Heterozygous loss	488,182
ADM318	7	68190947	68494133	Heterozygous gain	303,186
ADM318	7	68494133	69890750	Heterozygous loss	1,396,617
ADM318	7	113651659	115770646	Heterozygous loss	2,118,987
ADM318	7	115770646	116199682	Heterozygous gain	429,036
ADM318	7	116199682	159138663	Heterozygous loss	42,938,981
ADM318	9	0	33634769	Heterozygous loss	33,634,769
ADM318	12	7401023	13539813	Heterozygous loss	6,138,790
ADM318	12	121644530	124043314	Heterozygous loss	2,398,784
ADM320	4	43537072	43660029	Heterozygous gain	122,957
ADM320	13	19167959	115169878	cnLOH	96,001,919
ADM320	13	50506792	51537592	Homozygous loss	1,030,800
ADM321	2	97184612	97881267	Heterozygous loss	696,655
ADM321	13	48365397	51916546	Heterozygous loss	3,551,149
ADM321	13	59947829	80616281	cnLOH	20,668,452
ADM324	2	0	89131265	Heterozygous gain	89,131,265
ADM324	5	110309088	110423617	Heterozygous loss	114,529

ADM324	13	28680840	115169878	cnLOH	86,489,038
ADM324	13	49853633	51675055	Homozygous loss	1,821,422
ADM324	18	0	15400035	Heterozygous loss	15,400,035
ADM326	3	587676	6168074	Heterozygous loss	5,580,398
ADM326	11	65381226	134918837	Heterozygous gain	69,537,611
ADM326	13	50574980	51489153	Homozygous loss	914,173
ADM326	13	46834864	69243872	Heterozygous loss	22,409,008
ADM332	2	58226966	61741299	Heterozygous gain	3,514,333
ADM333	11	102205247	115622767	Heterozygous loss	13,417,520
ADM335	13	19322835	115049316	cnLOH	95,726,481
ADM335	13	51181705	51484913	Heterozygous loss	303,208
ADM335	13	50530223	51181705	Homozygous loss	651,482
ADM335	18	465212	684186	Heterozygous gain	218,974
ADM338	2	230595935	231741548	Heterozygous loss	1,145,613
ADM339	2	37279261	38487288	Heterozygous loss	1,208,027
ADM339	13	46753347	47279356	Heterozygous loss	526,009
ADM339	13	50488424	51423769	Homozygous loss	935,345
ADM339	13	49978529	51680099	Heterozygous loss	1,701,570
ADM340	3	46697394	48722199	Heterozygous loss	2,024,805
ADM340	3	124292641	124398800	Heterozygous loss	106,159
ADM340	10	104158981	104964128	Heterozygous loss	805,147
ADM340	13	50418048	72006388	Heterozygous loss	21,588,340
ADM342	7	86192037	87988002	Heterozygous loss	1,795,965
ADM342	7	89017685	89795072	Heterozygous loss	777,387
ADM342	7	92634503	95094234	Heterozygous loss	2,459,731
ADM342	7	107109007	126584321	Heterozygous loss	19,475,314
ADM350	8	19043956	19505048	Heterozygous gain	461,092
ADM350	9	0	400033	Heterozygous gain	400,033
ADM350	12	114260646	114560046	Heterozygous gain	299,400
ADM350	13	32892640	66380902	Heterozygous loss	33,488,262
ADM354	13	49751841	51935306	Heterozygous loss	2,183,465
ADM354	18	51063439	78008759	cnLOH	26,945,320
ADM359	14	69256151	107349540	Heterozygous loss	38,093,389
ADM364	13	49464162	51730052	Heterozygous loss	2,265,890
ADM366	11	79484470	89408083	Heterozygous loss	9,923,613
ADM366	21	0	48129895	Heterozygous gain	48,129,895

ADM366	21	17325955	17509863	Heterozygous loss	183,908
ADM373	10	6680890	6836560	Heterozygous gain	155,670
ADM373	11	76480376	123460951	Heterozygous loss	46,980,575
ADM384	13	50681963	52593904	Heterozygous loss	1,911,941
ADM391	13	50584367	51496705	Heterozygous loss	912,338
ADM391	13	48840454	49846413	Heterozygous loss	1,005,959
ADM391	16	88626687	89980058	Heterozygous loss	1,353,371
ADM398	11	76747454	114645502	Heterozygous loss	37,898,048
ADM398	13	46910907	51495892	Heterozygous loss	4,584,985
ADM399	13	46943036	63391011	Heterozygous loss	16,447,975
ADM401	6	140196360	140392319	Heterozygous gain	195,959
ADM404	12	0	133851895	Heterozygous gain	133,851,895
ADM407	13	48826435	52751336	Heterozygous loss	3,924,901
ADM407	X	87876014	155270560	cnLOH	67,394,546
ADM411	12	0	133851895	Heterozygous gain	133,851,895
ADM411	15	50909805	51035439	Heterozygous loss	125,634
ADM426	12	0	133851895	Heterozygous gain	133,851,895
ADM427	11	78281118	116683618	Heterozygous loss	38,402,500
ADM427	13	47313075	67803820	Heterozygous loss	20,490,745
ADM438	12	0	133851895	Heterozygous gain	133,851,895
ADM438	14	69259432	106327594	Heterozygous loss	37,068,162
ADM438	X	802528	1752567	Heterozygous gain	950,039
ADM441	1	93533372	105207577	cnLOH	11,674,205
ADM441	1	160340544	161509392	Heterozygous loss	1,168,848
ADM441	1	234565431	235389392	Heterozygous loss	823,961
ADM441	13	50592408	51089955	Homozygous loss	497,547
ADM441	13	38354526	63617135	Heterozygous loss	25,262,609
ADM446	2	198373367	231680988	Heterozygous loss	33,307,621
ADM455	2	0	67239895	Heterozygous gain	67,239,895
ADM455	2	87279109	87628556	Heterozygous gain	349,447
ADM455	8	6233832	6337916	Heterozygous gain	104,084
ADM455	18	0	15400035	Heterozygous loss	15,400,035
ADM456	2	240632721	243199373	Heterozygous loss	2,566,652
ADM456	4	0	18947065	Heterozygous loss	18,947,065
ADM456	8	93437435	102533324	Heterozygous gain	9,095,889
ADM456	8	125055709	125690170	Heterozygous gain	634,461

