Applications of next-generation technologies in the diagnosis of haematological diseases and cancer

Adam D Burns

A thesis submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

This research was carried out in collaboration with both the University of Oxford and the Oxford University Hospitals NHS Foundation Trust

September 2016

For Rachel, Oliver and Finlay

CONTENTS

ABSTRACT

The advent of massively-parallel next-generation sequencing (NGS) methods has provided researchers with a powerful tool with which to interrogate and characterise the molecular landscape of cancer genomes. Compared to existing methods of DNA sequencing, NGS platforms generate massive amounts of sequence data and, as a consequence, can reveal information not just on single nucleotide variations (SNVs), but also on copy-number aberrations, translocations and large insertions and deletions in a single experiment. Furthermore, targeted NGS provides the capability to focus on a small number of targets simultaneously, with high accuracy and sensitivity. The presence of specific molecular markers acts as predictors of disease outcome, survival rates and treatment response in individual patients. Screening for such markers has become routine practice in diagnostic laboratories using traditional methods of DNA analysis, are widely used in diagnostic laboratories around the world. Whilst these methods are proven and reliable, their limitations lie in the fact that they focus on only the most prevalent mutations in a particular cancer. The ability to investigate multiple gene targets within individual patients, to a high level of accuracy, and to monitor these changes over time will be a valuable tool in cancer diagnostics. As such, there is a potential use case for NGS techniques in routine diagnostics. Therefore, this thesis investigated the extent to which NGS platforms could be used in

a clinical setting for the diagnosis and risk-stratification of both lymphoid and myeloid malignancies.

A targeted next-generation sequencing panel was designed and validated against existing diagnostic methods. All mutations in the validation cohort were correctly identified. Both the specificity and sensitivity of the assay were determined and were considerably better than those of the current 'gold-standard' techniques. This panel has been fully validated and implemented into the diagnostic service at the John Radcliffe Hospital. The research applications of this panel were also demonstrated through the sequencing of a cohort of del(5q) MDS patients. It was not only found that mutations in *TP53* and *ASXL1* may be key drivers in the progression of del(5q) MDS into AML but also that 40% of del(5q) patients harboured at least one mutation. A number of mutations were below the limit of detection for Sanger sequencing, and so this study expands our knowledge of the del(5q) mutational landscape.

Whole genome sequencing of 42 CLL cases revealed a high level of molecular heterogeneity, with mutations in key CLL driver genes including *TP53*, *SF3B1*, *NOTCH1* and *ATM*. Both clinically relevant CNAs and translocations were detected in the cohort. Four mutation signatures were detected across the CLL genomes and are both associated with, and vary in their prevalence according to, specific clinical characteristics, including age and chemo-refractoriness. Mutations introduced as part of the SHM process in B-cells are present throughout the genome, including in patients

with unmutated IgHV genes. Regions of localised hypermutation are present in CLL, with a number affecting genes associated with coding mutations in CLL, including *ATM*, *KLHL6* and *MEGF9*. A number of mutation clusters are also identified in potentially regulatory regions of genes.

In summary, this thesis demonstrates that both whole genome sequencing and targeted sequencing panels can be introduced into diagnostics to aid the clinical decision-making process and also reveal important new findings that increase our understanding of the pathogenesis of leukaemia.

x

ACKNOWLEDGEMENTS

Firstly, I would like to extend my sincere thanks to my supervisors. I am extremely grateful to Professor Anna Schuh for allowing me the opportunity to perform this research as part of her laboratory. I consider it a great honour and privilege to have been able to work with her. Thanks to Dr Susan Brooks for her support, advice and editing skills during the writing process. Thanks also to Professor Jacqueline Boultwood for her continual support. It has been a privilege to collaborate with her research group during the course of this work.

Thanks also to my colleagues in the University of Oxford and the Molecular Haematology Department, in particular Adele Timbs, Dr Ruth Clifford and Pauline Robbe, whose support and reassurance has been invaluable.

I must also say a big thank you to all my family and friends for their patience, support and understanding.

Finally, and most importantly, I want to thank Rachel, Oliver and Finlay who have always encouraged and believed in me. Thank you for your endless patience and for putting up with me not always being around. I couldn't have done it without you. I love you.

INDIVIDUAL CONTRIBUTIONS

This thesis is the result of my work performed whilst employed as a research assistant within the Nuffield Division of Clinical Laboratory Sciences at the University of Oxford, with the additional contribution of others detailed below.

In chapter three of this thesis, the DNA from the clinical patient cohort, along with analysis of the sequencing data, was kindly provided by Dr Marta Fernandez-Mercado of the Bloodwise Molecular Haematology Unit.

In chapter 4, germline and tumour DNA for whole genome sequencing was extracted by Maite Cabes and her team at the Oxford Radcliffe Biobank. HiSeq 2000 sequencing of the samples, including data alignment and variant detection, was provided by Illumina, Chesterford Research Park, Little Chesterford, UK. Detection of translocations in the WGS data was performed in collaboration with Dr Alexander Kanapin of the Department of Oncology, University of Oxford. Identification of copy-number events was kindly provided by Dr Sam Knight of the Wellcome Trust Centre for Human Genetics. MiSeq sequencing to validate the presence of coding mutations, described in chapter 4, was conducted in collaboration with Reem Alsolami.

LIST OF ABBREVIATIONS

°C	degrees Celsius
ADD	ATRX-DNMT3-DNMT3L domain
AID	activation-induced cytidine deaminase
Ala	alanine
AGS	Aicardi-Goutières syndrome
ALL	acute lymphoblastic leukaemia
AML	acute myeloid leukaemia
ANK	ankyrin repeat domain
ANKRD12	ankyrin repeat domain 12
APS	adenosine 5' phosphosulfate
Arg	arginine
ARMS	amplification refractory mutation system
Asn	asparagine
Asp	aspartic acid
ASP	allele specific primer
ASXL1	additional sex combs like 1
ATM	ataxia telangiectasia mutated
ATMDS	α thalassemia-myelodysplastic syndrome
ATP	adenosine triphosphate
ATRA	all trans-retinoic acid
β2M	beta-2 microglobulin
BAM	binary alignment map file
BCL2	b-cell lymphoma 2
BCL6	b-cell lymphoma 6
BCOR	BCL6 corepressor
BCR	b-cell receptor

BIRC3	baculoviral IAP repeat containing 3
bp	base pair
BL	Burkitt's lymphoma
BTG2	BTG family member 2
BTK	Bruton's tyrosine kinase
BWA	Burrows-Wheeler alignment
CADPS2	Ca2+ dependent secretion activator 2
CBF-AML	core binding factor AML
CBFB	core binding factor beta subunit
CBLB	cbl proto-oncogene B
CBLC	cbl proto-oncogene C
CCD	charge-coupled device
CD5	cluster of differentiation 5
CD19	cluster of differentiation 19
CD20	cluster of differentiation 20
CD38	cluster of differentiation 38
CD49d	cluster of differentiation 49d
CDH12	cadherin 12
CDH18	cadherin 18
CDH19	cadherin 19
CDKN2A	cyclin-dependent kinase inhibitor 2a
CEBPα	CCAT enhancer binding protein alpha
CHD2	chromodomain helicase DNA binding protein 2
Chr	chromosome
CLL	chronic lymphocytic leukaemia
CLP	common lymphoid progenitor
CMML	chronic myelomonocytic leukaemia
CN	copy-number

CNA	copy number aberration
cn-AML	cytogenetically normal acute myeloid leukaemia
COSMIC	catalogue of somatic mutations in cancer
CR	complete remission
CTNNA2	catenin alpha 2
Cy5	cyanine-5
Cys	cysteine
ddATP	dideoxyadenosine-5'-triphosphate
ddCTP	dideoxycytidine-5'-triphosphate
ddGTP	dideoxyguanosine-5'-triphosphate
ddNTP	dideoxynucleotide
ddTTP	dideoxythymidine-5'-triphosphate
DDX3X	DEAD-box helicase 3 X-linked
dL	deciliter
DLBCL	diffuse large B-cell lymphoma
DNA	deoxyribonucleic acid
DNMT3a	DNA (cytosine-5)-methyltransferase 3a
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
EFS	event free survival
EGFR	epidermal growth-factor receptor
EGR2	early growth response 2
ENCODE	encyclopaedia of DNA elements
ETV6	ETS variant 6
EVI1	ectopic virus integration site 1
EZH2	enhancer of zeste homolog 2
FAB	French-American-British
FBXW7	f-box and WD repeat domain containing 7

FCR	fludarabine-cyclophosphamide-rituximab
FHIT	fragile histidine triad
FISH	fluorescent *in-situ* hybridisation
FL	follicular lymphoma
FLT3	fms related tyrosine kinase 3
FP	forward primer
FR	fludarabine refractory
FSTL5	follistatin like 5
g	gram
GATK	genome analysis toolkit
Gb	gigabase
gDNA	genomic DNA
GL	germline
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GPC5	glypican 5
GRIA2	glutamate ionotropic receptor AMPA type subunit 2
GRID2	glutamate ionotropic receptor delta type subunit 2
GTC	guanidine-thiocyanate-containing lysate
H3K4me1	Histone 3 Lysine 4 methylation
H3K4me3	Histone 3 Lysine 4 trimethylation
Hb	haemoglobin
HBD	histone binding domain
HCL	hydrochloric acid
HD	heterodimerisation domain
His	histidine
HIST1H1E	histone cluster 1

HOX	homeobox
HSC	haematopoietic stem cell
IDH1	isocitrate dehydrogenase 1
IDH2	isocitrate dehydrogenase 2
IgHV	immunoglobulin heavy chain variable
IGLL5	immunoglobulin lambda like polypeptide 5
IGV	integrative genomics viewer
Ile	isoleucine
IRF4	interferon regulatory factor 4
ITD	internal tandem duplication
ITPKB	inositol-triphosphate 3-kinase B
JAK2	janus kinase 2
JM	juxta-membrane domain
JMML	juvenile myelomonocytic leukaemia
$KHCO_3$	potassium bicarbonate
KLHL1	kelch like family member 1
KLHL6	kelch like family member 6
KMT2C	lysine methyltransferase 2c
Leu	leucine
LNR	Lin-Notch rich domain
LOH	loss of heterozygosity
LRP1B	low density lipoprotein receptor related protein 1b
Lys	lysine
M	molar
MALT	mucosa-associated lymphoid tissue
MAML	mastermind-like transcriptional coactivator 1
MAP3K14	mitogen-activated protein kinase kinase kinase 14
MAPK	mitogen-activated protein kinase

Mb	megabase
MCL	mantle cell lymphoma
MDR	minimally deleted region
MDS	myelodysplastic syndrome
MED12	mediator complex subunit 12
MEGF9	multiple EGF like domains 9
Met	methionine
mg	milligram
mg/L	milligrams per litre
MHC	major histocompatibility complex
mir-15-a	microRNA 15a
mir-16-a	microRNA 16a
MKL1	megakaryoblastic leukaemia 1
ml	millilitre
mm	millimetre
MM	multiple myeloma
MPD	myeloproliferative disease
MPP	multipotent progenitor cell
MRD	minimal residual disease
MYD88	myeloid differentiation primary response 88
MYH11	myosin heavy chain 11
NaCl	sodium chloride
NEG	negative regulation domain
NES	nuclear export sequence
ng	nanogram
NGS	next-generation sequencing
NH_4Cl	ammonium chloride
nm	nanometre

NMD	nonsense mediated decay
NPM1	nucleophosmin 1
OD	oligodimerisation domain
OS	overall survival
PAX5	paired box 5
PCLO	piccolo presynaptic cytomatrix protein
PCR	polymerase chain reaction
PDGFRA	platelet derived growth factor receptor alpha
PFS	progression free survival
PGM	personal genome machine
Phe	phenylalanine
PI3K	phosphoinositide 3-kinase
PML	promyelocytic leukaemia protein
POT1	protection of telomeres 1
PPI	pyrophosphate
PRD	proline rich domain
Pro	proline
PTEN	phosphatase and tensin homolog
PTP	picotiter plate
RA	refractory anemia
RAM	RBPJ-associated molecule domain
RARA	retinoic acid receptor alpha
RBC	red blood cell
RBM15	RNA binding motif protein 15
RD	read depth
RFS	relapse-free survival
RIN	RNA integrity number
RING	really interesting new gene

RIPK1	receptor-interacting serine-threonine kinase 1
RNA	ribonucleic acid
RNA-Seq	ribonucleic acid sequencing
RP	reverse primer
RPN1	ribophorin 1
RS	Richter's syndrome
RT-PCR	real-time polymerase chain reaction
RUNX1	runt-related transcription factor 1
RUNX1T1	RUNX1 translocation partner 1
SAMHD1	SAM domain and HD domain-containing protein 1
SBS	sequencing-by-synthesis
SDS	sodium dodecyl sulphate
SEM	standard error of the mean
Ser	serine
SF3B1	splicing factor 3b subunit 1
SHM	somatic hypermutation
SIFT	sorting intolerant from tolerant
SMS	single molecule sequencing
SMZL	splenic marginal zone lymphoma
SNP	single-nucleotide polymorphism
SRSF2	serine/arginine-rich splicing factor 2
STAT5	signal transducer and activator of transcription 5
T	tumour
TAD	trans-activation domain
TBE	tris-borate-EDTA
TET2	tet methylcytosine dioxygenase 2
TFS	treatment free survival
TGIF1	TGFB induced factor homeobox 1

Thr	threonine
TKD	tyrosine kinase domain
Tm	melting temperature
TP53	tumour protein 53
Trp	tryptophan
TSCA	TruSeq custom amplification
TTFT	time-to-first-treatment
Tyr	tyrosine
U2AF1	U2 small nuclear RNA auxilairy factor 1
UK	United Kingdom
µl	microliter
µM	microMolar
UPD	uniparental disomy
USA	United States of America
UTR	untranslated region
UV	ultra violet
V	volt
VAF	variant allele frequency
Val	valine
VCF	variant call file
w/v	weight to volume
WES	whole-exome sequencing
WGS	whole-genome sequencing
WHO	World Health Organisation
WT	wild-type
WT1	wilms tumour protein
XPO1	exportin 1
ZAP70	zeta chain of T cell receptor associated protein kinase 70kDa

ZFPM2	zinc finger protein FOG family member 2
ZMW	zero-mode wave guide
ZMYM3	zinc finger MYM-type containing 3

LIST OF FIGURES

LIST OF TABLES

Chapter 1 INTRODUCTION

1.1 Chronic Lymphocytic Leukaemia

Chronic lymphocytic leukaemia (CLL) is the most common form of adult leukaemia in the western world (Swerdlow *et al* 2008), accounting for one third of new leukaemia cases each year (Siegel *et al* 2012). CLL is a haematological malignancy originating in the bone marrow and affects the precursors of a specific type of white blood cell, the lymphocyte (Figure 1.1). CLL cells replicate at a higher rate and are resistant to apoptosis, causing them to out-compete the normal cells. CLL is characterised by both clinical and biological heterogeneity. The majority of CLL patients display few symptoms at initial diagnosis, after which the disease can progress into either an aggressive, chemo-resistant form with poor prognosis, or a relatively indolent form, with a life expectancy similar to that seen in the normal population (Gaidano *et al* 2012, Gruber and Wu 2014, Pekarsky *et al* 2010, Quesada *et al* 2013, Zenz, *et al* 2010a). CLL cells are defined by the presence of both CD5 and CD19 cell surface markers, with a diagnosis of CLL made when a patient presents with at least $5x10^9$/L lymphocytes (Hallek *et al* 2008). The median age of patients at diagnosis is 70 years, with more males affected than females by a ratio of 1.7:1 (Molica 2006, Watson *et al* 2008).

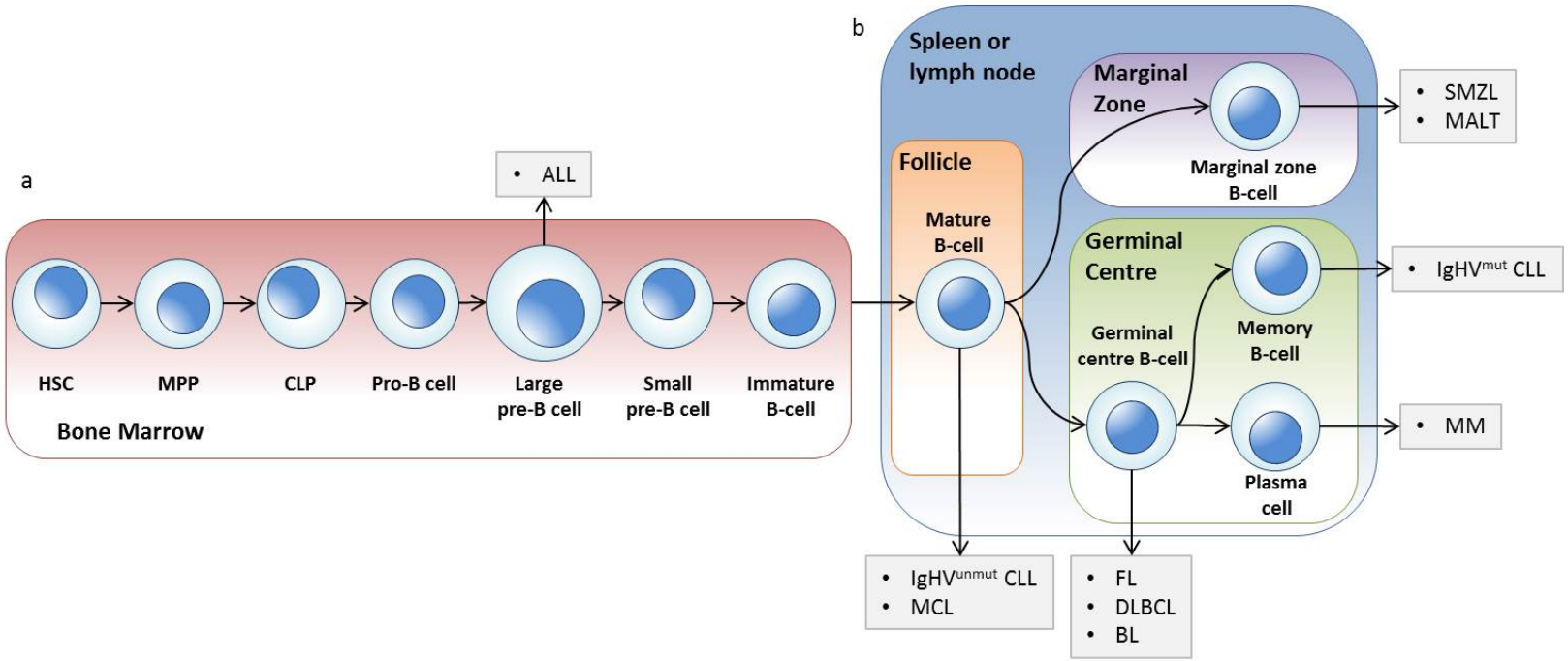

Figure 1.1 Stages of B-cell differentiation and associated malignancies.

Flowchart showing the different stages and environments involved in B-cell haematopoiesis. (**a**) Initial stages of development occur within the bone marrow, with haematopoietic stem cells (HSC) differentiating into immature B-cells. (**b**) Further development occurs within the spleen or lymph node. B-cell malignancies are associated with distinct stages of B-cell development. CLL originates in either mature B-cells, in the case of $IgHV^{unmut}$, or in memory B-cells as $IgHV^{mut}$ CLL. Abbreviations: HSC, Haematopoietic stem cell; MPP, multipotent progenitor cell; CLP, common lymphoid progenitor cell; ALL, acute lymphoblastic leukaemia; $IgHV^{unmut}$, immunoglobulin heavy chain variable unmutated chronic lymphocytic leukaemia; MCL, mantle cell lymphoma; FL, follicular lymphoma; DLBCL, diffuse large B-cell lymphoma; BL, Burkitt's lymphoma; MM, multiple myeloma; $IgHV^{mut}$, immunoglobulin heavy chain variable mutated chronic lymphocytic leukaemia; MALT, mucosa-associated lymphoid tissue lymphoma; SMZL, splenic marginal zone lymphoma. Adapted from (Rickert 2013).

1.1.1 Risk Stratification in CLL

There are two classification methods for CLL, the Rai staging system (Rai *et al* 1975), used primarily in North America, and the Binet staging system (Binet *et al* 1977) which is used in Europe. Both methods use a simple blood test, in combination with a physical examination, to classify CLL cases into either one of five groups (Rai) or one of three stages; A, B or C (Binet), each of which can be used as a prediction of overall survival (OS) rates.

1.1.1.1 Rai staging

First described in 1975 (Rai *et al* 1975), the Rai staging system initially arranged CLL cases into one of five risk categories (stages 0-IV) (Table 1.1). Patients presenting with lymphocytosis, defined as more than 15,000 lymphocytes per cubic millimetre (>15,000/mm^3), are categorised as being low risk stage 0. Stages I and II are intermediate risk, both presenting with lymphocytosis alongside enlargement of the lymph nodes (stage I), liver or spleen (both stage II). By contrast, the high risk stages III and IV may or may not present with organomegaly, being confirmed primarily by the presence of both lymphocytosis and either anaemia (haemoglobin <11g/dL) or thrombocytopaenia (platelets <100,000/mm^3) respectively. These groupings predict OS rates in new CLL cases (Table 1.1).

Table 1.1 Risk stratification in CLL according to Rai staging

Rai Stage	Biological Indicators	Median Survival (months)
0	Lymphocytosis	>150
I	Lymphocytosis with enlarged lymph nodes	101
II	Lymphocytosis with enlarged liver and/or spleen	71
III	Lymphocytosis and anaemia without thrombocytopaenia. Possible enlargement of any of lymph nodes, liver or spleen	19
IV	Lymphocytosis and thrombocytopaenia with or without anaemia	19

(Rai *et al* 1975).