ADM456	11	98387732	117646730	Heterozygous loss	19,258,998
ADM466	9	9698688	9934248	Heterozygous gain	235,560
ADM466	13	50603980	51637766	Heterozygous loss	1,033,786
ADM477	3	0	232795	Heterozygous loss	232,795
ADM477	8	93742900	146364022	Heterozygous gain	52,621,122
ADM477	13	49759449	51741697	Heterozygous loss	1,982,248
ADM477	13	95490389	115169878	Heterozygous loss	19,679,489
ADM477	17	0	18928388	cnLOH	18,928,388
ADM477	X	133082352	155270560	Heterozygous gain	22,188,208
ADM480	6	9255162	28647416	Heterozygous loss	19,392,254
ADM480	8	0	36209019	Heterozygous loss	36,209,019
ADM480	8	84167811	146364022	Heterozygous gain	62,196,211
ADM480	12	0	133851895	Heterozygous gain	133,851,895
ARC017	6	155374052	155553842	Heterozygous loss	179,790
ARC017	12	0	133851895	Heterozygous gain	133,851,895
ARC019	4	31534013	32529763	Heterozygous gain	995,750
ARC019	13	50588507	51414166	Heterozygous loss	825,659
ARC026	11	70157890	135006516	cnLOH	64,848,626
ARC031	11	76533209	126528088	Heterozygous loss	49,994,879
ARC031	13	50796740	51080462	Heterozygous loss	283,722
ARC033	5	88180862	118336938	Heterozygous loss	30,156,076
ARC033	11	103768480	116696801	Heterozygous loss	12,928,321
ARC033	15	41888607	102531392	cnLOH	60,642,785
ARC037	4	0	23456630	Heterozygous loss	23,456,630
ARC037	18	7134326	7486140	Heterozygous gain	351,814
ARC037	20	43381581	44741781	Heterozygous loss	1,360,200
ARC037	20	48715064	50707568	Heterozygous loss	1,992,504
ARC038	1	228187352	228700321	Heterozygous gain	512,969
ARC038	13	48512564	52063707	Heterozygous loss	3,551,143
ARC039	3	150626625	171622454	cnLOH	20,995,829
ARC039	5	134331847	146495112	cnLOH	12,163,265
ARC039	12	0	133851895	Heterozygous gain	133,851,895
ARC039	12	24560330	54957729	cnLOH	30,397,399
ARC040	2	185514156	185801578	Heterozygous loss	287,422
ARC040	13	50532211	51532183	Heterozygous loss	999,972
ARC042	1	46267006	46488466	Heterozygous gain	221,460

ARC042	1	89551479	89959124	Heterozygous loss	407,645
ARC042	1	117000082	117691995	Heterozygous loss	691,913
ARC042	2	214643726	214799916	Heterozygous loss	156,190
ARC042	10	50655778	50780608	Heterozygous loss	124,830
ARC042	10	51734843	52544426	Heterozygous loss	809,583
ARC042	11	107925606	108647072	Heterozygous loss	721,466
ARC042	12	123134802	123243147	Heterozygous loss	108,345
ARC042	13	41396640	41831907	Heterozygous loss	435,267
ARC042	13	42440038	43306463	Heterozygous loss	866,425
ARC042	13	48840454	52734999	Heterozygous loss	3,894,545
ARC043	3	148828990	148968248	Heterozygous loss	139,258
ARC043	11	104875043	117338197	Heterozygous loss	12,463,154
ARC043	12	0	133851895	Heterozygous gain	133,851,895
ARC047	13	41428764	43141393	Heterozygous loss	1,712,629
ARC047	13	50191759	52790700	Heterozygous loss	2,598,941
ARC052	13	114740209	115169878	Heterozygous gain	429,669
ARC052	20	48917835	50104554	Heterozygous loss	1,186,719
ARC052	22	23699954	25016180	Heterozygous gain	1,316,226
ARC053	12	0	133851895	Heterozygous gain	133,851,895
ARC065	11	94809270	96287529	Heterozygous loss	1,478,259
ARC065	13	49828462	49949301	Heterozygous loss	120,839
ARC065	13	50081759	51669653	Heterozygous loss	1,587,894
ARC071	13	50586838	51562506	Heterozygous loss	975,668
ARC077	11	79019270	114618601	Heterozygous loss	35,599,331
ARC082	12	0	133851895	Heterozygous gain	133,851,895
ARC085	1	0	27331645	Heterozygous gain	27,331,645
ARC085	2	10402736	16252614	Heterozygous gain	5,849,878
ARC085	2	37218864	39257051	Heterozygous loss	2,038,187
ARC085	2	40799142	48851276	Heterozygous loss	8,052,134
ARC085	2	70344367	87123114	Heterozygous loss	16,778,747
ARC085	9	0	2720504	Heterozygous loss	2,720,504
ARC085	9	6691733	103588431	Heterozygous loss	96,896,698
ARC085	10	27077859	39154535	Heterozygous loss	12,076,676
ARC085	13	87841682	96145617	Heterozygous loss	8,303,935
ARC085	14	34832611	39567838	Heterozygous loss	4,735,227
ARC085	17	0	18928388	Heterozygous loss	18,928,388