1.1.1.2 Binet staging

The Binet staging system, first described in 1977 (Binet *et al* 1977), is based upon the Rai method, with a few alterations. Patients are allocated to one of five risk categories based on a blood test and a physical examination. Stage 0 cases are defined as those with lymphocytosis (>4000/mm^3) without any lymph node enlargement or splenomegaly; stage I, both lymphocytosis and lymph node enlargement; stage II, lymphocytosis with splenomegaly; stage III, lymphocytosis and both enlarged spleen and lymph nodes; stage IV, lymphocytosis with either severe thrombocytopaenia (platelets <100,000/mm^3) or severe anaemia (<10g/dL haemoglobin (Hb)). Later work by the same group refined this schema (Binet *et al* 1981), dropping from five stages to three (A, B and C) and using total number of

organ enlargements regardless of location (Table 1.2). Patients with lymphocytosis and fewer than three enlarged organs are classified as group A. Group B patients present with lymphocytosis and at least three enlarged organs, while group C patients have anaemia and/or thrombocytopaenia, regardless of the number of enlarged organs. The difference in survival rates for these groups is marked, ranging from 10 years in group A, to two years for group C (Binet *et al* 1981).

Table 1.2 Binet staging criteria in CLL

Binet Group	Hb Level (g/dL)	Platelet count (k/mm3)	Enlarged Organs	Median OS
A	>10	>100	< 3	> 10 years
B	>10	>100	>3	7 years
C	<10	<100	Any number	2 years

(Binet *et al* 1981).

1.1.2 Prognostic Markers in CLL

1.1.2.1 IgHV Mutation Status

Somatic hypermutation (SHM) is an important part of the adaptive immune response. Upon activation by a foreign antigen, a mature B-cell enters the germinal centre of a lymph node, where the intracellular enzyme activation-induced cytidine deaminase (AID) introduces point mutations at the immunoglobulin-variable gene loci. These mutations result in the generation of a large antibody repertoire.

In 1999, work by both the Hamblin and Damle groups showed that CLL patients with hypermutated IgHV (<98% homology to the germline), were predicted to have significantly increased OS rates compared to un-mutated patients (Damle *et al* 1999, Hamblin *et al* 1999). More recent work has shown that hypermutated patients with the IgHV 3-21 subtype are now known to have the same poor prognosis as those with un-mutated IgHV (Thorsélius *et al* 2006, Tobin *et al* 2002).

1.1.2.2 Cytogenetics

Recurrent copy number aberrations (CNAs) are common in CLL, and can have a significant impact on the prognosis of the disease. Indeed, the presence or absence of a particular CNA can be used for risk stratification in new CLL cases (Döhner *et al* 2000, Pflug *et al* 2014). One of the most commonly used method of detecting CNAs is fluorescent *in-situ* hybridisation (FISH). This method involves the hybridisation of fluorescently labelled probes complementary to a specific target region of the genome. Subsequent visualisation with a microscope reveals information regarding the presence, absence or abnormal copy-number of the region in question. More recently molecular methods, including array-based assays, are gaining acceptance as alternatives to FISH. These arrays, by interrogating SNPs genome-wide, provide information on copy-number changes and loss of heterozygosity along the length of each chromosome. Furthermore, by comparing the genomic position of recurrent CNAs in a

cohort of patients, it is possible to identify which genes are critical to the pathogenesis of the disease.

1.1.2.2.1 Del(11q)

Full or partial deletions of the q-arm of chromosome 11 (del(11q)) are the second most frequent chromosomal aberration found in CLL, occurring in 10-20% of patients (Döhner *et al* 1997, Edelmann *et al* 2012, Hernández *et al* 2015, Neilson *et al* 1997, Pflug *et al* 2014, Puente *et al* 2015). These deletions are associated with rapid disease progression (Döhner *et al* 1997, Doneda *et al* 2003) represented by significantly lower OS rates than in patients with a normal karyotype (Döhner *et al* 1997, 2000, Neilson *et al* 1997) (Table 1.3). Despite del(11q) CLL being classified as a high-risk disease, recent data have shown that patients with del(11q) as sole cytogenetic abnormality respond well to monoclonal antibody based treatment regimes. For example, phase II and phase III trials of chemoimmunotherapy combining purine analogues with the anti-CD20 monoclonal antibody Rituximab, showed partial remission rates of up to 90% in del(11q) cases, with up to 40% achieving complete remission (Fischer *et al* 2011, 2012, Hallek *et al* 2010). Interestingly, recent data have shown that the percentage of del(11q) positive CLL cells present in a patient can affect the disease course. Patients with reduced numbers (<40%) of del(11q) containing cells show both longer time to first treatment (TTFT) (44 months vs 19 months) and OS (157 months vs 90

months) compared to those with higher (≥40%) levels of del(11q) cells (Hernández *et al* 2015). The chromosomal region most commonly deleted as a result of del(11q), known as the minimally deleted region (MDR) is a small section located at the q22.3 band of chromosome 11. Specifically, this is the location of the tumour suppressor gene *ATM* (Döhner *et al* 1997, Stilgenbauer *et al* 1996).

1.1.2.2.2 Trisomy 12

Gains of chromosome 12 (trisomy 12) are found in 10-17% of CLL cases (Döhner *et al* 2000, Edelmann *et al* 2012, Pflug *et al* 2014, Puente *et al* 2015) and confers an intermediate risk, with progression free survival times similar to patients with a normal karyotype (Döhner *et al* 2000, Van Dyke *et al* 2016) (Table 1.3). However, other studies have linked the presence of trisomy 12 in CLL to disease progression, specifically the development of an aggressive secondary lymphoma known as Richters syndrome (RS). In one study, using high density SNP arrays, trisomy 12 was present as the sole cytogenetic abnormality in 28% of CLL cases who went on to develop RS (Chigrinova *et al* 2013). The presence of trisomy 12 in more than 60% of CLL cells has been shown to predict for reduced progression free survival (González-Gascón *et al* 2015). Mutations in the *NOTCH1* gene have also been associated with trisomy 12 in CLL (Balatti *et al* 2012), leading to a significant reduction in OS rates compared to trisomy 12 cases with un-mutated *NOTCH1* (Del Giudice *et al* 2012).

1.1.2.2.3 Del(13q)

Deletions of the q-arm of chromosome 13 are the most frequent chromosomal aberration in CLL, occurring in 30-55% of cases (Döhner *et al* 2000, Van Dyke *et al* 2016, Edelmann *et al* 2012, Pflug *et al* 2014, Puente *et al* 2015). Del(13q) offers the best prognostic outlook with longer progression free and OS than patients with normal karyotypes (Döhner *et al* 2000, Van Dyke *et al* 2016) (Table 1.3). The minimally deleted region centres on the 13q14 region, which contains a number of genes, most notably the micro-RNAs (miRNA) mir-15a and mir-16-a (Mertens *et al* 2009). These miRNAs negatively regulate the activity of tumour protein 53 (*TP53*) by binding to the 3'UTR, targeting it for degradation. Patients with del(13q) exhibit reduced mir15a and mir16a expression levels (Mosca *et al* 2010), which effectively decreases the level of TP53 down-regulation (Lin *et al* 2014). Furthermore, the proportion of del(13q) positive CLL cells within a patient is also linked to the prognosis. Cases with a higher del(13q) load show significantly shorter treatment free survival (TFS) and OS (Huang *et al* 2016, Puiggros *et al* 2013).

1.1.2.2.4 Del(17p)

Deletions of the p-arm of chromosome 17 are found in 4-12% of CLL cases prior to frontline therapy and confer a poor prognosis, with rapidly progressing disease and reduced OS and TFS (Döhner *et al* 2000, Van Dyke *et al* 2016, Edelmann *et al* 2012, Pflug *et al* 2014, Puente *et al* 2015) (Table

1.3). These deletions focus on the 17p13 band, which also harbours the well-characterised tumour suppressor gene *TP53*. Patients with del(17p) demonstrate low response rates to many chemotherapy regimens (Byrd *et al* 2006), with combination therapies offering no significant improvement (Bosch *et al* 2008, 2009, Fischer *et al* 2011, Hallek *et al* 2010, Mauro *et al* 2014). However, treatment of del(17p) CLL with the Brutons tyrosine kinase (BTK) inhibitor Ibrutinib does show positive results, with a 30 month progression free survival of 96% and OS of 97% (Byrd *et al* 2015, O'Brien *et al* 2014).

Table 1.3 Overall and progression free survival rates associated with cytogenetic subgroups in CLL.

	Median Overall Survival (%)		Median Progression Free Survival (%)		
Genetic Abnormality	**5 year**	**10 year**	**3 year**	**5 year**	**10 year**
Del(17p)	22	N/R	7	N/R	N/R
Del(11q)	68	16	27	15	N/R
Gain(12)	77	30	50	36	13
Normal	80	50	60	42	16
Del(13q)	95	62	73	66	31

Abbreviations: N/R, not reached. (Döhner *et al* 2000).

1.1.2.3 ZAP70

ZAP70 is a protein tyrosine kinase, located at the q11.2 band of chromosome 2, which plays a role in B-cell development. Overexpression of ZAP70 enhances the activity of the B-cell receptor (BCR) pathway, a key mechanism for the survival of CLL cells. Early studies found that ZAP70

expression levels may provide a useful surrogate marker for IgHV mutation status, as those cases where >20% CLL cells showed high ZAP70 expression were also IgHV un-mutated (Crespo *et al* 2003, Dürig *et al* 2003, Wiestner *et al* 2003). Additional work identified it as a strong independent predictor of the need for treatment (Rassenti *et al* 2004). More recent data propose that the more aggressive disease course presented by this subset of CLL may be due to ZAP70 expression enhancing the migration capacity of the CLL cells, driving them towards the advantageous microenvironment of the bone marrow (Calpe *et al* 2013, Stamatopoulos *et al* 2009). The EGFR inhibitor Gefitinib, usually used to treat lung cancer, has shown some promise as a treatment regime, preferentially inducing cell death in ZAP70 positive cells (Dielschneider *et al* 2014).

1.1.2.4 CD38

CD38 is a transmembrane glycoprotein, encoded for by a gene located at 4p15, which plays a role in the regulation of intracellular Ca^{2+}. When expressed on the surface of >30% of CLL cells, CD38 is an independent predictor of shorter OS (Damle *et al* 1999). CD38 expression associates with unmutated IgHV genes and poor response to chemotherapy. Further studies found that patients with a lower threshold of 20% CD38+ cells also displayed shorter survival times and a more aggressive disease course (Ibrahim 2001). In a manner similar to ZAP70, CD38+ patients exhibited

enhanced migration of their CLL cells towards the bone marrow via CXCL12 signalling (Deaglio *et al* 2007, Vaisitti *et al* 2010).

1.1.2.5 CD49d

CD49d, an α-integrin subunit that forms part of the α4β1 lymphocyte homing receptor, has been shown to be a robust predictor of outcome in CLL patients. The gene encoding this protein is located at the q31.3 band of chromosome 2. High expression of CD49d ($CD19d^{high}$), defined as the presence of CD49d in ≥30% of CLL cells, is associated with both shorter OS and reduced time to treatment (Gattei *et al* 2008, Zucchetto *et al* 2006) compared to $CD49d^{low}$ CLL. Further studies found that $CD19d^{high}$ CLL cells would also express both CD38 and ZAP70, suggesting that CD49d may be linked to the increased migration of CLL cells into preferential micro-environments (Nückel *et al* 2009, Shanafelt *et al* 2008). CD38 and CD49d work in tandem at opposite ends of the same pathway to promote the survival of CLL cells via VCAM-1 (Zucchetto *et al* 2009). As an indicator of disease progression in CLL, the addition of CD49d expression to other prognostic markers can improve the power of the analysis (Majid *et al* 2011). It can also be used to further subdivide the mutated IgHV subgroup, predicting a shorter time to first treatment for $IgHV^{mut}/CD49d^{high}$ compared to $IgHV^{mut}/CD49d^{low}$ (Baumann *et al* 2016).

1.1.2.6 β2-Microglobulin

β2-microglobulin (β2M) is a component of the major histocompatibility complex (MHC) class I protein, present on the surface of all nucleated cells. It is also found as a free molecule in the surrounding serum. Levels of serum β2M were found to be elevated in CLL patients compared to age-matched controls, 5mg/L and 2mg/L respectively, and that levels >4mg/L inferred a much lower survival time over cases with <4mg/L (12 months and 43 months respectively) (Di Giovanni *et al* 1989). Increased β2M levels also indicate shorter time to first treatment, compounded by a lower probability of achieving complete remission with combination therapies such as Fludarabine-Cyclophosphamide-Rituximab (FCR) (Delgado *et al* 2009, Keating *et al* 2005). More recent studies, with larger cohorts, have confirmed β2M serum levels as an independent marker of shorter PFS, OS and poor treatment response, and correlated β2M serum levels with other markers of poor prognosis, including unmutated IgHV and high CD38 levels (Oscier *et al* 2010, Pratt *et al* 2016).

1.1.3 Recurrent Gene Mutations in CLL

CLL is a highly heterogeneous disease, both clinically and biologically. Despite the prognostic value of both the common chromosomal abnormalities and cell surface markers, more than 50% of patients will go on to develop disease progression and chemo-refractoriness (Tam *et al* 2014). This clinical heterogeneity is likely due to underlying molecular

differences between patients. A number of explorative genome-wide (Puente *et al* 2011, Schuh *et al* 2012) and exome-wide (Quesada *et al* 2012) sequencing efforts have identified a number of recurrently acquired mutations in the coding regions of genes. The presence of mutations in genes including *TP53* (Zenz *et al* 2008, 2009, 2010b), *SF3B1* (Quesada *et al* 2012, Rossi *et al* 2011, Wang *et al* 2011) and *NOTCH1* (Fabbri *et al* 2011, Rossi *et al* 2012, Rossi *et al* 2011) have been associated with chemo-refractoriness, advanced disease and poor prognosis.

1.1.3.1 NOTCH1

The NOTCH signalling pathway is involved in a number of crucial cell functions, including proliferation, cell differentiation and apoptosis. *NOTCH1* is a transmembrane protein that cleaves its intracellular domain (*NOTCH1*^{IC}) upon activation by an extracellular ligand. *NOTCH1*^{IC} translocates to the nucleus and forms a complex with *CBF1*, *MAML* and p300 to become a transcriptional activator for *NOTCH* target genes. *NOTCH1* is both upregulated and constitutively activated in CLL cells compared to normal controls, and is associated with apoptosis resistance (Rosati *et al* 2009). A 2bp deletion in the *NOTCH1* PEST domain (del7544_45) is found in 4-12% of CLL cases (Di Ianni *et al* 2009, Puente *et al* 2011, Rossi *et al* 2011, Sportoletti *et al* 2010) and results in early termination of the PEST domain required for protein degradation. As a result, *NOTCH1*^{IC} accumulates in the nucleus leading to constitutive activation of *NOTCH1* target genes (Figure 1.2). These *NOTCH1* mutations

are an independent predictor for shorter OS (Rossi *et al* 2011), with a 10-year survival of 21% compared to 56% with wild-type *NOTCH1* (Puente *et al* 2011). Additionally, *NOTCH1* mutations are found in 24% of trisomy 12 CLL samples and confer a 2.4x increase in the risk of death and a shortening of OS in this subgroup compared to trisomy 12 alone (Del Giudice *et al* 2012). *NOTCH1* mutations have also been associated with an increased risk of transformation of CLL into diffuse large B-cell lymphoma (DLBCL) (Villamor *et al* 2013).

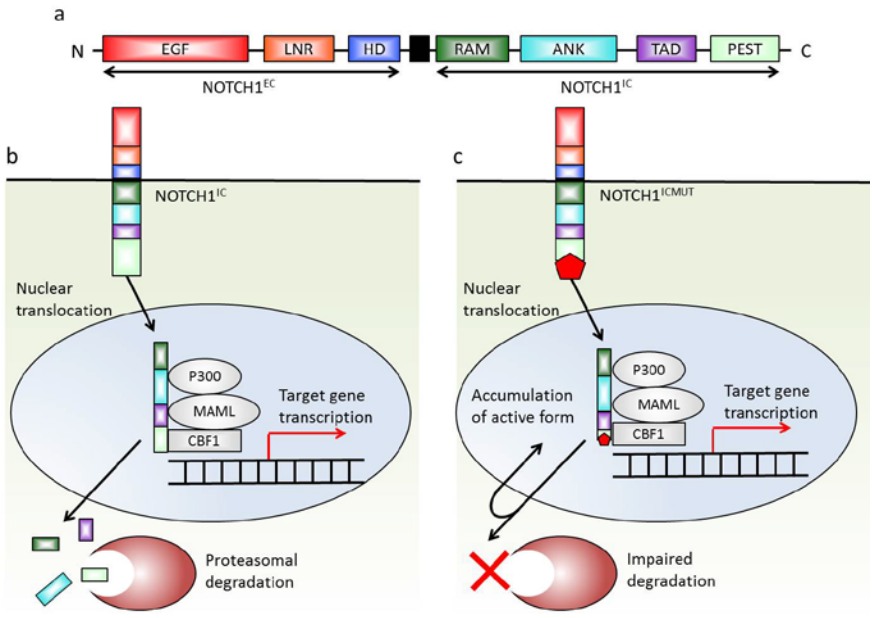

Figure 1.2 Constitutive activation of NOTCH1 in CLL.

(**a**) Schematic representation of the *NOTCH1* transcript. (**b**) Following target gene transcription, wild-type NOTCH1 is targeted for proteasomal degradation, (**c**) whilst *NOTCH1* with a disrupted PEST domain (*NOTCH1*^{ICMUT}) escapes degradation and accumulates in the nucleus. Abbreviations: *NOTCH1*^{EC}, extracellular domain; *NOTCH1*^{IC}, intracellular domain; EGF, epidermal growth factor like domain; LNR, Lin-Notch repeat domain; HD, heterodimerisation domain; RAM, RBPJ-associated molecule domain; ANK, ankyrin repeat domain; TAD, transactivation domain; PEST, Proline-Glutamate-Serine-Threonine rich domain.

1.1.3.2 TP53

TP53 is a tumour suppressor gene located at 17p13.1 in humans, composed of 11 exons, 10 of which are coding regions. The translated protein consists of an N-terminal transactivation domain (TAD), a highly conserved DNA-binding domain, and a C-terminal oligomerisation domain (Figure 1.3). *TP53* expression is increased in response to stress signals or following DNA damage, resulting in *TP53* mediated initiation of one of several key pathways in an attempt to save the cell; including activation of DNA repair proteins, arresting cell growth at the G1/S point, or the initiation of apoptosis as a last resort. Mutations in *TP53* (*TP53*^{mut}) are one of the most frequently seen genomic alterations in many human cancers. In CLL, *TP53* mutations are typically found in 8-15% of cases and are a marker of poor prognosis, being associated with a reduction in both OS and TTFT (Dufour *et al* 2013, Landau *et al* 2013, el Rouby *et al* 1993, Zenz, Eichhorst, *et al* 2010). The impact of *TP53*^{mut} on the disease course is the same as that inferred by del(17p) (Rossi *et al* 2009). A number of trials have linked the presence of *TP53*^{mut} with resistance to chemotherapy, particularly with Fludarabine based regimes (Gonzalez *et al* 2011, Zenz, Eichhorst, *et al* 2010). This is reflected in the increased incidence of *TP53*^{mut} in Fludarabine-refractory (FR) patient cohorts (Messina *et al* 2014, Schnaiter *et al* 2013). However, recent work indicates that *TP53*^{mut} patients do respond to treatment with the monoclonal antibody Alemtuzumab and novel B-cell receptor inhibitors such as BTK (Ibrutinib, Acalabrutinib) or phosphoinositide 3-kinase (PI3K) inhibitors (Idelalisib) and BCL-2 inhibitors

such as Venetoclax (Pettitt *et al* 2012, Sciumè *et al* 2015). Studies of *TP53*^{mut} CLL samples, using whole-exome sequencing techniques, have revealed that *TP53*^{mut} CLL cells tend to be subclonal at diagnosis before becoming the dominant clone at relapse. This suggests that firstly, *TP53* mutations contribute towards disease progression rather than initiation and secondly, that the presence of these mutations at low level at diagnosis is a reliable indicator of the rapid development of chemo-resistance (Landau *et al* 2013, Rossi *et al* 2014).

1.1.3.3 SF3B1

Splicing factor 3B subunit 1 (*SF3B1*) forms part of the U2 small nuclear ribonucleoprotein complex that plays a role in mRNA splicing. Mutations in *SF3B1* are common in CLL, present in 10-20% of cases (Jeromin *et al* 2014, Landau *et al* 2013, Messina *et al* 2014, Quesada *et al* 2012, Rossi, Bruscaggin, *et al* 2011, Stilgenbauer *et al* 2014). These mutations are typically missense variants clustered in the highly conserved HEAT domain of *SF3B1*, with 42-50% of mutations affecting the lysine residue at position 700 (Jeromin *et al* 2014, Rossi, Bruscaggin, *et al* 2011, Stilgenbauer *et al* 2014, Wang *et al* 2011). This recurrently affected region is predicted to form the inner surface of the *SF3B1* protein, and therefore may represent disruption of a binding site, leading to the altered splicing function seen in these samples (Lili Wang *et al* 2011). Mutations in *SF3B1* tend to be sub-

clonal that expand over time contributing to disease progression, particularly following chemotherapy (Landau *et al* 2013).