ARC085	18	41682523	51010358	Heterozygous loss	9,327,835
ARC085	20	17443135	18270250	Heterozygous loss	827,115
ARC087	13	50272984	51662741	Heterozygous loss	1,389,757
ARC087	17	0	19653846	cnLOH	19,653,846
ARC091	12	49079922	133851895	cnLOH	84,771,973
ARC093	13	49913914	51838923	Heterozygous loss	1,925,009
ARC093	X	94495681	95620533	Heterozygous gain	1,124,852
ARC095	5	20059415	20220818	Heterozygous loss	161,403
ARC095	13	49370620	49785789	Homozygous loss	415,169
ARC095	13	43905516	50335740	Heterozygous loss	6,430,224
ARC095	13	96325432	96998181	Heterozygous gain	672,749
ARC095	13	100279230	100573728	Heterozygous gain	294,498
ARC098	12	0	133851895	Heterozygous gain	133,851,895
ARC104	13	47573515	53030078	Heterozygous loss	5,456,563
ARC108	13	50703298	51386273	Heterozygous loss	682,975
ARC109	18	62871683	62987358	Heterozygous gain	115,675
ARC120	1	160085628	161353756	Heterozygous loss	1,268,128
ARC120	11	94962869	123071852	Heterozygous loss	28,108,983
ARC120	13	48520141	52450509	Heterozygous loss	3,930,368
ARC121	6	147891657	151673501	Heterozygous loss	3,781,844
ARC121	6	159058713	159407778	Heterozygous loss	349,065
ARC121	13	49346142	51444614	Heterozygous loss	2,098,472
ARC122	5	45790838	46097556	Heterozygous loss	306,718
ARC122	5	46202779	46404562	Heterozygous loss	201,783
ARC122	7	65195552	66463104	Heterozygous gain	1,267,552
ARC122	12	163819	26422277	Heterozygous gain	26,258,458
ARC122	13	47561317	76818454	Heterozygous loss	29,257,137
ARC122	15	80509584	93453610	Heterozygous gain	12,944,026
ARC122	15	93469878	102531392	Heterozygous loss	9,061,514
ARC134	13	50530223	51397172	Heterozygous loss	866,949
ARC137	3	98220041	98354900	Heterozygous loss	134,859
ARC137	3	99526266	100953381	Heterozygous loss	1,427,115
ARC137	3	101446336	102009304	Heterozygous loss	562,968
ARC137	3	104939476	106302333	Heterozygous loss	1,362,857
ARC137	13	48714976	49186401	Heterozygous loss	471,425
ARC137	13	46835594	47335865	Heterozygous loss	500,271

ARC137	13	50512578	51487735	Heterozygous loss	975,157
ARC138	2	36905829	39247190	Heterozygous loss	2,341,361
ARC138	2	168933371	170897070	Heterozygous loss	1,963,699
ARC138	11	106352758	119881557	Heterozygous loss	13,528,799
ARC138	13	111881509	112541980	Heterozygous loss	660,471
ARC138	13	47581433	72072949	Heterozygous loss	24,491,516
ARC138	18	39635593	78077022	cnLOH	38,441,429
ARC143	4	0	28796079	Heterozygous loss	28,796,079
ARC143	11	91250476	117890116	Heterozygous loss	26,639,640
ARC143	13	47581433	72072949	Heterozygous loss	24,491,516
ARC144	13	50143556	51350988	Heterozygous loss	1,207,432
ARC144	22	16851394	17299147	Heterozygous gain	447,753
ARC149	1	101785577	108723404	Heterozygous loss	6,937,827
ARC149	6	91184430	128795645	Heterozygous loss	37,611,215
ARC149	9	0	273153	Heterozygous gain	273,153
ARC149	11	82019934	89652696	Heterozygous loss	7,632,762
ARC149	13	31780755	70967847	Heterozygous loss	39,187,092
ARC150	12	0	133851895	Heterozygous gain	133,851,895
ARC150	12	0	133851895	Heterozygous gain	133,851,895
ARC154	4	97984423	98156737	Heterozygous loss	172,314
ARC154	12	163819	34853011	Heterozygous gain	34,689,192
ARC154	18	0	78077248	Heterozygous gain	78,077,248
ARC154	19	0	59128983	Heterozygous gain	59,128,983
ARC154	21	13122630	48129895	Heterozygous gain	35,007,265
ARC154	22	15291454	51304566	Heterozygous gain	36,013,112
ARC157	9	96367593	141213431	cnLOH	44,845,838
ARC157	13	41686983	51366944	Heterozygous loss	9,679,961
ARC159	13	46655930	67747694	Heterozygous loss	21,091,764
ARC160	3	189790998	190142548	Heterozygous loss	351,550
ARC160	22	20656811	21464479	Heterozygous gain	807,668
ARC166	6	105404380	109686228	Heterozygous loss	4,281,848
ARC166	13	50427605	51632119	Heterozygous loss	1,204,514
ARC166	13	41386246	44543762	Heterozygous loss	3,157,516
ARC171	13	48295262	52870512	Heterozygous loss	4,575,250
ARC171	13	52870512	73110135	Heterozygous gain	20,239,623
ARC171	X	88141996	155226929	cnLOH	67,084,933