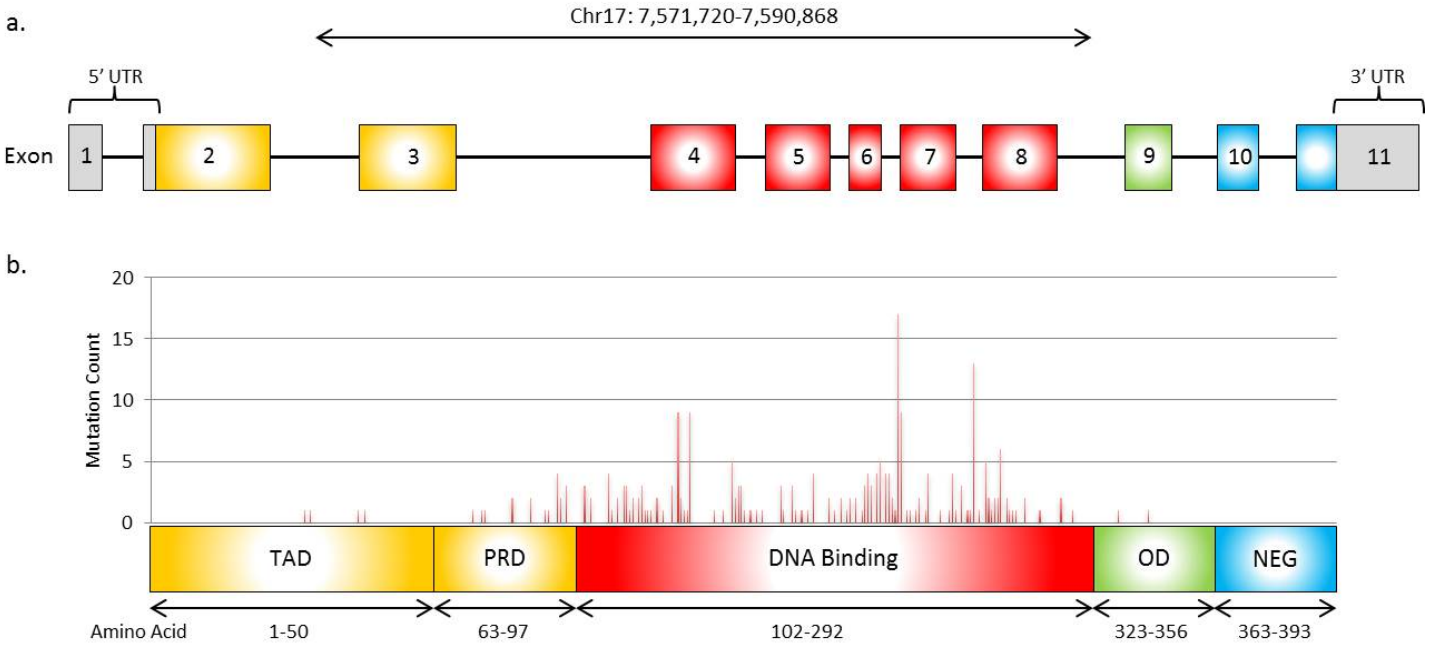

Figure 1.3 Distribution of TP53 mutations in chronic lymphocytic leukaemia

(**a**) Schematic representation of the *TP53* gene on chromosome 17. (**b**) Distribution of coding mutations reported in CLL according to the COSMIC database (http://cancer.sanger.ac.uk/cosmic, Forbes *et al* 2015) (**b**). The majority of reported mutations (78%, 292/304) are located within the DNA binding domain, encoded for by exons four to eight. Abbreviations: TAD, transactivation domain; PRD, proline-rich domain; OD, oligomerisation domain; NEG, negative regulation domain.

1.1.3.4 BIRC3

BIRC3 is a negative regulator of the NF-κB signalling pathway, a known pro-survival mechanism in CLL. The *BIRC3* protein, in combination with TRAF2 and TRAF3, negatively regulates the MAP3K14, which stimulates NF-κB via the non-canonical pathway. Mutations in *BIRC3* are enriched in Fludarabine-refractory CLL samples, occurring in 15-24% of cases (Messina *et al* 2014, Rossi *et al* 2012) and are predominantly frameshift variants that remove the c-terminal RING domain of the protein. This domain is required for the degradation of *MAP3K14*. These mutations lead to constitutive activation of the NF-κB pathway in CLL. Recent data indicate that BIRC3 disruption predicts for lower complete remission rates following first line treatment with Chlorambucil and Rituximab (Foà *et al* 2014).

1.1.4 Influence of Gene Mutations on Risk Stratification

Given the number of recurrently mutated genes recently identified in CLL, and their demonstrable impact on disease course, it seems reasonable to examine whether integration of these new prognostic markers into existing algorithms could offer an improvement in patient risk stratification. A study by Rossi *et al* (2013) devised a prognostic scoring system that integrated cytogenetic information with mutation screening data in four frequently mutated CLL genes (*TP53*, *BIRC3*, *NOTCH1* and *SF3B1*) in a cohort of newly diagnosed CLL cases. The authors generated a hierarchical model with four risk groups; high, intermediate, low and very low, with 10-year survival

rates of 37.7%, 48.5%, 70.7% and 84.2% respectively (Table 1.4). In addition, a recent study in relapsed/refractory CLL samples revealed that the presence of subclones with >1 mutation in a driver gene has an impact on the prognosis of the disease, being associated with poorer overall response to treatment (Guièze *et al* 2015).

Table 1.4 Prognostic risk categories in CLL.

Risk Group	Mutated Genes	Cytogenetics	10 Year Survival (%)
High	*TP53* and/or *BIRC3*	del(17p)	37.7
Intermediate	*NOTCH1* and/or *SF3B1*	del(11q)	48.5
Low	-	gain(12)	70.7
Very Low	-	del(13q)	84.2

10-year survival rates are expressed in comparison to those seen in a matched general population (Rossi *et al* 2013).

The inclusion of gene mutations was found to significantly improve survival prediction (Baliakas *et al* 2014), and highlights the need for mutation screening at the diagnosis stage.

1.2 Acute Myeloid Leukaemia

Acute myeloid leukaemia (AML) is characterised by the accumulation of immature, undifferentiated myeloid blast cells in the bone marrow and peripheral blood. A diagnosis of AML is confirmed when the proportion of these blasts exceeds the 20% threshold in either peripheral blood or bone marrow. AML is the most common myeloid malignancy in Europe, with an overall incidence of 3.7 new cases per 100,000 per year. This number increases with age, and is slightly more prevalent in males than females by a ratio of 1.2:1 (Visser *et al* 2012). Risk stratification in AML is based on the presence of cytogenetic abnormalities (Grimwade 2012), despite ~40% of cases having a normal karyotype (Grimwade *et al* 2010). The discovery of recurrently mutated genes in AML has added to the already considerable heterogeneous picture of this disease.

1.2.1 Clinical Classification

A number of different algorithms have been adopted for the classification of AML, involving a combination of morphological, cytochemical and molecular information. The French-American-British (FAB) group first devised a classification system for AML in 1976. The system uses a combination of the originating cell type and the level of cell differentiation to classify AML cases into one of eight subgroups (M0-M7) (Bennett *et al* 1976, 1985) (Table 1.5). The World Health Organisation (WHO) developed a

more comprehensive system that improves upon the FAB classification by integrating information regarding recurring chromosomal abnormalities and gene mutations into the FAB system. The WHO classification identifies seven AML subtypes (Vardiman *et al* 2002, 2009) (Table 1.6). The recurring genetic abnormality subgroup predominantly concerns the presence of translocations involving the *RUNX1*, *RARA*, *MLL* and *CBFB* transcription factors although point mutations in genes including *NPM1* and *CEBPα* are being added with each revision.

Table 1.5 FAB classification of AML.

FAB Group	Name
M0	Undifferentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia
M4	Acute myelomonocytic leukaemia
M4eos	Acute myelomonocytic leukaemia with eosinophilia
M5	Acute monocytic leukaemia
M6	Acute erythroid leukaemia
M7	Acute megakaryocytic leukaemia

(Bennett *et al* 1976).

1.2.2 Structural Variants in AML

In addition to classifying the subtype of AML at diagnosis, the presence of translocations can also add value in the context of risk stratification and treatment response predication.

Core binding factor AML (CBF-AML) comprises 15% of adult AML cases (Solh *et al* 2014) and is characterised by the presence of either the t(8;21) or inv(16) chromosomal abnormalities. The t(8;21) event fuses the *RUNX1* gene with *RUNX1T1*, whilst inv(16) joins the *MYH11* gene to *CBFB*. Both *RUNX1* and *CBFB* form part of the core binding factor complex, a regulator of haematopoiesis. These fusion proteins repress a number of genes involved in hematopoietic differentiation (Alcalay *et al* 2003, Kundu and Liu 2001, Mandoli *et al* 2014). CBF-AML cases display improved OS rates, with up to 61% of cases surviving 10 years from diagnosis (Grimwade *et al* 1998, 2010). Treatment of CBF-AML with high-dose Cytarabine (HiDAC) achieves a five year complete remission rate of 78% (Bloomfield *et al* 1998) and decreases the likelihood of relapse (Byrd *et al* 2004). Combination therapies can also significantly improve overall survival in this AML subtype (Borthakur *et al* 2008, Kaspers *et al* 2013).

Acute promyelocytic leukaemia (APL) is characterised by the presence of t(15;17), in which the promyelocytic leukaemia (*PML*) and retinoic acid receptor alpha (*RARA*) genes are fused together. APL accounts for 13% of new AML cases and is considered as favourable risk, with a 10-year survival rate of 81% (Grimwade *et al* 2010). Treatment of APL with all-trans retinoic acid (ATRA), often in combination with anthracyclines (Sanz *et al* 2015), achieves complete remission in up to 69% of cases (Castaigne *et al* 1990, Tallman *et al* 1997), with 77% of patients surviving to 10 years (Adès *et al* 2010).

AML with t(9;11) (*MLLT3*;*MLL*) is classified as intermediate risk, with 10 year survival rates of 39% (Grimwade *et al* 2010). The mixed lineage leukaemia gene (*MLL*) acts as a regulator of *HOX* gene expression via the

Table 1.6 Classification of acute myeloid leukaemia and related neoplasms according to 2008 WHO guidelines.

1.	**Acute myeloid leukaemia with recurrent genetic abnormalities**
	AML with t(8;21) (q22;q22); *RUNX1-RUNX1T1*
	AML inv(16) (p13.1q22) or t(16;16) (p13.1;q22); *CBFB-MYH11*
	APL with t(15;17)(q22;q12); *PML-RARA*
	AML with t(9;11)(p22;q23); *MLLT3-MLL*
	AML with t(6;9)(p23;q34); *DEK-NUP214*
	AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*
	AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1*
	*AML with mutated *NPM1*
	*AML with mutated *CEBPα*
2.	**Acute myeloid leukaemia with myelodysplasia-related changes**
3.	**Therapy-related myeloid neoplasms**
4.	**Acute myeloid leukaemia, not otherwise specified**
	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukaemia
	Acute monoblastic/monocytic leukaemia
	Acute erythroid leukaemia
	Pure erythroid leukaemia
	Erythroleukaemia, erythroid/myeloid
	Acute megakaryoblastic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis
5.	**Myeloid sarcoma**
6.	**Myeloid proliferations related to Downs syndrome**
	Transient abnormal myelopoiesis
	Myeloid leukaemia associated with Down syndrome
7.	**Blastic plasmacytoid dendritic cell neoplasm**

* Indicates provisional entity (Swerdlow *et al* 2008).

modification of histones and promoter binding (Hess 2004). The presence of t(9;11) disrupts the normal function of *MLL* and upregulates the *HOX* gene targets. Patients with t(9;11) generally display longer complete remission (CR) and EFS duration, along with improved OS rates in comparison to AML with other 11q translocations (Mrózek *et al* 1997, Rubnitz *et al* 2002). Despite this, t(9;11) AML is associated with high rates of relapse following chemotherapy (Chandra *et al* 2010).

AML with t(6;9) is also an intermediate risk disease, with a trend towards poorer outcome. Twenty seven percent of patients survive for 10 years after diagnosis (Grimwade *et al* 2010). This translocation fuses the 5' region of *DEK* on chromosome 6 to the 3' region of *CAN* on 9q34 (von Lindern *et al* 1992, Soekarman, von Lindern, Daenen, *et al* 1992, Soekarman, von Lindern, van der Plas, *et al* 1992). The *CAN* gene encodes for *NUP214*, a member of the FG-repeat containing nucleoporin family, whilst *DEK* is a proto-oncogene. The fusion protein initiates leukaemogenesis in a small subset of the haematopoietic stem cell (HSC) population (Oancea *et al* 2010).

By contrast, AML with inv(3)(q21;q26.2) represents a particularly aggressive disease course with poor survival (Fonatsch *et al* 1994, Grimwade *et al* 2010, Reiter *et al* 2000). These events affect the *EVI1* oncogene at 3q26.2 and *RPN1* at 3q21.

1.2.3 Recurrent Gene Mutations in AML

Despite the utility of chromosomal alterations in AML diagnosis, there remain 41% of cases presenting with cytogenetically normal AML (cn-AML) (Grimwade *et al* 2010), which thus fall into an indiscriminate, intermediate risk category. Over the course of the last decade, the addition of molecular findings, in particular mutations in a number of key genes, has added prognostic value to this category. The 2008 update to the WHO classification of myeloid neoplasms and acute leukemia (Swerdlow *et al* 2008) added both AML with *NPM1* mutations and AML with *CEBPα* mutations as two distinct, provisional, categories, reflecting the prognostic relevance of these genes. The next update to the classification is due in September 2016, and will likely add further molecular markers to the system.

1.2.3.1 NPM1

Nucleophosmin is a chaperone protein that continuously shuttles between the cytoplasm and the nucleus. It is encoded for by the *NPM1* gene on 5q35.1 and is translated into a protein, 294 amino acids in length. *NPM1* is induced in response to cellular stress, whereupon it protects *TP53* activity by repressing *HDM2* mediated *TP53* degradation (Colombo *et al* 2002, Kurki *et al* 2004). Insertions of 4bp sequences, causing a Trp288fs*12 frameshift, have been identified in 45-60% of cn-AML patients (Falini *et al* 2005, Schlenk *et al* 2008, Schnittger *et al* 2005, Christian Thiede *et al* 2006). These insertions cause the final seven amino acids of the C-terminal

domain to be substituted with 11 new residues (Figure 1.4). The new sequence contains an additional nuclear export amino acid sequence (NES); L-xxx-V-xx-V-x-L (Bolli *et al* 2007, Falini *et al* 2006), which facilitates the translocation of the *NPM1* protein into the cytoplasm. Whilst the location of the mutation in exon 12 remains constant, the inserted sequence can vary. The most common mutation is a TCTG insertion, accounting for 58% of *NPM1* mutated patients (Falini *et al* 2005), followed by CATG (3%) and CCTG (3%) insertions. On a functional level, the substitution of the tryptophan residue at position 288 with a cysteine, as is the case with the most commonly seen *NPM1* variants, induces increased sensitivity to chemotherapy in AML cells (M Huang *et al* 2013, Schneider *et al* 2009). As a result, *NPM1* mutations in cn-AML are associated with higher CR rates, longer PFS and longer OS in both adult and childhood cases, in the absence of concomitant *FLT3* mutations (Döhner *et al* 2005, Hollink *et al* 2009, Schlenk *et al* 2008, Schnittger *et al* 2005, Thiede *et al* 2006). Furthermore, monitoring the level of mutated *NPM1* in a post-treatment tumour cell population can indicate relapse in cn-AML patients (Krönke *et al* 2011, Shayegi *et al* 2013).

1.2.3.2 CEBPα

The gene encoding for CCAAT/enhancer binding protein alpha (*CEBPα*) consists of a single exon 2.6kb in length, located at 19q13.1, and is involved in the differentiation of myeloid cells. Two protein products are generated

from this gene, the full-length canonical p42 isoform, and a shorter p30 isoform that is not expressed in wild-type cells. Mutations in *CEBPα* have been identified in 7-10% of AML cases (Fasan *et al* 2014, Hollink *et al* 2011, Leroy *et al* 2005, Pabst *et al* 2001, Schlenk *et al* 2009, Wouters *et al* 2009), and are distributed throughout the coding region, with distinct clusters in both the C- and N-termini. Mutations in the N-terminal domain are predominantly frameshift insertions or deletions resulting in the expression of a truncated 20kD protein (Pabst *et al* 2001), whereas C-terminal variants are generally in-frame insertions or deletions, disrupting the DNA binding domain (Fasan *et al* 2014). Three patterns of mutation have been identified in *CEBPα*: 1; a single mutation on one allele (*CEBPα*sm), 2; mutations in both alleles (bi-allelic mutations, *CEBPα*dm), and 3; homozygous mutations due to loss of heterozygosity (LOH) (Fasan *et al* 2014, Wouters *et al* 2009). In 75% of *CEBPα*dm cases, mutations are found in both the C- and N-terminus domains (Fasan *et al* 2014). Studies show that *CEBPα*dm is a favourable prognostic marker, with improved rates of OS, EFS and relapse risk compared to wild-type *CEBPα* (Ahn *et al* 2016, Wouters *et al* 2009).

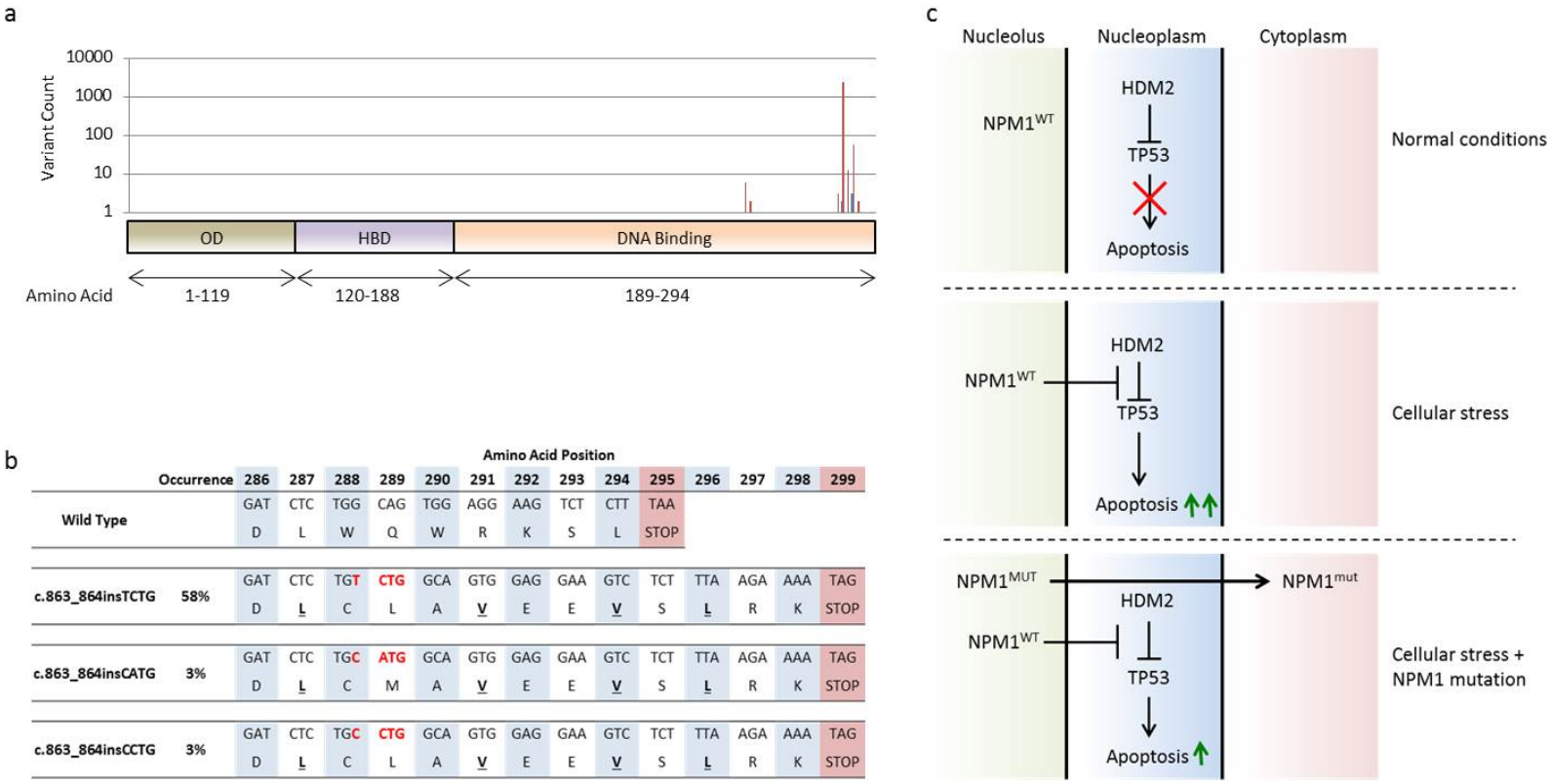

Figure 1.4 NPM1 mutations in AML.

(**a**) Location and frequency of indel (red) and missense (blue) variants in *NPM1* in AML as recorded in the COSMIC database. (**b**) Each of the three most common insertions in NPM1 (red) affects the tryptophan residue at position 288, extends the protein by four residues, and introduces an additional L-xxx-V-xx-V-x-L NES (underlined and bold residues). (**c**) Action of *NPM1* protein under normal and stress conditions, with and without *NPM1* mutations. Abbreviations: OD, oligomerisation domain; HBD, histone binding domain.

1.2.3.3 FLT3

Fms-related tyrosine kinase 3 (*FLT3*) is a receptor tyrosine kinase involved in the regulation of haematopoiesis. Mutations in *FLT3* take one of two forms, variable length in-frame internal tandem duplications (FLT3-ITD) in the juxtamembrane domain, or a single amino acid substitution of residue Asp835 in the activation loop (FLT3-TKD) (Abu-Duhier *et al* 2001). The incidence of FLT3-ITD varies from 20-34%, depending on clinical characteristics (Levis 2013), whilst FLT3-TKD mutations occur in 7-10% of cases (Fröhling *et al* 2002, Kottaridis *et al* 2001, Meshinchi *et al* 2006, Schlenk *et al* 2008, Christian Thiede *et al* 2002). Both the ITD and D835 mutations result in constitutive activation of *FLT3* (Abu-Duhier *et al* 2001, Yamamoto *et al* 2001) leading to the induction of downstream targets involved in signal transduction, cell proliferation and the differentiation and function of haematopoietic cells, including the signal transducer and activator of transcription 5 (*STAT5*) and mitogen activated protein kinase (*MAPK*) genes (Hayakawa *et al* 2000, Mizuki *et al* 2000, 2003, Zhang *et al* 2000). Mutations of FLT3 are considered to be a marker of poor prognosis with significant reductions of both PFS and OS (Fröhling *et al* 2002, Kottaridis *et al* 2001, Meshinchi *et al* 2006, Schlenk *et al* 2008, Christian Thiede *et al* 2002). A number of randomised trials have shown good response and complete remission rates with a variety of combination therapies, although these are often of short duration with relapse quickly following (Fiedler *et al* 2015, Mrózek *et al* 2012, Ravandi *et al* 2010, Stone *et al* 2012). Although not included in the current WHO classification system (Swerdlow *et al* 2008), it has been recommended that screening for mutations in *FLT3*, alongside both *NPM1*

and *CEBPα*, should be carried out for cn-AML patients who are likely to receive treatment other than low-dose chemotherapy (Döhner *et al* 2010).

1.2.3.4 DNMT3a

DNA methylation is an epigenetic mechanism through which the expression of individual genes can be carefully controlled. The addition of a methyl group (CH_3) to the fifth carbon atom of a cytosine residue results in the formation of 5-methylcytosine. These alterations predominantly occur at CpG islands, which often coincide with the location of gene promoter sequences. Methylation at these locations is associated with reduced expression levels of nearby genes (Deaton and Bird 2011). Aberrant DNA methylation is evident in a number of cancer types, either reducing the expression of tumour suppressors or increasing the expression of oncogenes. The methylation of cytosine residues is a result of the activity of the DNA methyltransferase family of enzymes. One of these genes, DNA cytosine-5-methyltransferase 3a (*DNMT3a*) is frequently mutated in AML, with mutations seen in 17-24% of cases (Gaidzik *et al* 2013, Hou *et al* 2012, LaRochelle *et al* 2011, Ley *et al* 2010, Luskin *et al* 2016, Shivarov *et al* 2013, Thol *et al* 2011, Yan *et al* 2011). Point mutations in *DNMT3a* are overwhelmingly concentrated on a single residue, Arg882, with 60% of all variants occurring at this position (Ley *et al* 2010). Patients with DNMT3a mutations are shown to have a significant reduction in global methylation levels (Qu *et al* 2014, Russler-Germain *et al* 2014, Yan *et al* 2011), leading to upregulation of downstream targets, including members of the

HOX gene family and *IDH1* (Qu *et al* 2014, Yan *et al* 2011). The utility of *DNMT3a* mutations as a prognostic factor, in terms of OS and PFS is unclear. A number of studies and meta-analyses demonstrated that the presence of a *DNMT3a* mutation predicted for reduced OS, TTF and RFS (Hou *et al* 2012, Ley *et al* 2010, Renneville *et al* 2012, Shivarov *et al* 2013, Thol *et al* 2011, Yan *et al* 2011), whilst conflicting studies report no impact on either OS or RFS (Gaidzik *et al* 2013, Marková *et al* 2012). One reason for this apparent discrepancy may be the effect of the induction therapy given as part of some clinical trials. It has been shown that increasing the dosage of anthracycline, in particular Daunorubicin, improves the OS of *DNMT3a* mutated AML, with 4-year survival rates of 33% compared to just 13% for patients given the standard dose (Luskin *et al* 2016, Patel *et al* 2012, Sehgal *et al* 2015).

1.2.3.5 WT1

Wilms tumour 1 (*WT1*) is a transcription factor encoded for by a 10 exon gene on 11p13. The expressed protein contains a four zinc-finger DNA binding region at the C-terminus, and a transactivation domain at the N-terminus. Wild-type WT1 proteins bind to the Tet Methylcytosine Dioxygenase 2 (*TET2*) gene (Rampal *et al* 2014) and recruit it to its downstream targets. *TET2* acts as a tumour suppressor by inhibiting leukaemic cell proliferation and colony formation (Yiping Wang *et al* 2015). Mutations in *WT1* are present in 5-12% of AML cases, including those with cn-AML (Becker *et al* 2010, Gaidzik *et al* 2009, Hou *et al* 2010, Krauth *et al* 2015,

Paschka *et al* 2008, Renneville, Boissel, Zurawski, *et al* 2009, Summers *et al* 2007, Virappane *et al* 2008). They are present throughout the coding region of the gene, although the majority cluster in exons seven and 9, which form part of the DNA binding domain. The majority of mutations are frameshift, introducing early termination codons and subsequently the loss of the C-terminus (Summers *et al* 2007, Virappane *et al* 2008). Truncated *WT1* proteins are subject to nonsense-mediated decay (NMD), suggesting that haploinsufficiency, rather than loss-of-function is the mechanism at work (Abbas *et al* 2010), whereby a reduction in functional WT1 protein levels reduces the ability of *TET2* to inhibit the proliferation of AML cells. Mutations in WT1 are regarded as a marker of poor prognosis, with mutated patients less likely to achieve CR after the first cycle of induction chemotherapy (Renneville, Boissel, Zurawski, *et al* 2009). It is unsurprising, therefore, that this leads to both higher relapse rates and shorter OS than in patients with wild-type *WT1* (Becker *et al* 2010, Paschka *et al* 2008, Virappane *et al* 2008).

1.3 Current Sequencing Methods in Diagnostics

1.3.1 Sanger Sequencing

In 1977, improving on earlier work (Sanger and Coulson 1975), the Sanger group developed a method for determining the nucleotide structure of a single stranded DNA molecule (Sanger *et al* 1977a). This method, commonly and hereafter referred to as 'Sanger sequencing', works in a similar manner to a standard

polymerase chain reaction (PCR), in which a DNA polymerase, an oligonucleotide 'primer' and deoxynucleotidetriphosphates (dNTPs) are used to amplify a specific nucleotide sequence, using single stranded DNA as a template. The fundamental difference between this and Sanger sequencing, is the introduction of modified di-deoxynucleotidetriphosphates (ddNTPs) to the reaction mix, which terminate DNA strand extension upon incorporation. Four ddNTPs are used, one each for adenine, cytosine, guanine and thymine (ddATP, ddCTP, ddGTP and ddTTP respectively), in combination with unmodified dNTPs, usually at a 100-fold lower concentration. This imbalance ensures that, at each round of amplification, a small proportion of DNA molecules are prevented from extending by the ddNTPs, whilst the remainder continue to be amplified. This generates a collection of DNA fragments, or amplicons, each of which differs in length by a single base pair. The pattern formed by these fragments, when separated on a polyacrylamide gel, represents the order of nucleotides in the original DNA template (Figure 1.5). In the original paper Sanger *et al* (1977a) performed four separate sequencing reactions, using the same template DNA and a different 32P radiolabelled ddNTP in each, and proceeded to successfully sequencing the genome of the phiX174 bacteriophage (Sanger *et al* 1977b), generating sequencing reads of between 150-200bp with an error rate of one every 50 bases, in the first real example of whole genome sequencing. This method of DNA sequencing facilitated new studies in a range of fields, with almost 1300 studies utilising it over the course of the next eight years.

During this time, Sanger sequencing remained a costly and laborious process, relying as it did on four individual reactions and hand-cast polyacrylamide gels. An

improvement to the method, therefore, was the development of dye-terminator DNA sequencing (L M Smith *et al* 1986), alongside the first automated DNA sequencing platforms, in which each ddNTP is labelled with a different fluorescent dye rather than 32P. This allowed the sequencing reaction to take place in a single assay, with the fragments being passed through a system of capillaries and across a detector. Analysis software is then used to determine the nucleotide order based on the colour pattern of the fragments.

Ultimately, this both greatly reduced the price and increased the throughput and caused the method to become widely established in laboratories around the world and is regarded as the 'gold standard' method for mutation detection.

Improvements to the chemistry used for Sanger sequencing have also led to an increase in the read length, with 700-1000bp routinely obtainable, alongside a reduction in the inherent error rate. However, the sensitivity of the standard method is only around 20%, that is, mutations present at a level below that are unlikely to be picked up. This is not an issue for constitutional genetics, where variants are expected to be present at either 50 or 100%, but a cause for concern in the study of cancer, where somatic driver mutations may be present at much lower levels.

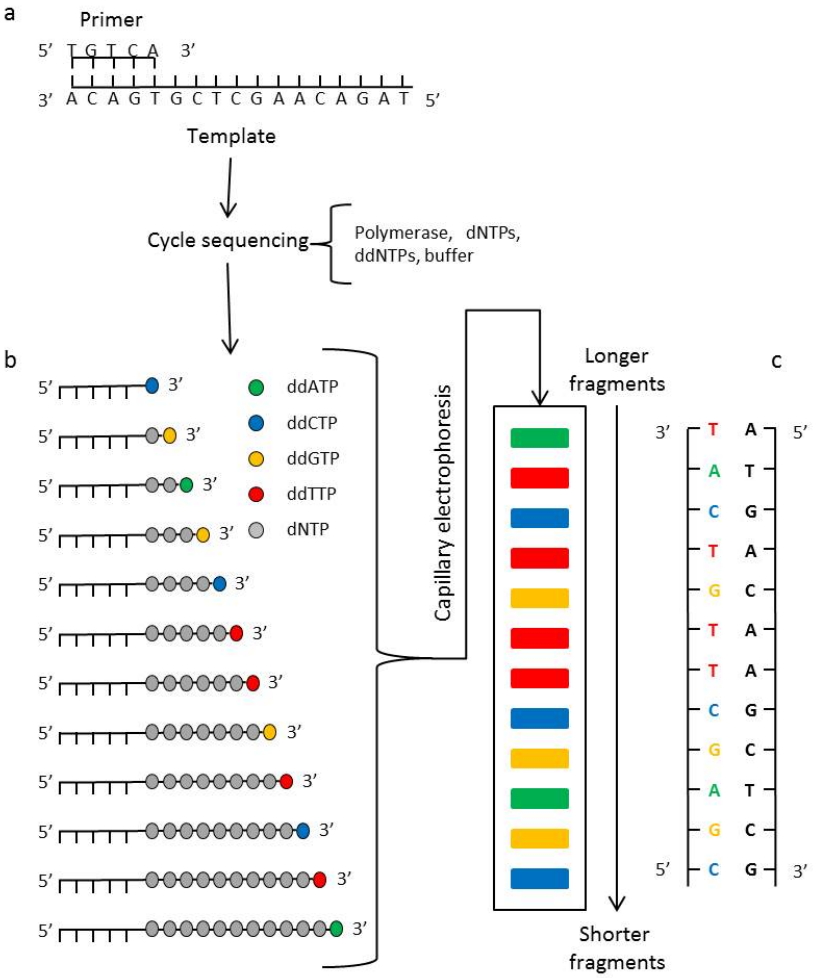

Figure 1.5 Outline of the Sanger sequencing method using fluorescent dye terminator chemistry.

(**a**) Template DNA is amplified in a cycle sequencing reaction using a combination of fluorescently labelled ddNTPS, unmodified dNTPs, Taq polymerase and a single sequencing primer. (**b**) The inclusion of modified ddNTPs in the reaction mix results in a collection of labelled PCR fragments of varying length. (**c**) These fragments are separated by capillary electrophoresis and passed between a laser and a detector, which determine the identity of each nucleotide in sequence based on the wavelength of light emitted by the ddNTP.

1.3.2 Pyrosequencing

First described in 1996 (Ronaghi *et al* 1996), pyrosequencing is a method of sequencing DNA templates using the 'sequencing-by-synthesis' principle, in which each base is identified during incorporation into the complementary strand. The region of interest is first amplified using a pair of specific primers, one of which is biotinylated, in a standard PCR reaction. The resulting amplicons are isolated using streptavidin beads and incubated with a sequencing primer, DNA polymerase, ATP sulfurylase, luciferase and apyrase enzymes, as well as the luciferin and adenosine 5' phosphosulfate (APS) substrates.

The assay works via the indirect detection of the DNA polymerase activity, during a sequencing reaction. When DNA polymerase incorporates a dNTP molecule, it also causes the release of pyrophosphate (PPI). During pyrosequencing, PPI is converted into adenosine triphosphate (ATP) by ATP sulfurylase in the presence of APS. Luciferase uses this ATP to oxidise the luciferin, producing oxyluciferin and releasing light (Figure 1.6). The light signal is detected using a charge-coupled device (CCD) camera, with the magnitude corresponding to the number of nucleotides incorporated during that sequencing cycle.

To identify which of the four possible nucleotides has been incorporated into the template, they are added to, and removed from, the reaction sequentially. This was achieved initially by washing the reaction mix between each nucleotide addition step. Ensuring that all previous nucleotides have been removed is critical to the accuracy of the assay. The later addition of an apyrase to the enzyme mix

(Ronaghi 1998) removed the need for these intermediate washes and paved the way for automated pyrosequencing systems.

The biggest advantage of this method is its ability to detect mutations at levels as low as 5% (Ogino *et al* 2005, Tsiatis *et al* 2010), compared to the 15-20% often quoted for Sanger sequencing. It is also able to generate accurate sequence data from the first base, unlike Sanger, which has a short 20-30bp interference window at the beginning of the read. There remain, however, a number of limitations to the method, not least of which is a restricted read length of around 100bp (Gharizadeh *et al* 2002). Additionally, the sensitivity of the CCD determines how well the system deals with homopolymer regions, as the strength of the light signal could overwhelm the detector. The quality of the sequence data often deteriorates downstream of these homopolymer regions.

Whilst these restrictions render this method unsuitable for de-novo sequencing, it is highly suited to more targeted approaches, where the mutation to be interrogated has been well defined. For example, colorectal cancer patients are often treated with *EGFR* inhibitors, such as Cetuximab, Panitumumab or Erlotinib.

However, mutations in the *BRAF* or *KRAS* genes render this therapy ineffective (Amado *et al* 2008, Karapetis *et al* 2008, Lièvre *et al* 2008). Pyrosequencing assays are commonly used in diagnostic laboratories to screen for these well characterised *BRAF* and *KRAS* mutations (Dufort *et al* 2009, Loupakis *et al* 2009, Ogino *et al* 2005, Packham *et al* 2009), with the information being used to help guide treatment decisions.

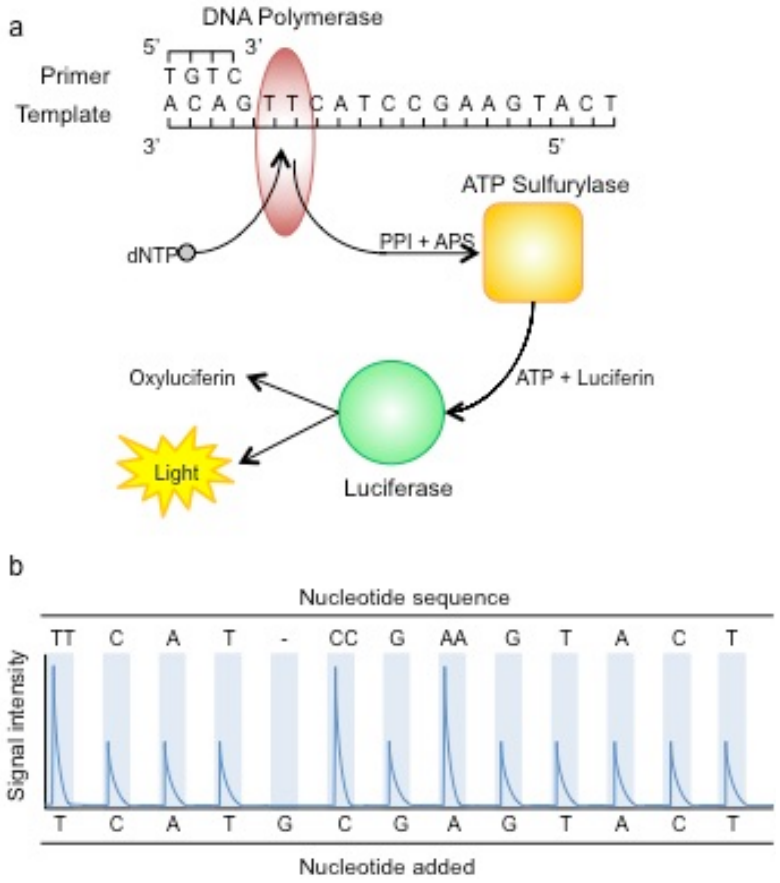

Figure 1.6 Outline of the pyrosequencing chemistry.

(**a**) Incorporation of a dNTP by DNA polymerase causes the release of pyrophosphate (PPI), which is converted into ATP by ATP sulfurylase, in the presence of APS. (**b**) Luciferase uses the ATP to oxidise luciferin, producing oxyluciferin and generating a visible light signal, the intensity of which is proportional to the number of bases incorporated. A CCD captures the light signal and converts it to a peak in a pyrogram.

1.3.3 Amplification Refractory Mutation System

Amplification Refractory Mutation System (ARMS) PCR is a robust and reliable method for the detection of single point mutations in DNA. The assay works on the principle that an oligonucleotide with a mismatched base at the 3' end will not function as a primer in a PCR reaction. Therefore, a primer that is complementary to a specific point mutation will only generate a PCR product in the presence of that mutation (Newton *et al* 1989).

Taking advantage of this, an ARMS-PCR assay will typically contain one common amplification primer, one primer whose 3' end is specific to the wild-type sequence and a third primer with a 3' modification specific to the targeted sequence. Generally, the PCR product generated by the wild-type/common primer pair will be significantly different in size than that produced by the mutant/common primer pair, in order to make visualisation of the products easier. In the presence of unmutated template DNA, the combination of these three primers will generate a single, wild-type PCR product. By contrast, in the presence of template DNA with a heterozygous mutation, the assay will generate both the wild-type amplicon and a shorter 'mutant' amplicon (Figure 1.7). The mutational status of the sample can therefore be determined by size-separating the PCR products using agarose gel electrophoresis. When fully optimised, these assays are able to detect the presence of mutated alleles present at levels as low as 0.5% (Huang *et al* 2013).

As reliable, sensitive and simple to use as these assays are, they rely on the investigator having prior knowledge of the mutation in question. As a result, they are widely used in cancer diagnostics where the variant has been well characterised and has prognostic value to the clinician; including *KRAS* mutations in colorectal cancer (Bolton *et al* 2015), *EGFR* Leu858Arg mutations in lung cancer (Morise *et al* 2014, Que *et al* 2016), *BRAF* Val600Glu mutations in melanoma and thyroid cancer (Huang *et al* 2013, Lu *et al* 2015), *JAK2* Val617Phe mutations in myeloid neoplasms (Kui *et al* 2013) and *NOTCH1* mutations in CLL (Bo *et al* 2014).

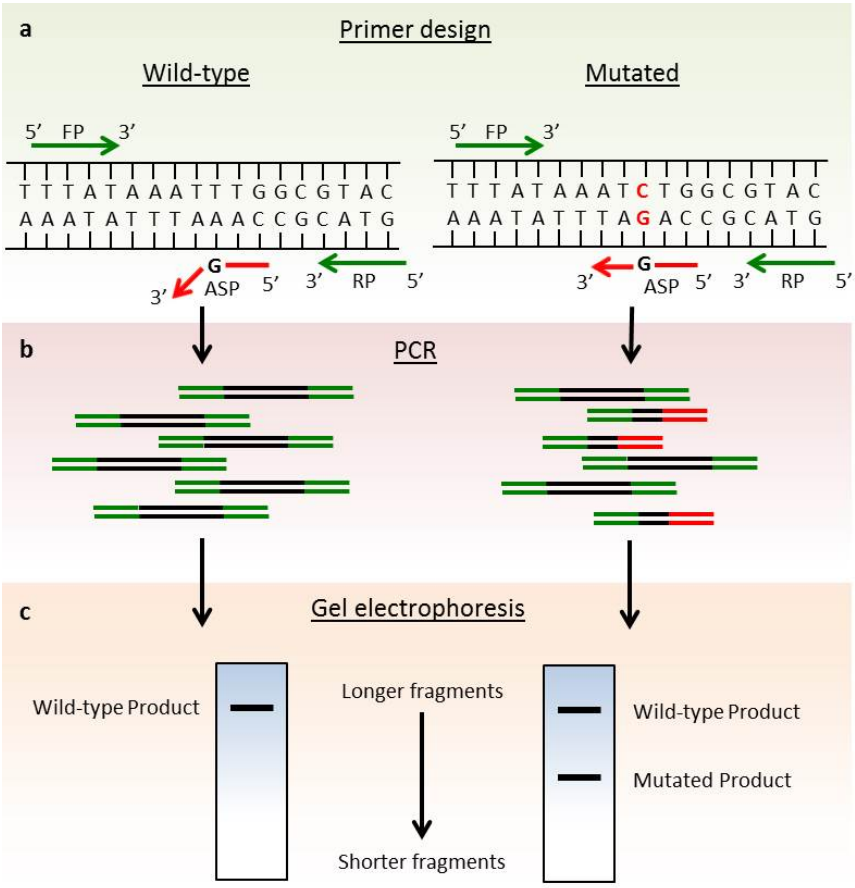

Figure 1.7 ARMS-PCR workflow.

(**a, b**) An allele specific primer, used in conjunction with wild-type forward and reverse primers, are used to generate amplicons by polymerase chain reaction. (c) These products can then be separated and visualised on an agarose gel. Abbreviations: FP, forward primer; RP, reverse primer; ASP, allele specific primer.

1.4 Development of 'Next-Generation' Sequencing

The publishing of the first drafts of the human genome in 2001 (McPherson *et al* 2001, Venter *et al* 2001) facilitated a greater understanding of our genetic profile. Three years later, the final draft was released (International Human Genome Sequencing Consortium, 2004), forming an important tool in molecular research and analysis to this day. This feat is all the more impressive considering the techniques available at the time. Both the International Human Genome Sequencing Consortium (2004) and the Venter group (2001) used dideoxy capillary sequencing methods, typically yielding 700-800 nucleotides of data per sequencing reaction. Putting that into context, the human genome is some three billion nucleotides in length.

In the years since a number of next-generation sequencing (NGS) methods have been developed. These systems offer a much greater level of sensitivity and throughput compared with the capillary methods that preceded them. They enabled researchers to sequence millions of DNA fragments simultaneously, with great sensitivity and accuracy.