ARC172	7	63520619	63790055	Heterozygous loss	269,436
ARC172	12	0	133851895	Heterozygous gain	133,851,895
ARC172	13	50421188	51803383	Heterozygous loss	1,382,195
ARC172	18	0	78077248	Heterozygous gain	78,077,248
ARC172	19	0	59128983	Heterozygous gain	59,128,983
ARC175	13	51448093	51581651	Heterozygous loss	133,558
ARC175	13	50619912	51448093	Homozygous loss	828,181
ARC176	4	74824186	74976170	Heterozygous gain	151,984
ARC178	13	19167959	115169878	cnLOH	96,001,919
ARC178	13	50375715	51572056	Homozygous loss	1,196,341
ARC179	3	141897839	142088433	Heterozygous gain	190,594
ARC179	13	20425929	115169878	cnLOH	94,743,949
ARC179	13	50637901	51417890	Homozygous loss	779,989
ARC180	3	0	362547	Heterozygous loss	362,547
ARC180	3	163299786	197891568	Heterozygous gain	34,591,782
ARC180	6	62613438	131895796	Heterozygous loss	69,282,358
ARC180	13	50734843	51123022	Heterozygous loss	388,179
ARC185	4	153327648	154332776	Heterozygous loss	1,005,128
ARC185	4	154392205	154502035	Heterozygous loss	109,830
ARC185	11	62527056	135006516	cnLOH	72,479,460
ARC185	13	50006991	51616960	Heterozygous loss	1,609,969
ARC185	15	20005287	44105483	Heterozygous loss	24,100,196
ARC185	18	0	14797890	Heterozygous loss	14,797,890
ARC186	13	47268255	52704392	Heterozygous loss	5,436,137
ARC186	13	50619259	51413068	Homozygous loss	793,809
ARC203	13	22489706	115169878	cnLOH	92,680,172
ARC203	13	51383220	51490801	Heterozygous loss	107,581
ARC203	13	50597377	51383220	Homozygous loss	785,843
ARC203	19	43547240	43865893	Heterozygous loss	318,653
ARC205	8	128070623	128551922	Heterozygous gain	481,299
ARC205	13	50554079	51524894	Homozygous loss	970,815
ARC205	13	38361428	55022761	Heterozygous loss	16,661,333
ARC205	17	73538530	75136209	Heterozygous loss	1,597,679
ARC206	12	0	133851895	Heterozygous gain	133,851,895
ARC206	18	0	78077248	Heterozygous gain	78,077,248
ARC206	19	0	59128983	Heterozygous gain	59,128,983

ARC221	2	233896082	235372205	Heterozygous loss	1,476,123
ARC221	10	91353407	94398013	Heterozygous loss	3,044,606
ARC221	12	0	133851895	Heterozygous gain	133,851,895
ARC221	X	41208588	44806300	Heterozygous loss	3,597,712
ARC227	1	225811708	226470900	Heterozygous loss	659,192
ARC227	5	84734558	104509804	Heterozygous loss	19,775,246
ARC227	13	48785422	52710942	Heterozygous loss	3,925,520
ARC227	16	76935165	77659796	Heterozygous gain	724,631
ARC230	12	0	133851895	Heterozygous gain	133,851,895
ARC234	12	0	133851895	Heterozygous gain	133,851,895
ARC234	13	46953385	61003623	Heterozygous loss	14,050,238
ARC240	13	48816827	50630624	Heterozygous loss	1,813,797
ARC240	18	65346514	78077248	cnLOH	12,730,734
ARC241	2	0	10137167	Heterozygous loss	10,137,167
ARC241	2	16358744	18037228	Heterozygous loss	1,678,484
ARC241	2	23580077	35808769	Heterozygous gain	12,228,692
ARC241	2	39475025	59517301	Heterozygous loss	20,042,276
ARC241	2	59517301	63460276	Heterozygous gain	3,942,975
ARC241	2	131656898	154274260	Heterozygous gain	22,617,362
ARC241	2	137158157	137524307	Heterozygous gain	366,150
ARC241	2	162381785	164651330	Heterozygous loss	2,269,545
ARC241	3	88327519	99382886	Heterozygous loss	11,055,367
ARC241	3	118073280	124128659	Heterozygous loss	6,055,379
ARC241	3	131713232	169089769	Heterozygous loss	37,376,537
ARC241	7	82968390	83775147	Heterozygous loss	806,757
ARC241	7	117512899	127345964	Heterozygous loss	9,833,065
ARC241	7	128845584	131944787	Heterozygous gain	3,099,203
ARC241	7	132894644	134482399	Heterozygous gain	1,587,755
ARC241	7	134521238	139161154	Heterozygous loss	4,639,916
ARC241	7	146780930	152741128	Heterozygous gain	5,960,198
ARC241	8	0	7409660	Heterozygous gain	7,409,660
ARC241	8	8104078	10167959	Heterozygous loss	2,063,881
ARC241	8	10167960	11891158	Heterozygous gain	1,723,198
ARC241	8	12504200	14220737	Heterozygous loss	1,716,537
ARC241	8	14220738	18763432	Heterozygous gain	4,542,694
ARC241	8	18763433	23220718	Heterozygous loss	4,457,285

ARC241	8	27576255	33378107	Heterozygous loss	5,801,852
ARC241	8	38479974	48602867	Heterozygous loss	10,122,893
ARC241	8	58820953	62385397	Heterozygous loss	3,564,444
ARC241	8	62385397	134952902	Heterozygous gain	72,567,505
ARC241	8	134952903	139929177	Heterozygous loss	4,976,274
ARC241	8	139929178	145290260	Heterozygous gain	5,361,082
ARC241	9	2751092	4606746	Heterozygous loss	1,855,654
ARC241	9	7948620	35630642	Heterozygous loss	27,682,022
ARC241	9	37031780	39179289	Heterozygous loss	2,147,509
ARC241	9	70961350	81426579	Heterozygous loss	10,465,229
ARC241	9	82841335	141095050	Heterozygous loss	58,253,715
ARC241	13	50284871	52098239	Heterozygous loss	1,813,368
ARC241	13	36288210	39151214	Heterozygous loss	2,863,004
ARC241	13	72065582	115169878	Heterozygous loss	43,104,296
ARC241	14	20496667	21683027	Heterozygous loss	1,186,360
ARC241	14	25160045	29204885	Heterozygous loss	4,044,840
ARC241	15	23624148	98107954	Heterozygous gain	74,483,806
ARC241	15	98107954	102531392	Heterozygous loss	4,423,438
ARC241	17	0	21563440	Heterozygous loss	21,563,440
ARC241	17	21685023	81057996	Heterozygous gain	59,372,973
ARC244	9	70961350	141213431	Heterozygous gain	70,252,081
ARC244	12	0	133851895	Heterozygous gain	133,851,895
ARC245	3	162426346	197891568	Heterozygous gain	35,465,222
ARC245	6	140934715	171115067	Heterozygous loss	30,180,352
ARC245	7	138797783	139319057	Heterozygous loss	521,274
ARC245	8	128335896	128733027	Heterozygous loss	397,131
ARC251	1	36589470	38064983	Heterozygous loss	1,475,513
ARC251	2	164431799	164904028	Heterozygous loss	472,229
ARC251	13	50533263	51450417	Homozygous loss	917,154
ARC251	13	49578332	51639536	Heterozygous loss	2,061,204
ARC252	13	48827964	51875075	Heterozygous loss	3,047,111
ARC254	4	0	36614382	Heterozygous loss	36,614,382
ARC254	5	160578066	180700260	Heterozygous gain	20,122,194
ARC254	7	92873348	93863307	Heterozygous gain	989,959
ARC254	11	60873551	135006516	cnLOH	74,132,965
ARC254	13	49816523	51709242	Heterozygous loss	1,892,719