1.4.1 Roche 454 Pyrosequencing

In 2005, the Margulies group described a method of generating >25 million bases worth of good quality data that utilised a combination of emulsion PCR and pyrosequencing (Margulies *et al* 2005). In contrast to other sequencing library preparation methods, in which the fragments are amplified in a single pooled

reaction, the 454 protocol clonally amplifies each fragment on the surface of DNA capture beads in an emulsion PCR reaction. Single stranded DNA molecules are generated from the genomic DNA source by either nebulisation or sonication. Following an end-repair reaction, adaptor sequences are ligated to each end of the fragments. These adaptors are complementary to the sequences found on the capture beads. The fragments are then diluted to a single molecule level and then mixed with the DNA capture beads in a manner that ensures that each capture beads contains only a single library fragment. Emulsion PCR, whereby each bead is isolated in its own reaction compartment, results in the amplification of the bound library fragment, with the bead surface containing $1x10^6$ clonal copies of the original fragment. The beads are then loaded onto a picotiter plate (PTP), the surface of which is covered with individual wells that are sufficiently small that only a single capture bead will fit into each one. The capture beads are packed into the wells alongside enzyme beads upon which the reaction enzymes (sulfurylase and luciferase) have been immobilised. The PTP functions as a flowcell, in that the four dNTPs are washed across it in sequence. Each nucleotide incorporation event is accompanied by the release of a chemi-luminescent light signal, which is detected and recorded by a CCD (Figure 1.8). On-instrument analysis software combines information regarding both the location and magnitude of the light signal, along with the identity of the dNTP distributed across the flowcell, to generate linear sequence data per well. Roche first marketed this technology in 2005 as the GS20 instrument, which has since evolved into the GSFLX+, with the

capacity to generate 700 megabases (Mb) of sequence data, with read lengths of 700bp, in just 23 hours.

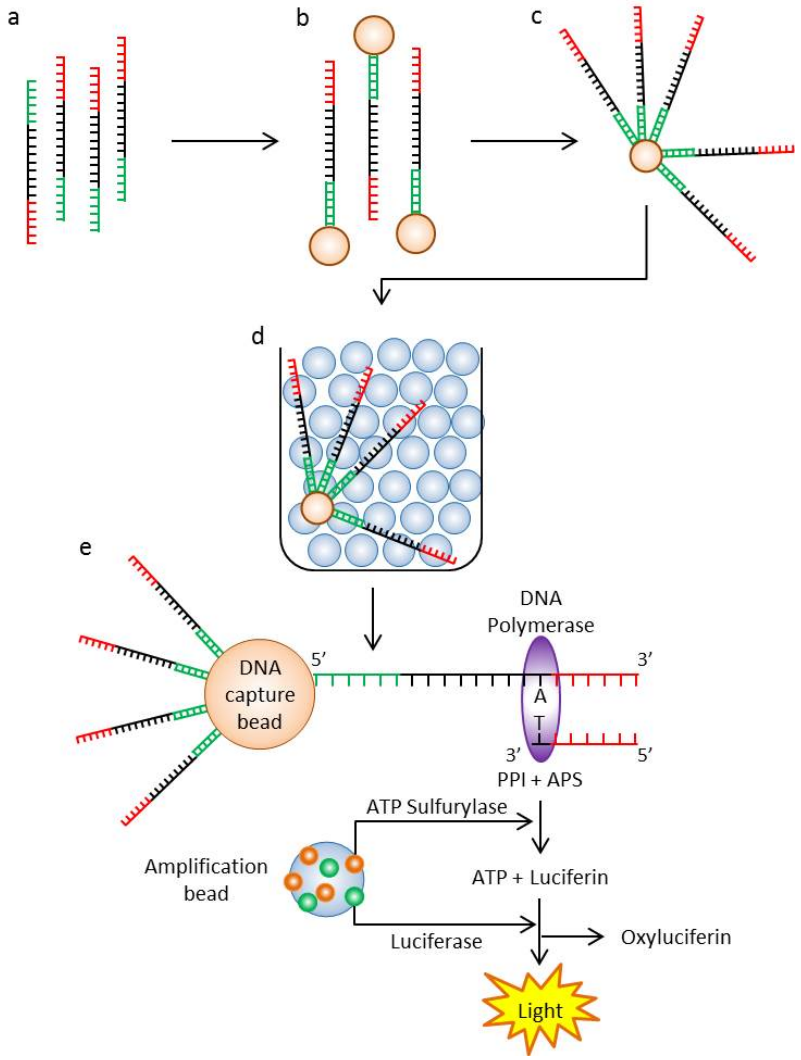

Figure 1.8 Roche 454 sequencing workflow.

(**a**) 454-specific adaptors are ligated to both the 5' and 3' ends of fragmented genomic DNA. (**b**) Single fragments are then ligated to DNA capture beads. (**c**) Emulsion PCR generates clonal copies of the bound fragment. (**d**) DNA capture beads are then loaded into individual wells of a PTP with (**e**) amplification beads covered in both sulfurylase (orange) and luciferase (green) enzymes, where the pyrosequencing reaction occurs. Adapted from Mardis *et al* (2008).

The increased read lengths afforded by the Roche chemistry, compared to the Illumina and Life Technologies platforms, make this method uniquely suited to certain applications, particularly de-novo sequencing projects in which no reference genome is available for post-sequencing mapping. However, this chemistry encounters some of the same issues as standard pyrosequencing assays. Sequencing errors are often introduced in regions containing large homopolymer domains, particularly those enriched for adenine and thymine bases (Luo *et al* 2012), as the proportional relationship between the signal intensity and the number of nucleotides incorporated becomes less reliable (Margulies *et al* 2005).

1.4.2 Illumina Sequencing-by-Synthesis

Bentley *et al* (2008) reported a method of using reversible terminator chemistry to produce between one and two billion bases of sequence data, a process known as 'sequencing-by-synthesis' (Figure 1.9). Template DNA is fragmented via sonication to produce a pool of fragments of between 350bp and 550bp in length. Following an end-repair step, Illumina specific adaptors, P5 and P7, are ligated to each end of the library fragments.

These adaptors are complementary to a lawn of oligonucleotides on the surface of the flow cell. The library fragments are then bound to the flow cell surface via an active heating and cooling step. The generation of clonal clusters is achieved by a process known as bridge amplification, in contrast to the emulsion PCR method of 454 sequencing. The immobilised library fragments are arched over and bound to

an adjacent complementary oligo. A series of amplification cycles generates up to 1,000 identical copies of each fragment in close proximity to each other, resulting in up to $50x10^6$ clonal clusters bound to the flow cell surface. Once loaded onto the sequencing instrument, these clusters are denatured and the fragments corresponding to the reverse strand are chemically cleaved away, leaving only the forward strand available for sequencing. Primers complementary to the P5 and P7 adaptor sequences are used to initiate the sequencing chemistry. This process utilises four different fluorescently labelled, reversible dye terminators that are concurrently washed across the flow cell and bind to the template DNA. In a manner reminiscent of Sanger sequencing, the modifications to the dye terminators allow for only a single base to be incorporated into each template molecule during each sequencing cycle. Once the next dye terminator has been incorporated into the template sequence, and the excess reagents have been washed away, the entire flow cell is imaged and the fluorescence is recorded. The next sequencing cycle begins with a chemical wash to first unblock the incorporated dye terminator, which facilitates the addition of the next base in the sequence.

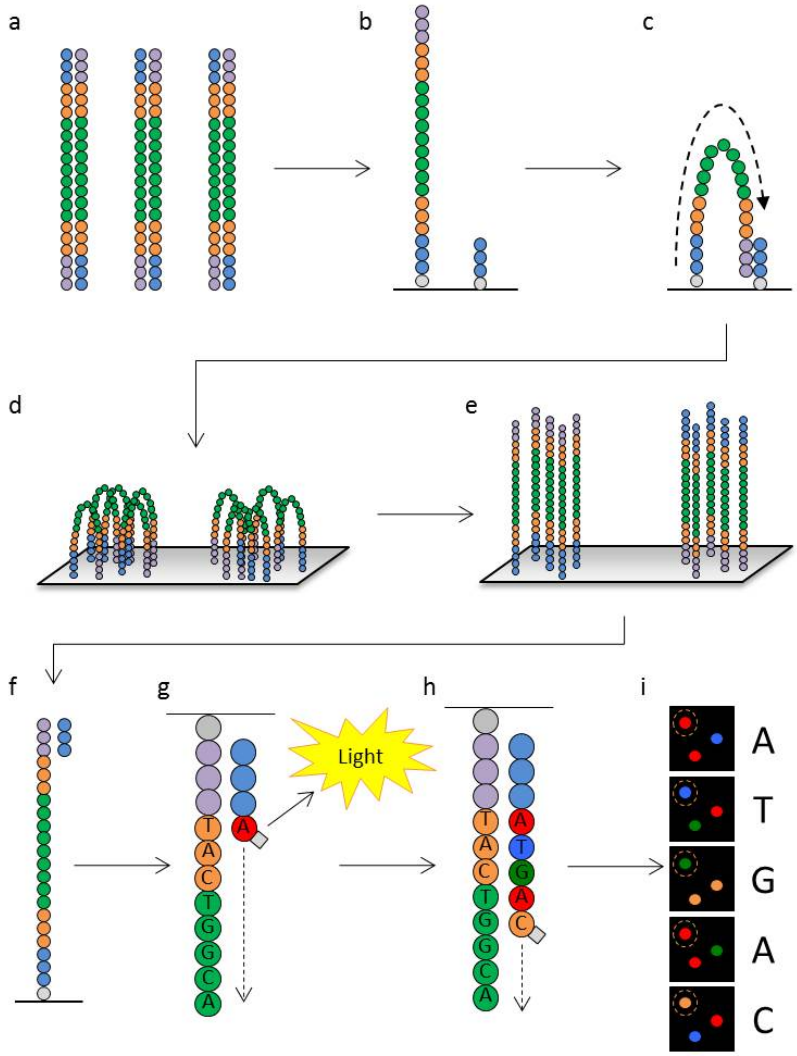

Figure 1.9 Illumina sequencing-by-synthesis workflow.

(**a**) Illumina specific adaptors (blue and purple) are ligated to fragmented genomic DNA, along with sample specific barcode sequences (orange). (**b**) Fragments are then annealed to the surface of the sequencing flowcell, and (**c**, **d**) amplified via a bridge amplification process. (**e**) The resulting clusters are then denatured to produce single stranded sequencing template. (**f**) A sequencing primer then binds to the adaptor sequence. (**g**) Reversible dye terminators are then washed across the surface of the flowcell, with each incorporation event releasing a chemi-luminescent signal. (**h**) The incorporated base is unblocked and the process is repeated along the length of the template. (**i**) A CCD records each signal, and the sequence of each cluster is determined based on the colour of the signal.

This continues in sequence until the full length of the forward read has been recorded. Employing the bridge amplification method a second time, which generates the reverse strand, and cleavage of the forward strand, enables paired-end sequencing.

By using this sequencing-by-synthesis method, the Illumina HiSeq 2500 instrument is able to generate 1000 gigabases (Gb) in a single sequencing run in less than a week. A single human genome, covered to a depth of at least 30x, is achievable in just 27 hours. In contrast to the Roche 454 platform, the read lengths achievable with the Illumina HiSeq are restricted to a maximum of 250bp. The recent development of the HiSeq 3000 and 4000 systems offers significant gains in capacity and accuracy over the HiSeq 2500. The introduction of a patterned flow cell, the surface of which is covered in nanowells at fixed locations, allows for uniformity in both the distribution and size of the clusters. This increases the accuracy of the sequencing and, therefore, improves the data output of the sequencer by reducing the number of reads discarded due to cross-talk between clusters. As a consequence of this, the HiSeq 4000 is able to generate up to 1500Gb of sequence data, and can sequence 12 human genomes in just four days.

Many research groups around the world have taken advantage of the quantity and quality of sequencing data obtainable through the Illumina platforms, in an effort to further characterise a number of cancers. These include both whole genome (Chapman *et al* 2011, Dulak *et al* 2013, Ho *et al* 2013, Link *et al* 2011, Morin *et al* 2013, Puente *et al* 2011, Rafnar *et al* 2011, Walter *et al* 2012, Wang *et al* 2014) and whole exome (Holmfeldt *et al* 2013, Landau *et al* 2013, Lawrence *et al* 2013,

Lohr *et al* 2012, 2014, Ojesina *et al* 2014) sequencing efforts that have helped to determine the complex molecular landscape of cancer by identifying new driver mutations, elucidating the impact of subclones on disease progression and discovering putative targets for future therapeutic regimens.

1.4.3 Emergence of Bench-top Instruments

Whilst the high-throughput capability of both the 454 FLX and HiSeq 2500 systems are well suited for exome and genome-wide explorative studies, for some particular applications there remains a need for a more focussed approach. For example, the use of targeted sequencing assays allows a researcher to isolate multiple specific regions of the genome for amplification and sequencing at once, returning only data that is relevant to the application in question. To address this need there has since been a push by Illumina, Roche and Life Technologies to produce smaller, more efficient and user-friendly 'bench-top' sequencers.

Illumina's solution is the MiSeq series of instruments. The MiSeq utilises the same sequencing-by-synthesis chemistry as the larger HiSeq packaged in a much more compact format. It is capable of generating up to 15Gb of sequence data from 2x300bp paired-end sequencing runs. Similarly, Roche has released the 454-junior instrument, a more compact version of its FLX+ sequencer. The 454-junior can generate 35Mb of data in 10 hours, with read lengths up to 400bp. Life Technologies developed the Ion Torrent personal genome machine (PGM), representing the first 'non-optical' semi-conductor sequencing system (Rothberg

et al 2011). In addition to a smaller physical footprint and data output, these instruments have much more automated and streamlined data analysis pathways. Software packages that are either on-instrument, in the case of the 454-Junior and Ion Torrent, or cloud based (MiSeq), analyse the data in real-time, aligning the sequence reads to a chosen genome to highlight any variants. These bench-top sequencers are ideal for targeted mutation analysis, whereby a number of candidate genes or mutational hotspots are simultaneously sequenced and analysed. This method of performing targeted sequencing is likely to be the primary application of NGS in clinical cancer diagnostics.

1.5 Third-Generation Sequencing Platforms

Despite the recent emergence and widespread acceptance of NGS platforms, a new generation of sequencing technologies is slowly emerging; so-called 'third-generation' sequencing. Exactly what distinguishes between a second and third-generation sequencing method remains somewhat ambiguous, but third-generation sequencing platforms are generally based upon the concept of single-molecule sequencing (SMS) without the need for template amplification. This reduces the likelihood of introducing errors into the template prior to sequencing, which can lead to false positive mutations.

In 2009, Pacific Biosciences described a sequencing method referred to as Single Molecule Real Time sequencing (SMRT) (Eid *et al* 2009). This chemistry utilises a flowcell covered in small reaction wells, referred to as zero-mode wave guides

(ZMWs), on the bottom of which a single DNA polymerase molecule has been immobilised. Each ZMW is illuminated from underneath, however the wavelength of the light is too large to pass fully through the ZMW, resulting in a detection volume of 20 zeptolitres (10^{-21} litres) at the bottom of the well. Incorporation of fluorescently labelled nucleotides releases a light signal, which is detected by a CCD and converted into a base call depending on the colour of the signal. The small reaction volume ensures that only incorporated nucleotides are detected, reducing the background noise level 1000-fold over other methods. This method has been utilised in a number of studies where long-range sequencing reads and high sensitivity have been crucial to the outcome, including in the detection of HIV-1 drug resistance mutations (Da Wei Huang *et al* 2016), non-invasive testing for colorectal cancer (CRC) (Russo *et al* 2015) and full-length MHC class I sequencing (Westbrook *et al* 2015).

One alternative method, currently in development by Oxford Nanopore Technologies, directly detects a single DNA molecule as it passes through a specifically bio-engineered nanopore. Each nanopore is based around a modified α-haemolysin ion channel, with both an exonuclease and a cyclodextrin sensor attached. These pores are embedded in a synthetic lipid bi-layer, which enables a current to be passed across it by altering the salt concentration on either side. Double-stranded DNA molecules are unwound at the entrance of the pore, with one strand then fed through by the current. During the transition through the pore, individual nucleotides are cleaved from the strand by the exonuclease. These cleaved nucleotides produce a disruption in the ionic flow through the pore,

altering the current in a manner proportional to the size, and therefore the identity, of the nucleotide in question. In this manner, the specific sequence of the DNA molecule can be inferred. A hairpin adaptor attached to the template DNA allows for 'paired-end' sequencing if required. This chemistry is expected to facilitate rapid sequencing of long DNA templates, without the need for error prone amplification, and the ability to handle complex or repetitive regions of the genome more accurately than earlier sequencing methods. Despite the early promise of nanopore sequencing, it is apparent that nanopore sequencing is still in its infancy, with studies highlighting high error rates and low alignment rates (Goodwin *et al* 2015, Mikheyev and Tin 2014). Efforts are now focussing on correcting these errors downstream of the sequencing process to generate more accurate data (Goodwin *et al* 2015, Madoui *et al* 2015).

1.6 Aims of this Thesis

The work presented in this thesis aims to demonstrate that next-generation sequencing platforms can play important roles, both diagnostically and on a research basis, in the context of haematological malignancies. These aims will be achieved by:

1. Designing and validating a targeted sequencing panel to screen for clinically relevant gene mutations in patients with known or suspected myeloid malignancies.

2. Demonstrating the clinical utility of the targeted sequencing panel by applying it to a cohort of clinical patient samples.

3. Describing the molecular landscape of the del(5q) MDS subtype using the targeted sequencing panel.

4. Performing whole genome sequencing of a cohort of CLL patients to fully characterise the molecular landscape, with particular reference to somatically acquired mutations, structural rearrangements and mutational signatures.

Chapter 2 MATERIALS AND METHODS

This chapter describes the protocols used in this work. Additional protocols, including primer sequences used for specific projects, are detailed in the relevant chapters.

2.1 Extraction of Genomic DNA

2.1.1 Extraction from Buffy Coat Samples

Manual extraction of genomic DNA from isolated Buffy coat samples was kindly performed by Maite Cabes of the Oxford Radcliffe Biobank, using phenol chloroform as follows: 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 five minutes. 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), and mixed by pipetting. The sample was incubated at 37^{o}C overnight. The addition of phenol to the reaction generates an emulsion that isolates the proteins in the mixture. The

phenol can then be separated via centrifugation and removed. 700µl of phenol (Sigma-Aldrich Corporation, St. Louis, MO, USA) was added to the solution, followed by centrifugation at 28,341g for five minutes. The aqueous upper layer was then transferred to a new 0.5ml Eppendorf tube (Sigma-Aldrich Corporation, St. Louis, MO, USA) also containing 700µl phenol and the centrifugation repeated. The two wash steps were repeated, replacing the phenol with 700µl of chloroform (Sigma-Aldrich Corporation, St. Louis, MO, USA). The DNA was extracted by adding the final upper aqueous layer to a new 0.5ml Eppendorf tube containing 800µl of chilled (4°C) 100% ethanol (VWR International, Radnor, PA, USA) and 1µl 7.5M ammonium acetate (Sigma-Aldrich Corporation, St. Louis, MO, USA). The mixture was centrifuged at 13,226g for five minutes to collect the pellet, and washed in 500µl 70% ethanol. The pellet was then air dried and re-suspended in nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA).

2.1.2 Extraction from Saliva Samples

Extraction of germline DNA from saliva samples was kindly performed by Maite Cabes of the Oxford Radcliffe Biobank using the Oragene DNA kits (DNA Genotek Inc, Ottawa, ON, Canada) as follows: each saliva sample was inverted 10 times to ensure adequate mixing and then incubated at 50°C for one hour. The sample was transferred to a 15ml Falcon tube (Thermo Fisher Scientific, Waltham, MA, USA) containing a $1/25^{th}$ 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 out impurities. The sample was then centrifuged at 3,500xg for 10 minutes. The supernatant was transferred to a new 15ml Falcon tube (Thermo Fisher Scientific, Waltham, MA, USA) and the pellet 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 then allowed to incubate at room temperature for 10 minutes, before centrifuging at 3,500xg for a further 10 minutes. The supernatant was removed and discarded and the pellet washed by adding 1ml of 70% ethanol (VWR International, Radnor, PA, USA) and incubating at room temperature for one minute. The supernatant was removed, taking care not to disturb the pelletised DNA. The DNA was re-hydrated by adding 0.5ml of TE buffer (10mM Tris-HCL, 1mM disodium EDTA, pH 8.0) (Sigma-Aldrich Corporation, St. Louis, MO, USA) followed by vortexing for 30 seconds. The DNA was then incubated at room temperature overnight, followed by a final 30 second vortex mix and then transferred to a labelled 1.5ml Eppendorf tube (Sigma-Aldrich Corporation, St. Louis, MO, USA).

2.2 Extraction of Total RNA

Total RNA was manually extracted from blood samples using the Qiagen RNeasy mini kit (Qiagen N.V. Hilden, Germany). The extraction procedure occurs in two parts, the first produces a guanidine-thiocyanate–containing (GTC) lysate, which can be stored at -80°C for up to six months. The second stage isolates the total

RNA. Whole blood was transferred to a 50ml Falcon tube (Thermo Fisher Scientific, Waltham, MA, USA), and the volume made up to 50ml with red blood cell (RBC) lysis buffer (Qiagen N.V. Hilden, Germany). Rotation of the tube for 10 minutes on a rotary mixer ensured efficient lysis of the sample. The lysed sample was centrifuged at 1509g for 10 minutes. The pellet was re-suspended in 20ml of RBC lysis buffer (Qiagen N.V., Hilden, Germany), mixed by rotation and centrifuged as above. The resulting cell pellet was re-suspended in the appropriate volume of RBC lysis buffer (Qiagen N.V., Hilden, Germany) in order that 1ml contained no more than $1x10^7$ cells. One millilitre of this suspension was transferred to a new tube and centrifuged at 13,226g for 2mins 30sec. The supernatant was removed and the pellet re-suspended in a solution containing 600µl RLT buffer (Qiagen N.V., Hilden, Germany) and 6µl β-mercaptoethanol/ml (Sigma-Aldrich Corporation, St. Louis, MO, USA). Buffer RLT enhances the binding of RNA to the membrane of the RNeasy spin column (Qiagen N.V., Hilden, Germany). The sample was homogenised by hand using a 20-gauge needle. The volume was doubled with 70% ethanol (VWR International, Radnor, PA, USA), and then washed through the RNeasy spin column, by centrifuging at 13,226g for 15 seconds, in two lots of 700µl. The sample was then washed first with 700µl of a stringent wash buffer (RW1) (Qiagen N.V., Hilden, Germany) and then twice with 500µl of a non-stringent wash buffer (RPE) (Qiagen N.V., Hilden, Germany), spinning as before. Once the membrane of the RNeasy spin column had air-dried, 30µl of nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA) was spun through the

column to elute the RNA into a 1.5ml Eppendorf tube (Thermo Fisher Scientific, Waltham, MA, USA).