ARC264	12	0	133851895	Heterozygous gain	133,851,895
ARC264	14	69259582	107349540	Heterozygous loss	38,089,958
ARC280	1	153005002	153456568	Heterozygous gain	451,566
ARC280	2	0	92304211	Heterozygous gain	92,304,211
ARC280	6	105656696	171115067	Heterozygous loss	65,458,371
ARC280	8	130798606	146364022	Heterozygous loss	15,565,416
ARC280	9	9261503	38774065	Heterozygous loss	29,512,562
ARC280	9	96745707	141213431	Heterozygous loss	44,467,724
ARC280	10	104630575	106146873	Heterozygous loss	1,516,298
ARC280	10	124537565	126575828	Heterozygous loss	2,038,263
ARC280	13	49206215	51572757	Heterozygous loss	2,366,542
ARC280	14	42393055	44736968	Heterozygous loss	2,343,913
ARC280	17	0	21563440	Heterozygous loss	21,563,440
ARC280	18	0	50706606	Heterozygous loss	50,706,606
ARC280	19	20827749	20986033	Heterozygous gain	158,284
ARC280	20	0	26316921	Heterozygous loss	26,316,921
ARC282	2	34694489	34835157	Heterozygous loss	140,668
ARC286	13	45047752	52078337	Heterozygous loss	7,030,585
ARC292	6	130906219	145822249	Heterozygous loss	14,916,030
ARC292	10	10678598	15909158	Heterozygous loss	5,230,560
ARC292	11	57585753	62324504	Heterozygous loss	4,738,751
ARC292	11	78395212	92290618	Heterozygous loss	13,895,406
ARC292	11	95490539	99376638	Heterozygous loss	3,886,099
ARC292	11	105060755	115657397	Heterozygous loss	10,596,642
ARC292	13	48884607	51361934	Homozygous loss	2,477,327
ARC292	13	41312511	52002256	Heterozygous loss	10,689,745
ARC297	12	0	133851895	Heterozygous gain	133,851,895
ARC299	1	93870381	94289849	Heterozygous loss	419,468
ARC301	2	82753005	83187612	Heterozygous gain	434,607
ARC301	3	60179160	60468566	Heterozygous loss	289,406
ARC301	13	28789830	115169878	cnLOH	86,380,048
ARC301	13	48741749	52611967	Heterozygous loss	3,870,218
ARC301	X	22849432	23118512	Heterozygous loss	269,080
ARC304	2	0	15678585	Heterozygous gain	15,678,585
ARC304	2	42992463	79943663	Heterozygous gain	36,951,200
ARC304	3	19626218	31062993	Heterozygous loss	11,436,775

ARC304	3	34421866	37378795	Heterozygous loss	2,956,929
ARC304	3	43151605	43612450	Heterozygous loss	460,845
ARC304	3	45637039	49127046	Heterozygous loss	3,490,007
ARC304	3	59964597	62282627	Heterozygous loss	2,318,030
ARC304	3	72842006	75352276	Heterozygous loss	2,510,270
ARC304	4	0	39570573	Heterozygous loss	39,570,573
ARC304	4	39570573	43033272	Heterozygous gain	3,462,699
ARC304	4	43033272	43182715	Heterozygous loss	149,443
ARC304	4	44136700	49325091	Heterozygous loss	5,188,391
ARC304	4	124761656	125144027	Heterozygous loss	382,371
ARC304	4	125144027	152776818	Heterozygous gain	27,632,791
ARC304	4	153037799	191154276	Heterozygous gain	38,116,477
ARC304	8	58697353	71697573	Heterozygous loss	13,000,220
ARC304	11	84200504	84471311	Heterozygous loss	270,807
ARC304	13	113848191	114325855	Heterozygous loss	477,664
ARC304	13	63847239	64851011	Heterozygous loss	1,003,772
ARC304	13	54791893	59255570	Heterozygous loss	4,463,677
ARC304	13	59255570	63847239	Heterozygous gain	4,591,669
ARC304	13	35949965	40521211	Heterozygous loss	4,571,246
ARC304	13	64851011	89189709	Heterozygous gain	24,338,698
ARC304	13	19167959	28463805	Heterozygous loss	9,295,846
ARC304	15	93376188	102531392	Heterozygous loss	9,155,204
ARC304	17	0	22252439	Heterozygous loss	22,252,439
ARC305	11	81489861	96287529	Heterozygous loss	14,797,668
ARC305	11	96439090	117547807	Heterozygous loss	21,108,717
ARC305	13	18159566	115169878	cnLOH	97,010,312
ARC305	13	49824565	51596325	Heterozygous loss	1,771,760
ARC305	13	50738084	51064512	Homozygous loss	326,428
ARC305	15	27470154	27817005	Heterozygous gain	346,851
ARC310	4	52682617	65222229	cnLOH	12,539,612
ARC310	13	19167959	115169878	cnLOH	96,001,919
ARC313	13	49733827	52790700	Heterozygous loss	3,056,873
ARC314	17	14101133	15485379	Heterozygous gain	1,384,246
ARC331	2	0	79768208	Heterozygous gain	79,768,208
ARC331	4	0	27173815	Heterozygous loss	27,173,815
ARC331	11	99423037	135006516	Heterozygous loss	35,583,479