2.3 Calculation of DNA Quality and Concentration

2.3.1 Nanodrop Spectrophotometry

The concentration and quality of DNA samples was determined using a Nanodrop 2000 UV-vis spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). This instrument generates data about both the concentration and purity of DNA samples in 15 seconds using only 1µl of undiluted material. In addition, the material used for measurement is recoverable. With the measurement arm raised, 1.5µl of test sample was pipetted onto the pedestal. Closing the measurement arm sandwiched the sample between the light source and the detector. The Nanodrop 2000 software package v1.6.19 (ThermoFisher Scientific, Waltham, MA, USA) was used to take the measurement and record the results for each sample. To determine a baseline measurement, 1µl of the solution used to re-suspend the DNA sample was used as a blank sample prior to the first test sample. The sample pedestal was wiped clean with distilled water between each test sample. DNA samples with a 260/280 ratio of ≥1.8, and RNA samples with 260/280 ratios ≥2.0 were considered pure enough for downstream analysis.

2.3.2 Qubit Fluorometric Quantitation

The Qubit fluorometric quantitation method utilises a fluorescent dye to quantify DNA samples. This approach is more accurate than spectrophotometric based methods, as it is able to distinguish between DNA, RNA and other contaminates, which may affect spectrophotometric readings. A Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) was used to determine the concentration of extracted DNA samples. A working dilution of Qubit solution was prepared by diluting the Qubit dsDNA reagent (Thermo Fisher Scientific, Waltham, MA, USA) by a factor of 1:200 in Qubit dsDNA buffer (Thermo Fisher Scientific, Waltham, MA, USA). For each of the two standards, 190µl of Qubit solution is required, along with 199µl for each test sample. Ten microliters of each standard was added to separate Eppendorf tubes (Thermo Fisher Scientific, Waltham, MA, USA) and mixed with 190µl of the Qubit working solution. For each test sample, 199µl of Qubit solution (Thermo Fisher Scientific, Waltham, MA, USA) was added to 1µl of DNA. All tubes were mixed by vortexing and allowed to stand at room temperature for two minutes. Measurements of the standards were taken first to establish a standard curve, followed by each sample in turn.

2.4 Calculation of RNA Quality and Concentration

2.4.1 Qubit RNA Quantitation

An estimate of the concentration of an RNA sample can be made using a Qubit RNA assay kit and the Qubit fluorometer. This assay is highly selective for single-

stranded RNA molecules over double-stranded DNA. Working dilutions of the RNA samples, standards and the Qubit reagent (Thermo Fisher Scientific, Waltham, MA, USA)were created as per the DNA assay (described in section 2.3.2). 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 solution (Thermo Fisher Scientific, Waltham, MA, USA) was added to 1µl of RNA. All tubes were mixed by vortexing and allowed to stand at room temperature for two minutes. Measurements of the standards were taken first to establish a standard curve, followed by each sample in turn.

2.4.2 Agilent Bioanalyzer

The quality, or integrity, of an RNA sample is an important factor in many experiments, particularly in next-generation sequencing where one of the first steps of library preparation involves degrading the sample. The Agilent bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) provides an accurate assessment of both the quality and quantity of an RNA sample prior to downstream processing. The integrity of the RNA is reflected in the RNA integrity number (RIN) assigned to each sample run, on a scale of one to ten. Lower RIN scores (<4) indicate a higher level of RNA degradation than higher ones (>8). For whole transcriptome sequencing, only samples with a RIN >7 were selected. For each sample, 1µl of RNA was loaded into a single well of an Agilent RNA 6000 Nano chip and processed using a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara,

CA, USA) as per the manufacturer's instructions (http://www.agilent.com/cs/library/usermanuals/Public/G2938-90034_RNA6000Nano_KG.pdf).

2.5 DNA Sequence Analysis

2.5.1 Polymerase Chain Reaction

A number of experiments in this study employed the polymerase chain reaction (PCR). Whilst the specific conditions were optimised based on the primer and target sequences of each assay, the reaction mix remained the same. Each 25µl reaction consisted of 100-250ng of genomic DNA, 12.5µl of 2x Qiagen Multiplex Mastermix (Qiagen N.V., 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. All primers were designed, and melting temperatures (Tm) were calculated, using Primer3 software (http://primer3.ut.ee). All primers were supplied by Sigma-Aldrich Corporation (St. Louis, MO, USA). 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) on a T3 Biometra thermocycler (Biometra GmbH, Göttingen, Germany). The standard cycling conditions were:

1 cycle of:	15 minutes at 97°C
30-40 cycles of	30 seconds at 92°C

	30 seconds at Tm°C
	20 seconds at 72°C
1 cycle of:	10 minutes at 72°C
1 cycle of:	Hold at 4°C

2.5.2 Agarose Gel Electrophoresis

PCR products were separated according to size using agarose gel electrophoresis, which uses an electrical field to move the negatively charged DNA towards a positive electrode through an agarose gel matrix. Smaller DNA fragments are able to move through the matrix more efficiently than larger fragments, and so will migrate further. 3% (w/v) gels were made by dissolving 6g of agarose (Sigma-Aldrich Corporation, St. Louis, MO, USA) in 200ml of 1x Tris-Borate-EDTA (TBE) (Geneflow Ltd, Lichfield, UK) with 0.5µg/ml ethidium bromide (Sigma-Aldrich Corporation, St. Louis, MO, USA). 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 run with the DNA marker appropriate for the fragment sizes in a 1x TBE running buffer (Geneflow Ltd, Lichfield, UK). A voltage of 120V was applied across the gel for 45 to 60 minutes. The gels were placed on a Bio-Rad 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).

2.5.3 Purification of PCR Product

Following PCR amplification, 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 28,341g for 40 minutes to collect the pellet, and then spun inverted for 10 seconds at a very low speed (1.6g) on a stack of paper towels to remove the supernatant. The pellet was re-suspended in 12µl of nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA) and allowed to stand for 10 minutes to re-hydrate.

2.5.4 Sanger Sequencing

DNA sequencing was based on the chain termination method first described by Prof. Sanger and colleagues (Sanger, Nicklen, *et al* 1977). The sequencing reactions were performed on an Applied Biosystems 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA). PCR reactions were performed as described in section 2.5.1.

2.5.4.1 Chain Terminator PCR

Each 10µl Sanger sequencing PCR reaction contained 1µl of purified PCR product (as described in section 2.5.3) and 9µl of reaction mix consisting of: 2µl 2.5x BigDye Terminator 3.1 ready reaction premix (Applied Biosystems, Foster City, CA,

USA), 1µl 5x BigDye sequencing buffer (Applied Biosystems, Foster City, CA, USA), 1µl of 3.2 microlmolar (µM) sequencing primer (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 5µl of nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA). Cycle sequencing reactions were performed in 0.2ml 96-well plates, under mineral oil, on a T3 Biometra thermocycler (Biometra GmbH, Göttingen, Germany). The cycling conditions were:

1 cycle of:	1 minutes at 96°C
25 cycles of	10 seconds at 96°C
	5 seconds at 50°C
	4 minutes at 60°C
1 cycle of:	Hold at 4°C

2.5.4.2 Ethanol EDTA Precipitation

5µl of 125mM EDTA pH 6.0 (Sigma-Aldrich Corporation, St. Louis, MO, USA) and 60µl of cold 100% ethanol (VWR International, Radnor, PA, USA) was added to each cycle sequencing reaction. The sample was inverted 10 times by hand to mix and then incubated at room temperature for 15 minutes. The plate was centrifuged at 2296g for 30 minutes at 4°C. The plate was then inverted to remove the supernatant and washed again in 60µl of ice-cold 70% ethanol (VWR International, Radnor, PA, USA). 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.

2.5.4.3 DNA Sequence Data Analysis

Sequencing analysis software (v5.4, Applied Biosystems, Foster City, CA, USA) was used to analyse the raw sequencing data in the first instance. This 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. The ABI file extension contains information on both the DNA sequence and the chromatogram for each sample. For the identification of variants in the DNA sample, Mutation Surveyor (SoftGenetics, State College, PA, USA) software was used. This software takes analysed ABI files and performs a direct base-to-base comparison against a FASTA sequence file of the wild-type sequence. A FASTA file is a simple text-based format for representing a specific nucleotide sequence.

2.5.5 DNA Fragment Analysis

The insertion or deletion of a short stretch of DNA can be detected effectively via a DNA fragment analysis assay. Primers are designed to amplify a region of DNA spanning the site of the suspected insertion or deletion. One of the two primers has a fluorescent dye, such as cyanine-5 (Cy5), conjugated to its 5’ end. When

excited by a laser, these dyes emit a light signal, allowing the PCR fragment to be compared to a size standard on an automated DNA sequencer. Samples containing a deletion or insertion will appear shorter or longer than the wild-type, respectively.

2.5.5.1 Detection of 4bp Insertion in *NPM1*

Exon 12 of the *NPM1* gene was amplified using primers (forward: CY5-ACT CTC TGG TGG TAG AAT GAA, reverse: GGC ATT TTG GAC AAC ACA TTC) previously designed and published by Scholl and colleagues (Scholl *et al* 2007). The forward primer had Cy5 dye conjugated to its 5' end. Extracted DNA, as described in section 2.1, was amplified using the following amplification conditions: 95°C for 15 minutes, followed by 40 cycles of 30 seconds at 92°C, 30 seconds at 58°C and 20 seconds at 72°C, with a final extension step of 10 minutes at 72°C. The resulting PCR product was diluted 1:10 in nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA). 2µl of this dilution was mixed with 40µl sample loading solution (Beckman Coulter, Brea, CA, USA) and 0.5µl GenomeLab DNA size standard 600 (Beckman Coulter, Brea, CA, USA) and subjected to capillary electrophoresis on a CEQ8000 Genetic Analysis System (Beckman Coulter, Brea, CA, USA). Data analysis was performed using CEQ analysis software version 9.0.25 (Beckman Coulter, Brea, CA, USA) using default settings.

2.5.5.2 Detection of Internal Tandem Duplications in *FLT3*

The fragment analysis protocol (section 2.5.5) was also used to screen for internal tandem duplications (ITDs) in the FMS-like tyrosine kinase-3 (*FLT3*) gene. The primer sequences were modified from those described by Murphy *et al* (2003) by conjugating WellRED fluorescent dyes (Sigma-Aldrich Corporation, St. Louis, MO, USA) to the 5' ends. The primer sequences were; forward: D4-GCA ATT TAG GTA TGA AAG CCA GC, reverse: D3-CTT TCA GCA TTT TGA CGG CAA CC, where D4 and D3 are WellRED dye codes. The standard PCR reagent mix was used (see section 2.5.1 above), with the following amplification conditions: 95°C for 15 minutes, followed by 35 cycles of 30 seconds at 95°C, one minute at 56°C and two minutes at 72°C, with a final extension step of 10 minutes at 72°C. The resulting PCR product was prepared and sequenced as per section 2.5.4.

2.5.6 Pyrosequencing Assay for JAK2^{V617F} Mutation

Pyrosequencing is a sequencing-by-synthesis assay by which short DNA molecules can be interrogated for the presence of small indels or single nucleotide polymorphisms (Ronaghi *et al* 1996). This method was used to detect the presence of the common Val617Phe point mutation in myeloid leukaemia cases. The Pyromark Assay Design software tool (v2.0, Qiagen, Hilden, Germany) was used to design a number of potential primer sequences, with the following found to be the most efficient; forward: biotin-GAA GCA GCA AGT ATG ATG AGC A, reverse: TGC TCT GAG AAA GGC ATT AGA A and the sequencing primer: TCT CGT CTC CAC AGA. The standard PCR amplification mix was used, as described in section 2.5.1 above,

with the forward and reverse primers, and with the addition of 2.5µl 5x Q-solution (Qiagen N.V., Hilden, Germany). Cycling conditions were 97°C for 15 minutes, followed by 35 cycles of 30 seconds at 97°C, 90 seconds at 62°C and two minutes at 72°C, with a final extension step of 10 minutes at 72°C. The pyrosequencing assay was performed on a Pyromark Q24 instrument (Qiagen N.V., Hilden, Germany) as per the manufacturer's protocol (https://www.qiagen.com/us/resources/resourcedetail?id=59f0275d-e60f-4517-b786-b0e0ca13952e&lang=en) The assay design software defined the dispensation order for the assay.

2.5.7 Real-time PCR Assay for JAK2^{V617F} Mutation

Real-time PCR is a method commonly used to quantitatively determine the proportion of single nucleotide polymorphisms in DNA samples through Taqman allelic discrimination (Livak 1999). In this assay, allele specific probes, labelled with 5' reporter and 3' quencher dyes, for both the wild-type and variant alleles, are incorporated into an otherwise standard PCR reaction. Incorporation of the probes into the PCR amplicon results in the cleavage of both the reporter and quencher dyes. This removes the suppression of the reporter signal by the quencher, and enables the detection of the reporter signal. The strength of the signal is proportional to the copy number of the target sequence. The Ipsogen *JAK2* MutaQuant kit (Qiagen N.V., Hilden, Germany) was used to determine the level of V617F *JAK2* mutations. PCR reactions were setup in a 100-well rotor by a CAS1200

liquid handling instrument (Qiagen N.V., Hilden, Germany). Each reaction mix contained 6.25µl 2x Taqman Universal PCR Master Mix (Applied Biosystems, Life Technologies, Carlsbad, CA), 0.5µl 25x primer/probe mix (Qiagen N.V., Hilden, Germany), 3.25µl nuclease-free water (ThermoFisher Scientific, Waltham, MA, USA) and 25ng of sample DNA. The samples were amplified on a Rotor-Gene 6000 instrument (Qiagen N.V., Hilden, Germany) using the following PCR conditions: 50°C for two minutes, 95°C for 10 minutes, followed by 50 cycles of 95°C for 15 seconds and 62°C for one minute, with signal acquisition in the FAM channel during the 62°C step. Each sample was tested for both wild-type and mutant alleles. Analysis of the raw data was performed using the Rotor-Gene Q software package (Qiagen N.V., Hilden, Germany). The cycle threshold was set at 0.03 with the slope corrected, as per the manufacturer's guidelines (https://www.qiagen.com/us/resources/resourcedetail?id=58d4a7d9-287f-4b01-85c3-5cb83db2228b&lang=en).

Raw data tables for both wild type and V617F assays were exported into Excel (Microsoft, Redmond, WA) to facilitate further analysis. The standard curves were plotted (y = mean ct, x = Log_{10} CN, where CN is gene copy number/5µl) for both the wild-type and V617F standard samples, and values for both the slope of the line (Y) and coefficient of determination (R^2) were extracted. The copy number for *JAK2*^{V617F} was calculated as:

$$\frac{mean\ Ct^{JAK2V617F} - standard\ curve\ intercept^{JAK2V617F}}{standard\ curve\ slope^{JAK2V617F}}$$

JAK2 wild-type copy number was calculated as:

$$\frac{mean\ ct^{JAK2WT} - standard\ curve\ intercept^{JAK2WT}}{standard\ curve\ slope^{JAK2WT}}$$

Final results were determined as a percentage of *JAK2*^{V617F} allele load, calculated by:

$$\frac{copy\ number^{JAK2V617F}}{copy\ number^{JAK2V617F} + copy\ number^{JAK2WT}}$$

2.5.8 Targeted Next-Generation Sequencing

Targeted next-generation sequencing provides a convenient method of screening for a number of mutations simultaneously in a time and cost effective manner, and is starting to contribute to the diagnosis and management in a range of diseases. The Illumina TruSeq Custom Amplicon assay (TSCA) (Illumina, San Diego, CA, USA) was used to amplify and sequence a large number of targets in multiple patients simultaneously (Figure 2.1). Dual-barcoded libraries were generated from 250ng of input DNA as per the manufacturer's protocol (https://support.illumina.com/content/dam/illumina-support/documentation/chemistry_documentation/samplepreps_truseq/truseqcustomamplicon/truseq-custom-amplicon-15-protocol-1000000005006-01.pdf). Up to 20 patient libraries were subjected to 2x150bp paired-end sequencing per sequencing run on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA). The Stampy and

Comment [a1]: Ref

Smith-Waterman (Smith and Waterman 1981) algorithms were used to align the sequence reads to the hg19 human genome build in parallel, with both Platypus and GATK used for variant calling (Figure 2.2). Additionally, Pindel was applied to the ISW-generated BAM files to screen for larger indel events. Annotation of each variant was carried out using VariantStudio software (Illumina, San Diego, CA, USA). The design of individual sequencing panels is discussed in depth in the relevant chapters.

Comment [a2]: Ref

Comment [a3]: Ref

Comment [a4]: Ref

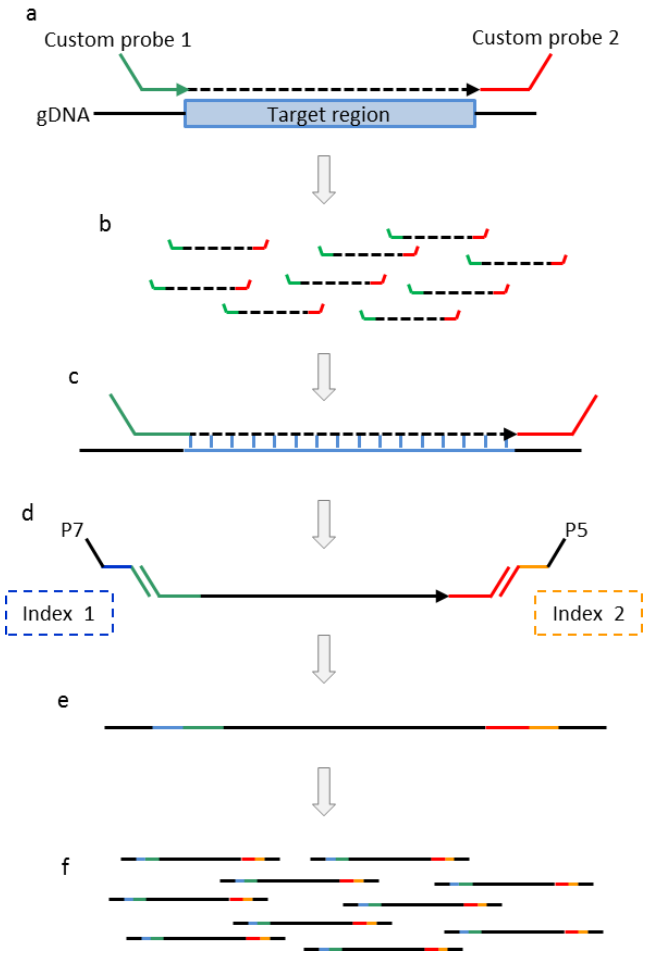

Figure 2.1 TruSeq Custom Amplicon Assay Workflow

(**a**, **b**) A series of probes, which are complementary to the DNA sequence immediately upstream and downstream of the target locus are designed and pooled together. (**c**) Genomic DNA is heat denatured and slowly cooled, allowing the probe pool to bind the specific targets. This is followed by an elongation and ligation process, to generate double-stranded target regions. (**d**) A polymerase chain reaction, using primers containing both sample-specific index sequences and the P5 and P7 sequences required for cluster generation, amplifies the target regions. (**e**, **f**) The resulting amplicon pool is then quantified and diluted, before loading onto a MiSeq instrument. DNA sequencing is then performed, as described in section 1.4.2. Abbreviations: gDNA, genomic DNA.

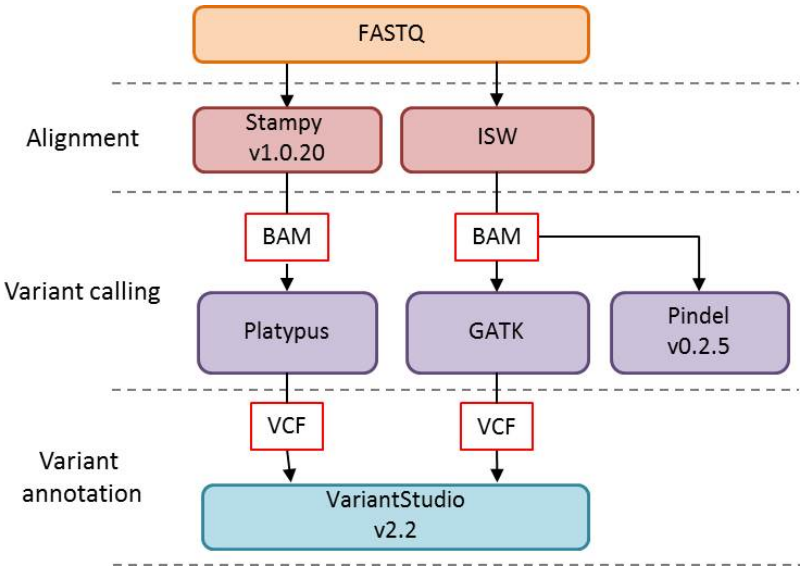

Figure 2.2 Analysis workflow for TruSeq custom amplicon sequencing data.

FASTQ files from the MiSeq instrument are aligned to the hg19 build of the human genome using both ISW and Stampy, simultaneously. The resulting BAM files are then passed to GATK and Platypus, respectively for calling of SNVs and small indels. Additionally, BAM files generated by ISW are used as input for Pindel, which detects larger indels than either Platypus or GATK. The VCF output from Platypus and Stampy are annotated using VariantStudio. Abbreviations; ISW, Illumina Smith-Waterman aligner; BAM, binary alignment/map file; VCF, variant call file; GATK, genome analysis toolkit.