ARC331	13	48507525	52790700	Heterozygous loss	4,283,175
ARC331	18	0	15102421	Heterozygous loss	15,102,421
ARC334	13	50451677	51597808	Heterozygous loss	1,146,131
ARC344	8	10716240	10959041	Heterozygous gain	242,801
ARC344	9	0	35731006	Heterozygous gain	35,731,006
ARC344	13	21729085	115169878	cnLOH	93,440,793
ARC344	13	50587397	51057650	Homozygous loss	470,253
ARC344	17	0	21563440	Heterozygous loss	21,563,440
ARC344	20	9753409	26316921	Heterozygous loss	16,563,512
ARC344	20	31724133	36184748	Heterozygous loss	4,460,615
ARC344	20	37447598	40421766	Heterozygous loss	2,974,168
ARC344	X	6449368	8148906	Heterozygous gain	1,699,538
ARC379	11	76864788	116695042	Heterozygous loss	39,830,254
ARC393	3	17856262	31098473	Heterozygous loss	13,242,211
ARC393	3	31719053	33837957	Heterozygous loss	2,118,904
ARC393	6	85087091	87503885	Heterozygous loss	2,416,794
ARC393	6	91583614	92548764	Heterozygous loss	965,150
ARC393	6	93574588	93920649	Heterozygous loss	346,061
ARC393	6	94522310	96140970	Heterozygous loss	1,618,660
ARC393	6	97887760	98179041	Heterozygous loss	291,281
ARC393	6	98650395	127854237	Heterozygous loss	29,203,842
ARC393	6	128426135	129897407	Heterozygous loss	1,471,272
ARC393	8	77266438	77839392	Heterozygous loss	572,954
ARC393	8	77985854	78383248	Heterozygous loss	397,394
ARC393	11	22488644	24455306	Heterozygous loss	1,966,662
ARC393	11	27292107	27473516	Heterozygous loss	181,409
ARC393	11	55839144	61107749	Heterozygous loss	5,268,605
ARC393	11	89822105	89998197	Heterozygous loss	176,092
ARC393	11	91206606	122448278	Heterozygous loss	31,241,672
ARC393	11	123489354	126051928	Heterozygous loss	2,562,574
ARC393	11	126809291	127080863	Heterozygous loss	271,572
ARC393	11	131290369	133324297	Heterozygous loss	2,033,928
ARC400	6	10009510	17176845	Heterozygous loss	7,167,335
ARC400	6	105465122	112511335	Heterozygous loss	7,046,213
ARC400	6	148870649	152776864	Heterozygous loss	3,906,215
ARC400	6	155203191	157466966	Heterozygous loss	2,263,775

ARC400	12	0	133851895	Heterozygous gain	133,851,895
ARC400	13	39087053	42796356	Heterozygous loss	3,709,303
ARC400	13	46880911	52790700	Heterozygous loss	5,909,789
ARC417	16	77570764	77989996	Heterozygous gain	419,232
ARC417	17	0	20857549	Heterozygous loss	20,857,549
ARC417	17	25268096	25969529	Heterozygous loss	701,433
ARC417	17	27044452	27463250	Heterozygous loss	418,798
ARC417	17	27509201	27680369	Heterozygous loss	171,168
ARC417	18	0	15400035	Heterozygous loss	15,400,035
ARC417	18	19397454	29087503	Heterozygous loss	9,690,049
ARC484	8	95518824	146364022	Heterozygous loss	50,845,198
ARC484	11	90115229	92462025	Heterozygous loss	2,346,796
ARC484	11	92462023	104514598	Heterozygous gain	12,052,575
ARC484	11	104514598	122978274	Heterozygous loss	18,463,676
ARC484	11	122978274	125073371	Heterozygous gain	2,095,097
ARC484	11	125112751	128180048	Heterozygous loss	3,067,297
ARC484	11	128180048	132992412	Heterozygous gain	4,812,364
ARC484	11	132992412	134991927	Heterozygous loss	1,999,515
ARC484	13	19055298	24558883	Heterozygous gain	5,503,585
ARC484	13	24558883	28622711	Heterozygous loss	4,063,828
ARC484	13	28622711	35996914	Heterozygous gain	7,374,203
ARC484	13	35996914	44459769	Heterozygous loss	8,462,855
ARC484	13	44459769	48119664	Heterozygous gain	3,659,895
ARC484	13	48119664	63236583	Heterozygous loss	15,116,919
ARC484	13	63236583	88229617	Heterozygous gain	24,993,034
ARC484	13	88229617	92790136	Heterozygous gain	4,560,519
ARC484	13	92790136	115169878	Heterozygous loss	22,379,742
ARC484	14	69262437	106051813	Heterozygous loss	36,789,376
ARC484	17	0	16062092	cnLOH	16,062,092
ARC490	6	12576234	127466517	cnLOH	114,890,283
ARC490	11	107650358	107811628	Heterozygous gain	161,270
ARC495	13	50584367	50977946	Heterozygous loss	393,579
ARC502	10	135235268	135374676	Heterozygous gain	139,408
ARC502	13	48842845	51891532	Heterozygous loss	3,048,687
ARC528	13	40351981	41409219	Heterozygous loss	1,057,238
ARC528	13	50531923	51644288	Homozygous loss	1,112,366

ARC528	13	50282420	51964904	Heterozygous loss	1,682,484
ARC533	13	55146605	55481157	Heterozygous loss	334,552
ARC554	2	86295905	86464391	Heterozygous gain	168,486
ARC554	11	76810142	116449573	Heterozygous loss	39,639,431
ARC554	13	49824565	52733533	Heterozygous loss	2,908,968
ARC554	20	18149173	18500661	Heterozygous gain	351,488
ARC566	3	0	12198021	Heterozygous loss	12,198,021
ARC566	3	162626585	197891568	Heterozygous gain	35,264,983
ARC566	12	11265767	16729175	Heterozygous loss	5,463,408
ARC566	15	40672051	43053754	Heterozygous loss	2,381,703
ARC574	11	108080945	108184654	Heterozygous loss	103,709
ARC574	13	49974128	52344014	Heterozygous loss	2,369,886
ARC574	13	36244190	40577345	Heterozygous loss	4,333,155
ARC574	13	19167959	26953054	Heterozygous loss	7,785,095
ARC574	18	0	15400035	Heterozygous loss	15,400,035
ARC574	21	14338959	48129895	Heterozygous gain	33,790,936
ARC576	11	96439090	115878087	Heterozygous loss	19,438,997
ARC576	13	49211748	52883671	Heterozygous loss	3,671,923
ARC579	6	45257874	50256199	Heterozygous loss	4,998,325
ARC579	6	50967579	53893444	Heterozygous loss	2,925,865
ARC579	6	54739811	55011461	Heterozygous loss	271,650
ARC579	6	85861615	86506338	Heterozygous loss	644,723
ARC579	6	87868430	90341170	Heterozygous loss	2,472,740
ARC579	6	147034813	148584333	Heterozygous loss	1,549,520
ARC579	9	5098265	5231015	Heterozygous gain	132,750
ARC579	14	41842009	41993354	Heterozygous loss	151,345
ARC579	18	1753016	2319401	Heterozygous loss	566,385
ARC579	18	2960021	6417139	Heterozygous loss	3,457,118
ARC579	18	8487774	9604743	Heterozygous loss	1,116,969
ARC579	18	11478238	15400035	Heterozygous loss	3,921,797
ARC579	X	8606069	11102142	Heterozygous loss	2,496,073
ARC579	X	93993809	155270560	Heterozygous gain	61,276,751
ARC582	4	147227993	147538773	Heterozygous gain	310,780
ARC582	9	121543917	121781307	Heterozygous loss	237,390
ARC582	12	0	133851895	Heterozygous gain	133,851,895
ARC592	13	19972736	115169878	cnLOH	95,197,142

ARC592	13	49874254	51110749	Heterozygous loss	1,236,495
ARC596	7	152592393	152750778	Heterozygous loss	158,385
ARC601	12	0	133851895	Heterozygous gain	133,851,895
ARC601	15	40337869	41146872	Heterozygous loss	809,003
ARC601	15	42050092	42896292	Heterozygous loss	846,200
ARC601	18	64169202	64281111	Heterozygous gain	111,909
ARC602	1	227782279	235253833	Heterozygous loss	7,471,554
ARC602	11	102650868	125549040	Heterozygous loss	22,898,172
ARC602	14	34977758	35900214	Heterozygous loss	922,456
ARC602	17	0	102914	Heterozygous loss	102,914
ARC603	8	86841171	146364022	Heterozygous gain	59,522,851
ARC603	13	50540730	51404486	Heterozygous loss	863,756
ARC603	17	0	18928388	Heterozygous loss	18,928,388
ARC603	17	19141867	20215768	Heterozygous gain	1,073,901
ARC603	20	0	21517630	Heterozygous loss	21,517,630
ARC604	12	86079455	86441373	Heterozygous loss	361,918
ARC609	1	26452854	29211800	Heterozygous loss	2,758,946
ARC609	3	51768658	54446830	Heterozygous loss	2,678,172
ARC609	8	0	27605868	Heterozygous loss	27,605,868
ARC609	10	90943361	95877263	Heterozygous loss	4,933,902
ARC609	10	102594103	106474286	Heterozygous loss	3,880,183
ARC609	12	0	75896153	Heterozygous gain	75,896,153
ARC609	12	9852187	56887951	cnLOH	47,035,764
ARC618	12	0	133851895	Heterozygous gain	133,851,895
ARC618	14	69259202	106049154	Heterozygous loss	36,789,952
ARC621	1	234731228	235247828	Heterozygous loss	516,600
ARC621	3	1742170	1934142	Heterozygous loss	191,972
ARC621	16	15489524	18201888	Heterozygous gain	2,712,364
ARC627	12	0	133851895	Heterozygous gain	133,851,895
ARC627	13	50461625	51667628	Heterozygous loss	1,206,003
ARC627	19	0	59128983	Heterozygous gain	59,128,983
ARC628	12	0	133851895	Heterozygous gain	133,851,895
ARC631	1	219842956	220698566	Heterozygous loss	855,610
ARC631	1	228671404	235417107	Heterozygous loss	6,745,703
ARC631	11	85430388	123687230	Heterozygous loss	38,256,842
ARC631	13	49788290	52712340	Heterozygous loss	2,924,050