Comment [a5]: Refs

2.5.9 Whole Genome Sequencing

Whole genome sequencing of the 42 CLL patients in chapter four was provided by Illumina Inc. (Little Chesterford, UK). Library preparation, DNA sequencing, data alignment and variant calling were performed by Illumina Inc. The resulting VCF and BAM files were returned to Oxford, whereupon I performed additional data analysis and annotation. Matched germline and tumour sequencing libraries were prepared from 2µg of input DNA for paired-end sequencing. DNA fragments were generated by random shearing using a Covaris M220 instrument (Covaris Inc, Woburn, MA, USA). Each library was prepared using the TruSeq DNA PCR Free

Library Preparation kit (Illumina, San Diego, CA, USA) as per manufacturer's protocol (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqdnapcrfree/truseq-dna-pcr-free-library-prep-protocol-guide-15075699-a.pdf). Flowcell clusters were generated using a cBot instrument (Illumina, San Diego, CA, USA) with the Illumina V3 Cluster Generation kit (Illumina, San Diego, CA, USA). The libraries were subjected to 2x35bp paired-end sequencing on a HiSeq 2000 instrument (Illumina, San Diego, CA, USA). Alignment of the sequencing reads to the hg19 human genome build was performed using iSAAC v01.13.04.04 (Raczy *et al* 2013). Single nucleotide variants (SNVs) and indels were called using Starling v2.0.3 and Strelka v2.0.10 (Saunders *et al* 2012) for germline and tumour reads respectively. 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 easily visualised using software such as the integrated genomics viewer (Broad Institute, Cambridge, MA, USA). Annotation of the variants was performed with VariantStudio v2.2 (Illumina, San Diego, CA, USA) (Figure 2.3). Variants were further filtered to omit any with a read depth <10x, or a quality score <30, from further analysis. Translocations were identified using the Delly program (Rausch *et al* 2012), with tumour and germline BAM files used as input. Copy number aberration analysis was performed using the Nexus 7 software (BioDiscovery Inc, El Segundo, CA, USA).

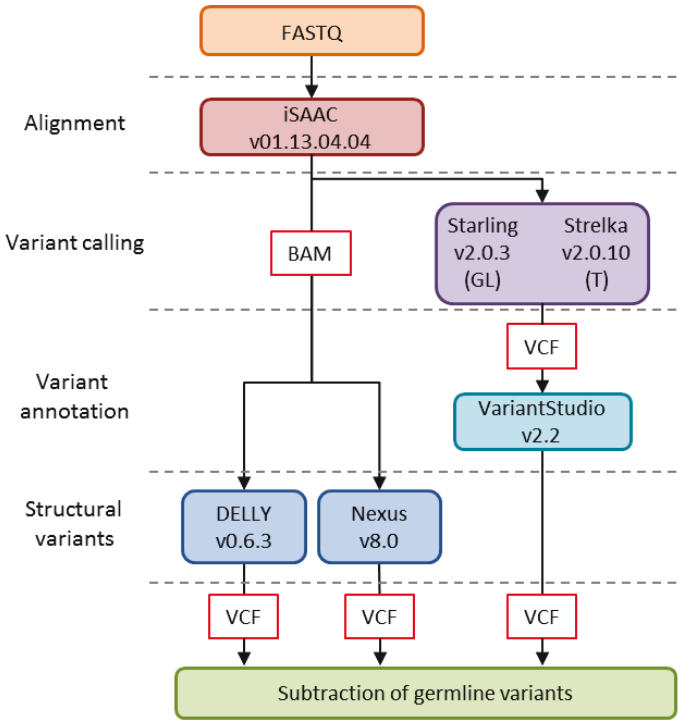

Figure 2.3 Analysis workflow for whole genome sequencing data.

FASTQ output files from the HiSeq 2500 are first aligned to the hg19 build of the human reference genome using the iSAAC aligner, followed by SNV and small indel detection using Starling and Strelka for germline and tumour samples, respectively. The resulting VCFs were then annotated using VariantStudio. The iSAAC aligner also generates BAM files, which were used as input for both DELLY and Nexus, to screen for translocations and copy-number aberrations, respectively. Subtraction of all variants present in germline DNA from the tumour DNA results in a final list of somatically acquired, tumour specific mutations and structural alterations. Abbreviations: GL, germline; T, tumour; VCF, variant call file; BAM, binary alignment/map file.

2.6 RNA Sequence Analysis

2.6.1 Whole Transcriptome Sequencing

Dr Basile Stamatopoulos kindly performed whole transcriptome sequencing on 32 CLL patients in chapter four. For the whole transcriptome sequencing of CLL patients, sequencing libraries were prepared using the TruSeq Stranded Total RNA

Sample Preparation kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqstrandedtotalrna/truseq-stranded-total-rna-sample-prep-guide-15031048-e.pdf). Libraries were pooled and subjected to 2x76bp paired-end sequencing on a HiSeq 2500 instrument (Illumina, San Diego, CA, USA) using the FASTQ only mode. Running the instrument in this mode means that no data analysis (either alignment to a reference genome or variant calling) is performed, rather, the output is a FASTQ formatted file containing raw DNA sequence. Sequence data was aligned to the hg19 build of the human genome using the TopHat algorithm (Trapnell *et al* 2009).

2.7 Statistical Analysis and Data Visualisation

Statistical analyses were performed using SPSS version 22 (IBM Corporation, Armonk, NY, USA) and Prism version 5 (GraphPad Software Inc, La Jolla, CA, USA). Categorical variables were compared using either the χ^2-test or Fishers exact test as appropriate. P values were considered significant at the 0.05 level. Circos plots were generated using the Circos software platform version 0.69 (Krzywinski *et al* 2009).

Chapter 3 DESIGN AND VALIDATION OF A TARGETED NEXT-GENERATION SEQUENCING PANEL FOR MYELOID MALIGNANCIES

3.1 Introduction

Acute myeloid leukaemia is a biologically complex disease, with treatment choices heavily influenced by both molecular and cytogenetic factors. In the first instance, AML cases can be classified into one of three risk groups on the basis of chromosomal abnormalities, as described in section 1.2.1. Despite the clear clinical utility of cytogenetic information in risk stratification in AML, around 40% of patients present with a normal karyotype (Grimwade *et al* 2010) and thus fall into a heterogeneous intermediate risk category.

A number of molecular abnormalities have been shown to have prognostic importance in AML, as described in section 1.2.3. In particular, the presence of an internal tandem duplication in the *FLT3* gene (*FLT3-ITD*) confers a poor prognosis (Fröhling *et al* 2002, Kottaridis *et al* 2001, Christian Thiede *et al* 2002, Whitman *et al* 2001), and can shift a patient with otherwise normal cytogenetics from the intermediate to the high risk group. Mutations in *NPM1*, on the other hand, confer a better prognosis, with increased OS and response to therapy in the absence of *FLT3*-ITDs (Döhner *et al* 2005, Hollink *et al* 2009, Schlenk *et al* 2008, Schnittger *et al* 2005, Thiede *et al* 2006). Other recurrently mutated genes identified in AML

include *ASXL1, CEBPα, DNMT3a, IDH1* and *IDH2*, although the full prognostic value of these remains to be fully determined (Table 3.1).

Traditional methods of screening for these mutations, including Sanger sequencing, pyrosequencing and allele-specific PCR (see section 1.3), are widely used in diagnosis, treatment selection, response monitoring and relapse prediction in laboratories around the world. Whilst all these methods are proven and reliable, their limitations lie in the fact that they are targeted at only the most prevalent mutations in a particular cancer. The ability to investigate as many candidate genes involved as possible in a particular patient's cancer, to a high level of accuracy, and to monitor these changes over time, will be a valuable tool in cancer diagnostics.

Table 3.1 Prevalence and prognostic impact of recurrently mutated genes in AML

Gene	Mutation	Prevalence	Prognostic Value	Reference
FLT3	Variable length ITD	20-30%	Shorter overall and disease free survival	(Fröhling *et al* 2002, Kottaridis *et al* 2001, Christian Thiede *et al* 2002, Whitman *et al* 2001)
NPM1	4bp insertion in exon 12	48-64% in normal karyotype, 8% with additional cytogenetic abnormalities	Increased overall survival and reduced relapse risk in absence of FLT3-ITD	(K Döhner *et al* 2005, Gale *et al* 2008, Schnittger *et al* 2005, Christian Thiede *et al* 2006, Verhaak *et al* 2005)
ASXL1	Frameshift or nonsense variants in exon 12	17%	Shorter event-free survival and shorter overall survival	(Schnittger *et al* 2013)
CEBPα	Frameshift or nonsense variants in exon 1	7-9%	Longer relapse free survival and overall survival in absence of FLT3-ITD	(Dufour *et al* 2010, Helbig *et al* 2014, Pabst *et al* 2001, Renneville *et al* 2009, Snaddon *et al* 2003)
DNMT3a	Missense mutations affecting arginine 882	18-30%	Shorter overall survival and increased risk of relapse	(Hou *et al* 2012, Ley *et al* 2010, Marková *et al* 2012, Thol *et al* 2011)
IDH1	Missense mutations affecting arginine 132	5-6%	Unknown	(Chou *et al* 2010, Schnittger *et al* 2010)
IDH2	Missense mutations affecting arginine residues 140 and 172	12%	Unknown	(Thol *et al* 2010)

3.2 Aims and Objectives

The aim of this work is to design and validate a multi-gene next-generation sequencing panel to be used in the diagnostic screening of patients with known or suspected acute myeloid leukemia, myelodysplastic syndromes or myeloproliferative neoplasm.

3.3 Materials and Methods

3.3.1 TruSeq Custom Amplicon Panel Design

A TruSeq Custom Amplicon (TSCA) sequencing panel (Illumina, San Diego, CA, USA) was designed to target the exons of 25 genes (Figure 3.1, Table 3.2). The panel was defined using the online DesignStudio pipeline (http://designstudio.illumina.com). In genes with well-defined mutational hotspots, only these regions were targeted; otherwise the entire coding sequence of the gene was sequenced. Each target region was designed to cover both the exon in question, and included a 10bp margin on either side. This was designed to detect any mutations affecting splice sites within genes. It was also designed to avoid regions containing known polymorphisms, the presence of which can affect the ability of the probes to bind to the target sequence. The final design of the assay covered 46,604bp of genomic sequence (96% of the desired region), spread across 322 amplicons. The missing regions included exon three of *NRAS,* exons nine and 10 of *EZH2* and exon 31 of *ATRX*, a total of 1,844bp. Subsequent updates of this panel, with improved chemistry and synthesis methods, now include all of these regions.

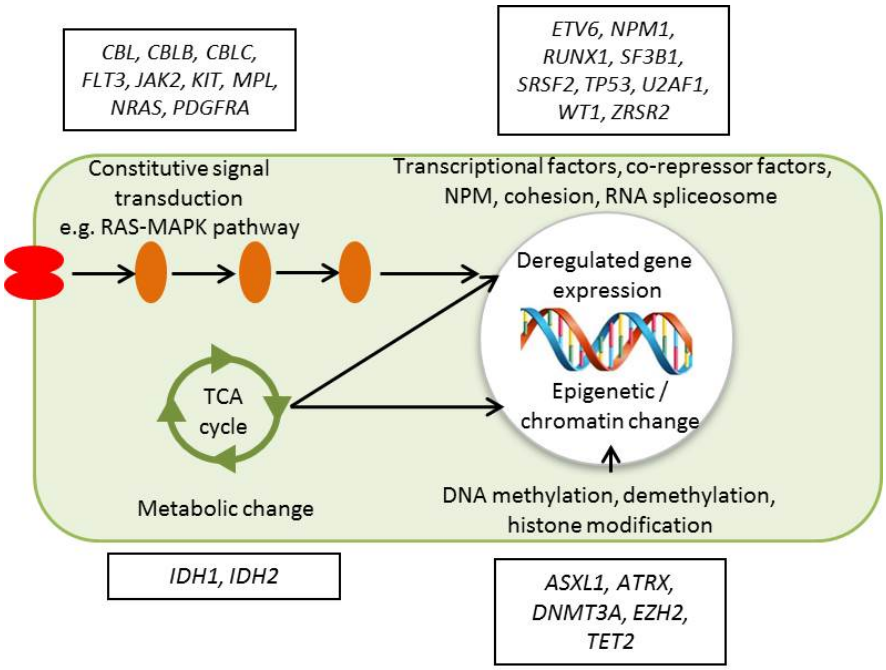

Figure 3.1 Overview of the genes targeted by the TSCA panel.

The genes targeted by the panel play roles in one of four biological processes; signalling pathways (*JAK2, FLT3, KIT*), metabolic pathways (*IDH1, IDH2*), DNA modification (*ASXL1, DNMT3A*) or gene expression (*NPM1*). Target genes were selected based on their presence in the literature as being prognostic markers, or as having implications in the therapy selection in AML, MDS or MPD.

Table 3.2 Details of genomic regions targeted by the myeloid TSCA panel.

Gene	Location	Chromosomal Coordinates (hg19)	Targeted Exons	Malignancy (% recurrence)	References
ASXL1	20q11.21	30,946,147-31,027,122	12	MDS (11-21%) AML (10%)	(Abdel-Wahab *et al* 2011, Boultwood *et al* 2010, Gelsi-Boyer *et al* 2009, Metzeler *et al* 2011)
ATRX	Xq21.1	76,760,356-77,041,719	8 to 10 and 17 to 31	ATMDS	(Gibbons *et al* 2008, Herbaux *et al* 2015, Steensma *et al* 2004)
CBL	11q23.3	119,076,986-119,178,859	8 and 9	MDS (5%)	(Grand *et al* 2009)
CBLB	3q13.11	105,377,109-105,587,887	9 and 10	MDS (1%)	(Aranaz *et al* 2012)
CBLC	19q13.32	45,281,126-45,303,903	9 and 10	MDS (1%)	(Aranaz *et al* 2012)
DNMT3A	2p23.3	25,455,830-25,564,784	23	MDS (8%) AML (4-20%)	(Ley *et al* 2010, M J Walter *et al* 2011, Yamashita *et al* 2010)
ETV6	12p13.2	11,802,788-12,048,325	All eight exons		
EZH2	7q36.1	148,504,464-148,581,441	2 to 20	MDS (7-12%)	(Abdel-Wahab *et al* 2011, Ernst *et al* 2010)
FLT3	13q12.2	28,577,411-28,674,729	14, 15 and 20	AML (25-37%)	(Patel *et al* 2012, Stirewalt and Radich 2003)
IDH1	2q34	209,100,953-209,119,806	4	AML (8%)	(Mardis *et al* 2009, Patel *et al* 2012)
IDH2	15q26.1	90,627,212-90,645,708	4	AML (8%)	(Patel *et al* 2012)
JAK2	9p24.1	4,985,245-5,128,183	12 and 14	MPN (48-80%)	(Baxter *et al* 2005, James *et al* 2005, Kralovics *et al* 2005, Levine *et al* 2005)

KIT	4q12	55,524,095-55,606,881	2, 8-11, 13, 17	AML (24%)	(Klein *et al* 2015)
MPL	1p34.2	43,803,475-43,820,135	10	MPN (8.5%)	(Beer *et al* 2008)
NPM1	5q35.1	170,814,708-170,837,888	12	AML (24-29%)	(Mardis *et al* 2009, Patel *et al* 2012)
NRAS	1p13.2	115,247,085-115,259,515	2 and 3	AML (9%)	(Mardis *et al* 2009)
PDGFRA	4q12	55,095,264-55,164,412	12, 14 and 18		
RUNX1	21q22.12	36,193,574-36,260,987	3 to8	MDS (20%)	(Harada and Harada 2009)
SF3B1	2q33.1	198,256,698-19,8299,771	15 and 16	MDS (13-20%)	(Papaemmanuil *et al* 2011, Yoshida *et al* 2011)
SRSF2	17q25.1	74,730,197-74,733,493	1	MDS (10%)	(Yoshida *et al* 2011)
TET2	4q24	106,067,842-106,200,960	3 to 11	MDS (19-26%)	(Delhommeau *et al* 2009, Langemeijer *et al* 2009)
TP53	17p13.1	7,571,720-7,590,868	4 to 9	MDS (5-10%)	(Pellagatti and Boultwood 2015)
U2AF1	21q22.3	44,513,066-44,527,688	2 and 6	MDS (9%)	(Graubert *et al* 2011)
WT1	11p13	32,409,322-32,457,081	7 and 9	AML (11%)	(Paschka *et al* 2008)
ZRSR2	Xp22.2	15,808,574-15,841,382	All 11 exons	MDS (4%)	(Yoshida *et al* 2011)

Most target genes contained well-defined mutational hotspots, and so only these regions were targeted by the panel. In contrast, for genes such as *ETV6* and *ZRSR2* in which there are no common mutation hotspots, the complete coding sequence was sequenced. All co-ordinates refer to hg19 genome build.

3.3.2 Patient Samples

3.3.2.1 Validation Cohort

DNA samples from nine patients with confirmed MDS were selected for the validation cohort. All samples had been previously screened for common MDS mutations at diagnosis by the Oxford University Hospitals Trust Molecular Diagnostics unit using either Sanger sequencing (as described in section 2.5.4, pyrosequencing as described in section 2.5.6 or DNA fragment analysis as described in section 2.5.5. The nine samples harboured a total of 17 variants across 11 genes (*ASXL1, DNMT3a, EZH2, FLT3, IDH1, IDH2, KIT, NPM1, NRAS, RUNX1* and *TP53*) and included missense (n=5), nonsense (n=3) and insertion/deletions (n=9).

3.3.2.2 Test Cohort

DNA samples from 43 untreated MDS cases, each harbouring a 5q deletion, were selected for mutational screening from collaborating institutes (St Johannes Hospital, Duisburg, Germany; Heinrich Heine University, Düsseldorf, Germany; University of Navarra, Pamplona, Spain; and Royal Bournemouth Hospital, UK). The ethics committees of all institutes involved approved the work, and informed consent was obtained from all patients, in line with the declaration of Helsinki. These included 22 patients classified as 5q- syndromes, nine with refractory anaemia (RA) and 12 with

advanced MDS (defined as having bone marrow with >5% blast cell content).

3.4 Results

3.4.1 Targeted Sequencing Panel Performance

3.4.1.1 Mutations Identified in the Validation Cohort

Dual-barcoded sequencing libraries were generated, sequenced and analysed according to section 2.5.8. Five missense, three nonsense and seven frameshift mutations were identified in the validation cohort (15 out of 17, 88.2%). Of note, the short frameshift mutations in both *ASXL1* and *NPM1* (1bp and 5bp respectively) were correctly detected (Table 3.3). All mutations were identified by GATK through the BaseSpace pipeline, with the exception of the *TP53* deletion in Test009, which was generated by Platypus, and the *FLT3*-ITDs, which were detected by Pindel.

Analysis of the aligned sequencing reads using the Integrated Genomic Viewer (IGV, Broad Institute) for one of the validation samples (Test009) revealed a dramatically reduced read depth of 30x across *TP53*, compared to >1000x within this sample at other positions and across the other samples. This suggested an inability to align the *TP53* sequence reads to the reference genome, which can be indicative of a deletion in the region.

Table 3.3 Summary of mutations present in TSCA panel validation samples.

Sample ID	Gene	Mutation	Coding change	Sequencing Depth	VAF
TEST001	*ASXL1*	c.1925het_insA	p.Gly643ArgfsX13	239	0.6
TEST001	*EZH2*	c.290dup	p.Leu98IlefsX28	895	0.32
TEST001	*EZH2*	c.748C>CT	p.Gln250X	437	0.42
TEST001	*NRAS*	c.35G>GA	p.Gly12Asp	955	0.32
TEST001	*RUNX1*	c.422C>CA	p.Ser141X	361	0.31
TEST002	*ASXL1*	c.1748G>GA	p.Trp583X	159	0.28
TEST003	*IDH2*	c.13775G>GA	p.Arg140Gln	1382	0.51
TEST003	*NPM1*	c.860_863dup	p.Trp288CysfsX12	836	0.23
TEST003	*FLT3*	79bp insertion	-	756	0.42
TEST004	*NPM1*	c.860_863dup	p.Trp288CysfsX12	734	0.25
TEST004	*FLT3*	64bp insertion	-	1294	0.12
TEST005	*NPM1*	c.860_863dup	p.Trp288CysfsX12	527	0.15
TEST006	*DNMT3a*	c.2844G>GA	p.Arg882Val	250	0.44
TEST007	*KIT*	c.2447A>AG	p.Asp816Val	248	0.39
TEST008	*NPM1*	c.860_863dup	p.Trp288CysfsX12	875	0.2
TEST008	*IDH1*	c.6694C>CT	p.Arg132Cys	1390	0.48
TEST009	*TP53*	c.489_471del	p.Lys164AlafsX12	731	0.95

Sequencing depth = number of sequencing reads covering the mutation. VAF = variant allele frequency

The sequence data for Test009 was subsequently submitted to the Stampy (Lunter and Goodson 2011) and Platypus (Rimmer *et al* 2014) analysis pipeline, which revealed a 19bp deletion in exon five of *TP53*. The initial read alignment and variant calling failed to align any reads containing deletions to the reference genome, resulting in a much lower read depth across this locus (~30x). By comparison, re-analysis of the same data using

the Stampy and Platypus pipeline resulted in a greater number of aligned reads, giving a higher read depth (>700x) and successful identification of the 19bp deletion (Figure 3.2).

The initial data analysis also missed the two internal tandem duplications in *FLT3* in samples Test003 and Test004, again likely due to the alignment algorithms inability to successfully map the reads containing the additional sequence. Another alignment and variant calling algorithm, Pindel (Ye *et al* 2009), was used to detect these insertions.

In addition to the known control mutations, six mutations affecting five genes in five samples were found (Table 3.4). One of the mutations, a *TET2* Cys1464X nonsense variant in Test001, was visible in earlier Sanger sequencing traces; however at the time, the variant had not been picked up by the analysis software due to the high level of background noise. The Cys1464X variant was detected and called from the MiSeq data with a VAF of 0.47 (200/423 reads). The variant can be seen in the Sanger sequencing trace (Figure 3.3), but was not identified by the Mutation Surveyor software (SoftGenetics, State College, PA, USA). All remaining additional mutations were confirmed by Sanger sequencing and fragment analysis.