ARC634	4	189686691	189922063	Heterozygous gain	235,372
ARC634	4	190380303	190640778	Heterozygous gain	260,475
ARC639	3	142896167	197891568	Heterozygous gain	54,995,401
ARC639	12	0	133851895	Heterozygous gain	133,851,895
ARC639	12	111356844	113234337	Heterozygous gain	1,877,493
ARC643	12	0	133851895	Heterozygous gain	133,851,895
ARC646	1	149952169	249250621	cnLOH	99,298,452
ARC646	11	86721678	115886487	Heterozygous loss	29,164,809
ARC649	11	58093902	64503271	Heterozygous loss	6,409,369
ARC649	11	104692000	116556519	Heterozygous loss	11,864,519
ARC649	X	95928611	155270560	cnLOH	59,341,949
ARC679	4	56171914	57132122	Heterozygous loss	960,208
ARC679	4	79519074	79832607	Heterozygous loss	313,533
ARC679	4	82325347	85730322	Heterozygous loss	3,404,975
ARC679	4	94460048	95581702	Heterozygous loss	1,121,654
ARC679	4	95728497	96181324	Heterozygous loss	452,827
ARC679	4	97758879	98254857	Heterozygous loss	495,978
ARC679	4	104404113	104876898	Heterozygous loss	472,785
ARC679	4	106107627	106289099	Heterozygous loss	181,472
ARC679	4	106892200	107921186	Heterozygous loss	1,028,986
ARC679	4	113844145	114581080	Heterozygous loss	736,935
ARC679	4	114719440	118995072	Heterozygous loss	4,275,632
ARC679	4	119005199	119404287	Heterozygous loss	399,088
ARC679	4	124006110	125967493	Heterozygous loss	1,961,383
ARC679	4	126026319	126830788	Heterozygous loss	804,469
ARC679	4	126901347	129577132	Heterozygous loss	2,675,785
ARC679	4	129581981	129804563	Heterozygous loss	222,582
ARC679	4	130262000	134854112	Heterozygous loss	4,592,112
ARC679	4	176528793	190986343	Heterozygous loss	14,457,550
ARC679	7	75415883	75810760	Heterozygous loss	394,877
ARC679	7	75833104	76955071	Heterozygous loss	1,121,967
ARC679	7	77438129	78770868	Heterozygous loss	1,332,739
ARC679	7	127678875	127783687	Heterozygous loss	104,812
ARC679	7	135586884	145048066	Heterozygous loss	9,461,182
ARC679	7	148994310	149688281	Heterozygous loss	693,971
ARC679	11	194441	509145	Heterozygous loss	314,704

ARC679	11	4286931	4978493	Heterozygous loss	691,562
ARC679	11	15398591	16656985	Heterozygous loss	1,258,394
ARC679	11	18773828	24119828	Heterozygous loss	5,346,000
ARC679	11	25127685	26262104	Heterozygous loss	1,134,419
ARC679	11	26672534	27381921	Heterozygous loss	709,387
ARC679	11	27427151	27747024	Heterozygous loss	319,873
ARC679	11	29284186	33794537	Heterozygous loss	4,510,351
ARC679	11	34460804	44071880	Heterozygous loss	9,611,076
ARC679	11	48286243	51591253	Heterozygous loss	3,305,010
ARC679	11	54707899	56359650	Heterozygous loss	1,651,751
ARC679	11	58773859	60605359	Heterozygous loss	1,831,500
ARC679	11	60681103	60925013	Heterozygous loss	243,910
ARC679	11	61012817	61971332	Heterozygous loss	958,515
ARC679	11	70809884	71678185	Heterozygous loss	868,301
ARC679	11	74338370	78569791	Heterozygous loss	4,231,421
ARC679	11	78860950	93421744	Heterozygous loss	14,560,794
ARC679	11	96439090	99070431	Heterozygous loss	2,631,341
ARC679	11	99396082	101330098	Heterozygous loss	1,934,016
ARC679	11	105496100	105621641	Heterozygous loss	125,541
ARC679	11	105774042	109914969	Heterozygous loss	4,140,927
ARC679	11	109927911	117698622	Heterozygous loss	7,770,711
ARC679	11	119401005	120072554	Heterozygous loss	671,549
ARC679	11	122427611	127920662	Heterozygous loss	5,493,051
ARC679	11	129034606	129693872	Heterozygous loss	659,266
ARC679	11	129712704	131956142	Heterozygous loss	2,243,438
ARC679	13	49614445	51879274	Heterozygous loss	2,264,829
ARC679	22	16851394	17365233	Heterozygous loss	513,839
ARC679	22	26207454	26668349	Heterozygous loss	460,895
ARC679	22	27189219	28195043	Heterozygous loss	1,005,824
ARC679	22	28448277	29439737	Heterozygous loss	991,460
ARC679	22	32435943	32904586	Heterozygous loss	468,643
ARC679	22	32944259	35384298	Heterozygous loss	2,440,039
ARC679	22	36247830	36481902	Heterozygous loss	234,072
ARC679	22	43924068	45321861	Heterozygous loss	1,397,793
ARC679	22	47063126	47419434	Heterozygous loss	356,308
ARC679	22	48020393	50025247	Heterozygous loss	2,004,854

ARC682	12	0	133851895	Heterozygous gain	133,851,895
ARC683	6	0	47225909	cnLOH	47,225,909
ARC683	12	14309323	14479008	Heterozygous gain	169,685
ARC683	15	75916234	102531392	cnLOH	26,615,158
ARC689	8	127275421	128714605	Heterozygous gain	1,439,184
ARC697	1	26561943	38348349	Heterozygous loss	11,786,406
ARC697	12	0	133851895	Heterozygous gain	133,851,895
ARC697	21	0	48129895	Heterozygous gain	48,129,895
ARC697	X	1726690	2036900	Heterozygous gain	310,210
ARC706	12	0	133851895	Heterozygous gain	133,851,895
ARC714	11	93426030	118833506	Heterozygous loss	25,407,476
ARC714	13	48699397	52710141	Heterozygous loss	4,010,744
ARC724	3	176654560	176889637	Heterozygous gain	235,077
ARC724	13	48740246	51537592	Heterozygous loss	2,797,346
ARC745	6	80671037	124830621	Heterozygous loss	44,159,584
ARC745	8	53716520	54376866	Heterozygous gain	660,346
ARC745	11	106703048	115337486	Heterozygous loss	8,634,438

cnLOH=copy neutral loss of heterozygosity