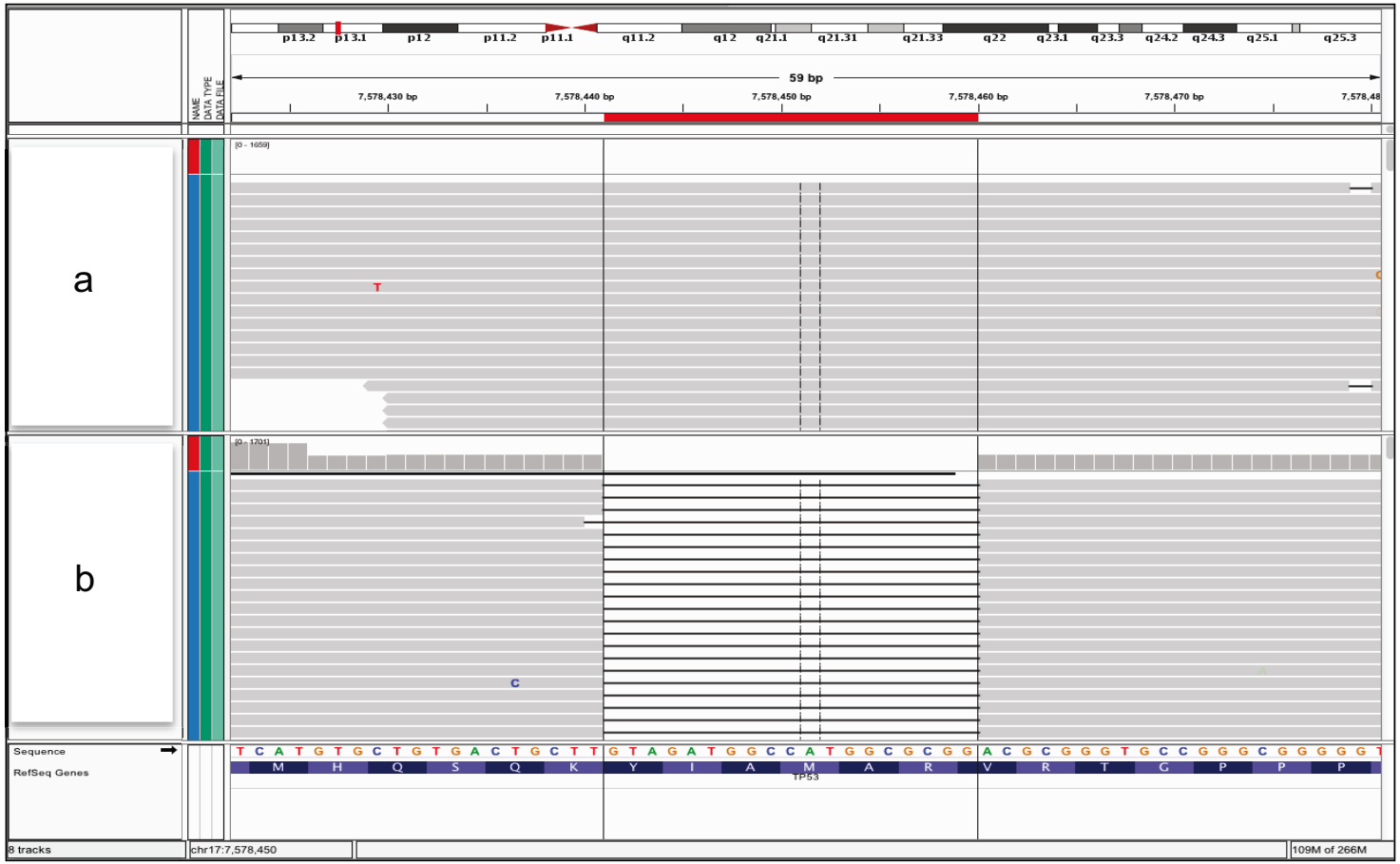

Figure 3.2 Comparison of NGS read alignments covering across the 19bp TP53 deletion in Test009

(**a**) The inability of the BWA aligner to map reads containing the deletion is represented by a lower read depth, (**b**) compared to the Stampy and Platypus pipeline results (**b**). Individual sequencing reads are represented by single horizontal grey bars, with the total read depth represented by vertical grey bars above each base.

Table 3.4 Summary of the additional mutations detected in the validation cohort

Sample ID	Gene	Mutation	Coding change	Sequencing depth	VAF
TEST001	*TET2*	c.4412C>CA	p.Cys1464X	423	0.47
TEST005	*TET2*	c.3767_3771dup	p.Leu1258AlafsX10	2968	0.25
TEST006	*NPM1*	c.860_863dup	p.Trp288CysfsX12	999	0.18
TEST007	*RUNX1*	c.291delC	p.Leu71SerfsX24	395	0.51
TEST007	*SF3B1*	c.2098T>TC	p.Lys700Glu	2053	0.45
TEST008	*FLT3*	c.2508_2510del	p.Ile836del	2271	0.48

Sequencing depth = number of sequencing reads covering the mutation. VAF = variant allele frequency.

3.4.1.2 Assay Limit of Detection

To determine the limits of detection of the assay, the variant allele frequencies (VAF) for mutations in *JAK2* from the targeted sequencing panel were compared with those obtained from a quantitative real-time PCR assay. For the test cohort, seven samples harbouring a previously identified *JAK2* Val617Phe (*JAK2*^{V617F}) mutation were selected, with a VAF ranging from one to 24%.

Real-time PCR assays were performed using the commercially available *JAK2* Muta*Quant*^{TM} kit (Qiagen N.V., Hilden, Germany), which distinguishes between *JAK2* wild type and the *JAK2*^{V617F} alleles through Taqman allelic discrimination. Allele specific probes, labelled with 5' reporter and 3' quencher dyes, for both wild-type and *JAK2*^{V617F} alleles are used to amplify the region of interest. The *JAK2*^{V617F} VAF can be calculated from the fluorescent levels of each assay.

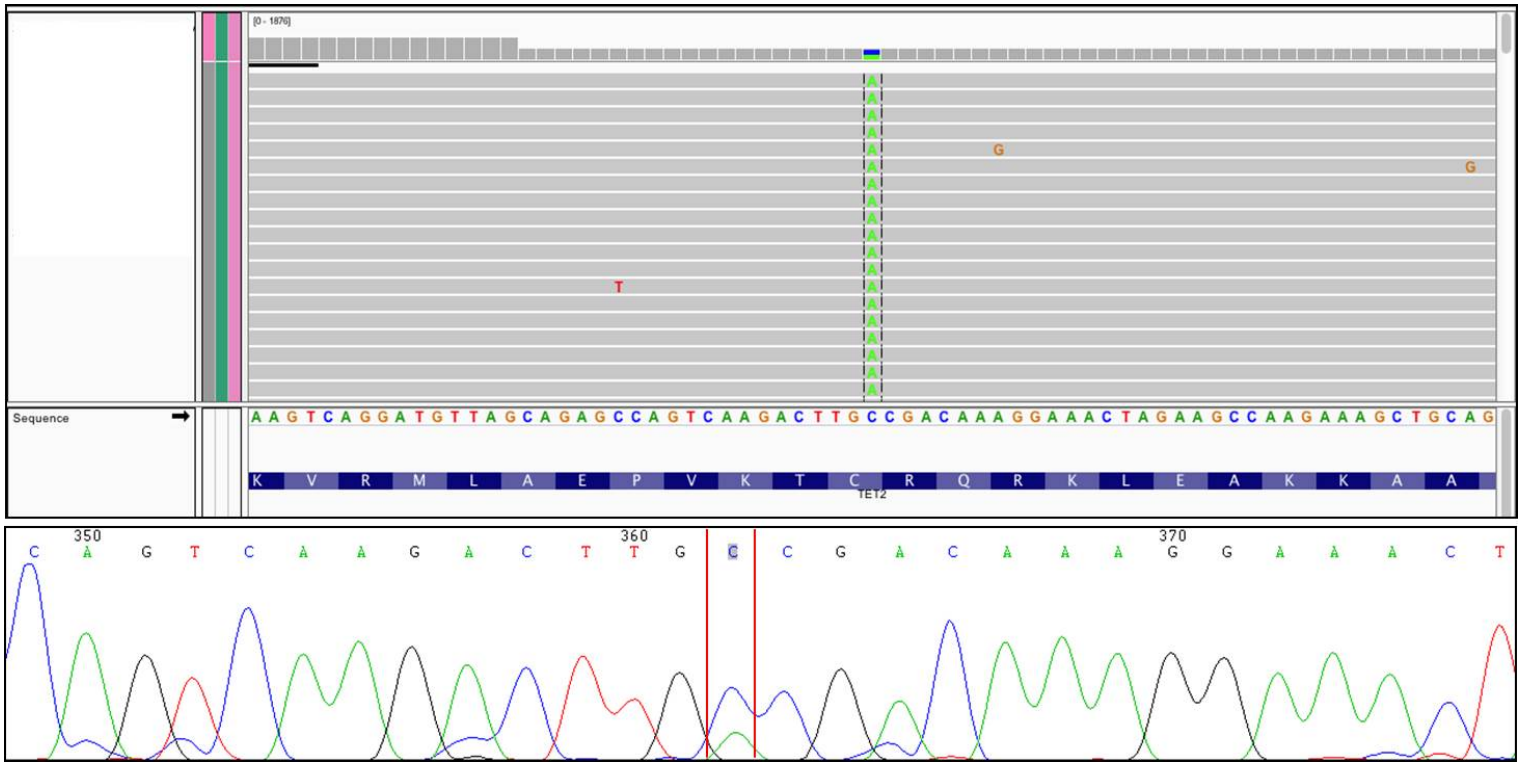

Figure 3.3 Detection of the *TET2* Cys1464X mutation in sample Test001 by both Sanger and NGS.

The mutated C>A nucleotide can be seen in the MiSeq sequencing traces (top pane) and at a low level in the Sanger sequencing trace (green peak at the indicated position, bottom pane). This mutation was missed during analysis of the Sanger data due to the high background noise level.

Table 3.5 *JAK2* V617F variant allele frequencies in RT-PCR and MiSeq data

	RT-PCR			MiSeq			
Sample ID	***JAK2* WT Copy Number**	**V617F Copy Number**	**VAF**	**Total Depth**	**Reference Depth**	**Variant Depth**	**VAF**
JAK2_A	60035	1865	0.03	8205	7564	628	0.08
JAK2_B	51617	5929	0.10	7924	6540	1366	0.17
JAK2_C	58408	9834	0.14	7883	5859	2015	0.26
JAK2_D	52331	7917	0.13	7828	6342	1472	0.19
JAK2_E	59013	852	0.01	7411	7219	177	0.02
JAK2_F	50564	10411	0.17	8139	6390	1719	0.21
JAK2_G	36490	11804	0.24	7637	5290	2333	0.31

The variant allele frequency of the $JAK2^{V617F}$ positive samples, as determined by the real-time PCR assay, ranged from 1-24%. All mutations with a VAF >3% (6/7, 86%) were successfully aligned and called in the MiSeq data. The remaining mutation (1% VAF) was present in the sequencing reads, but was below the detection limit of the variant calling software.

3.4.1.3 Calculation of Assay Background Interference

The background noise level of the assay was calculated by investigating the sequencing read composition at 31 SNP loci over 14 chromosomes in 15 samples (Table 3.6). The SNPs were all initially identified by the data analysis pipeline, are bi-allelic and are all recorded in dbSNP135 as being non-pathogenic. At each locus (n=465), the level of background noise was determined by expressing the number of sequencing reads containing any of the alternate nucleotides (3 in the case of homozygous SNPs, two in the case of heterozygous SNPs) as a percentage of the total number of calls at that position.

The mean level of background noise in the assay was thus determined as 0.31% (range 0.0-0.8%) across all SNP loci in all samples, and was consistently low, both between the SNPs (mean 0.31%, range 0.1-0.8%), and between the samples (mean 0.33%, range 0.25-0.55%) (Figure 3.4). Interestingly, the background level at heterozygous loci was lower than that at homozygous loci (0.2% and 0.4% respectively). The data for the highest background measurements (2.4 and 4.0%, in different samples and at different loci) are likely skewed due to low sequence coverage (80 and 24 reads respectively).

Table 3.6 Details of 31 SNPs used for assay background noise analysis.

Chromosome	Genomic Co-ordinates (hg19)	dbSNP ID	Alleles
1	115,256,669	rs969273	C/T
2	198,266,943	rs788019	C/T
3	105,439,026	rs2305035	C/T
3	105,438,957	rs2305036	A/C
3	105,453,034	rs6768096	A/G
3	105,438,806	rs7624400	A/G
4	55,141,055	rs1873778	C/T
4	106,196,951	rs2454206	A/G
4	55,593,464	rs3822214	A/C
5	112,228,667	rs430665	C/T
7	148,508,833	rs2072407	C/T
7	148,506,064	rs740949	A/G
9	5,073,770	rs77375493	G/T
12	11,803,228	rs2541135	C/T
12	11,803,220	rs2856305	A/C
13	28,607,916	rs2491222	A/T
13	28,607,989	rs2491223	G/T
17	7,579,472	rs1042522	C/G
17	7,578,115	rs1625895	C/T
17	74,733,099	rs237057	C/T
19	7,806,868	rs1544767	A/T
20	31,025,231	rs2295763	A/G
20	31,025,163	rs2295764	C/T
20	31,024,274	rs4911231	C/T
20	31,022,959	rs6058694	C/T
21	36,206,932	rs11702841	A/G
X	15,838,366	rs2301724	C/T
X	76,937,963	rs3088074	C/G
X	76,940,534	rs35268552	A/G
X	15,808,795	rs4830994	C/T
X	133,511,988	rs5978001	A/G

All co-ordinates refer to hg19 genome build.

Taken together, the sensitivity of the panel is calculated at 1-3%, depending on both the particular locus examined and the variant caller software used for analysis.

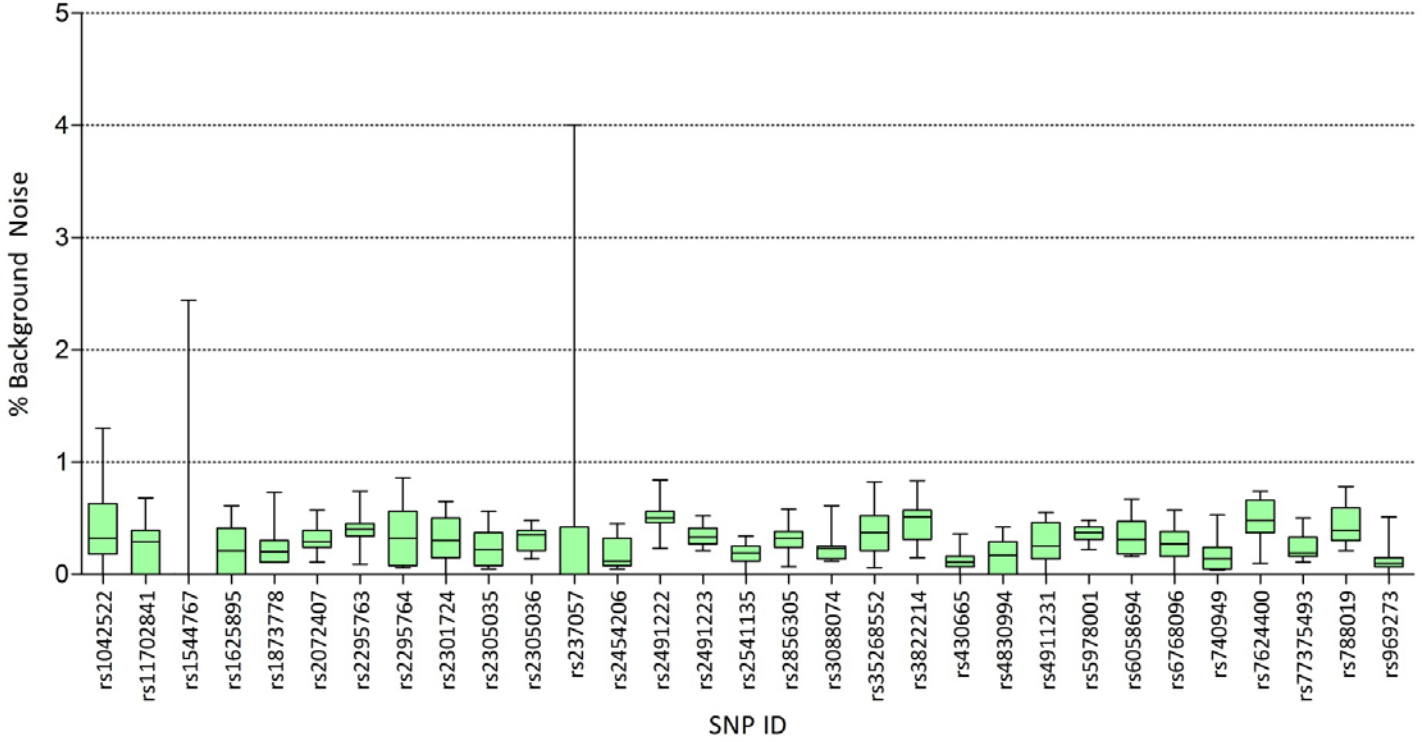

Figure 3.4 Comparison of background sequencing calls in NGS data.

Each boxplot represents the distribution of background sequencing calls at 31 SNP loci in 15 AML samples, represented as a percentage of total sequencing reads. The x-axis represents each of the 31 loci analysed. The y-axis shows the percentage of sequencing reads containing alternate base calls.

3.5 The Mutational Landscape of del(5q) MDS

The first application of this panel was the characterisation of the mutational landscape of a cohort of 43 early or advanced del(5q) myelodysplastic syndrome cases (Fernandez-Mercado *et al* 2013). DNA for the sample cohort, comprising 22 patients diagnosed with 5q- syndrome, nine with refractory anaemia and 12 with advanced MDS, was kindly provided by Dr Marta Fernandez-Mercado of the LLR Molecular Haematology Unit at the University of Oxford. I was responsible for sequencing the cohort using the myeloid panel, along with the initial data analysis, including alignment of the data to the hg19 reference genome, generation of a list of variants and filtering based on the quality of the calls. Dr Fernandez-Mercado provided further analysis in the context of the mutational landscape of del(5q) MDS.

A total of 29 mutations were detected in the cohort (Appendix A). Overall, 45% of early and 66.7% of advanced cases presented at least one mutation (Figure 3.5). Genes with the highest mutation frequency among the advanced cases were *TP53* and *ASXL1* (mutations present in 25% of cases each). Mutations in these genes occurred at a much lower frequency in the 5q- cases (4.5% and 13.6% respectively), suggesting a potential role for these genes as drivers of AML progression in del(5q) MDS.

Comment [a6]: Check this

The 5q- syndrome is widely considered to be relatively genetically stable compared to other MDS subtypes, a fact that is reflected in its relatively good prognosis. Previous studies have described mutations in a limited

number of genes, including *TP53* (Sebaa *et al* 2012), *JAK2* (Ingram *et al* 2006) and *ASXL1* in del(5q) MDS. In this study, we determined that over 40% of 5q- syndrome patients harbour a mutation in genes including *TET2, SF3B1, RUNX1, WT1* and *ASXL1*.

Table 3.7 List of mutations detected in 43 MDS cases

Sample ID	Gene	Genomic coordinates (hg19)	DNA change	Amino acid change	Variant Reads	Total Reads	VAF
MDS16	*RUNX1*	chr21:36,259,324	A>AG	Leu29Ser	23	72	0.32
MDS15	*SF3B1*	chr2:198,266,834	T>TC	Lys700Glu	170	1496	0.11
MDS07	*DNMT3A*	chr2:25,457,242	C>CA	Arg882Leu	92	1176	0.08
MDS08	*ASXL1*	chr20:31,022,449	insG	Gly646TrpfsX12	174	389	0.45
MDS08	*WT1*	chr11:32,413,565	C>CT	Arg462Gln	174	355	0.49
MDS14	*TET2*	chr4:106,193,748	C>CT	Gly646TrpfsX12	309	685	0.45
MDS12	*ASXL1*	chr20:31,022,449	insG	Gly646TrpfsX12	37	351	0.11
MDS12	*SF3B1*	chr2:198,266,834	T>TC	Lys700Glu	620	1549	0.40
MDS12	*TET2*	chr4:106,164,896	insA	Tyr1255X	41	771	0.05
MDS06	*TET2*	chr4:106,197,552	C>CT	Pro1962Leu	303	602	0.50
MDS11	*ASXL1*	chr20:31,022,902	G>GA	Trp796X	144	402	0.36
MDS10	*TP53*	chr17:7,578,413	C>CG	Val173Leu	109	265	0.41
MDS29	*JAK2*	chr9:5,073,770	G>GT	Val617Phe	-	-	-
MDS34	*JAK2*	chr9:5,073,770	G>GT	Val617Phe	-	-	-
MDS29	*DNMT3A*	chr2:25,457,176	G>GA	Pro904Leu	198	450	0.44
MDS28	*U2AF1*	chr21:44,514,777	T>TC	Gln157Arg	242	632	0.38
MDS30	*CBL*	chr11:119,149,332	C>CT	Ala447Val	99	227	0.44
MDS26	*TP53*	chr17:7,577,553	A>AG	Met243Thr	327	1166	0.28
MDS37	*TP53*	chr17:7,577,120	C>CT	Arg273His	620	754	0.82
MDS42	*ASXL1*	chr20:31,023,821	G>GT	Glu1102Asp	366	806	0.45
MDS42	*CBL*	chr11:119,149,004	G>GT	Trp408Cys	267	278	0.96
MDS36	*ASXL1*	chr20:31,024,704	G>GA	Gly1397Ser	875	1755	0.50
MDS33	*TET2*	chr4:106,196,850	insCATG	Glu1728Aspfs*13	121	713	0.17
MDS43	*TET2*	chr4:106,164,880	G>GT	Glu1250X	313	1145	0.27
MDS43	*ASXL1*	chr20:31,022,449	insG	Glu646TrpfsX12	470	1097	0.43
MDS39	*TP53*	chr17:7,578,190	T>TC	Typ220Cys	48	124	0.39
MDS39	*TP53*	chr17:7,578,275	G>GA	Gln192X	99	201	0.49
MDS38	*TP53*	chr17:7,577,538	C>CA	Arg248Leu	1168	2648	0.44
MDS38	*TP53*	chr17:7,577,568	C>CT	Cys238Tyr	998	2664	0.37

Total reads indicates the number of sequencing reads covering each locus, and includes both reference and variant calls. VAF is shown as a fraction of variant reads to reference reads. Abbreviations: VAF, variant allele frequency. All co-ordinates refer to hg19 genome build.

Six of the mutations detected are present at levels below the detection level of Sanger sequencing (15-20%). Sanger sequencing has been the gold standard of DNA sequencing for many years, and the vast majority of sequencing studies published to date have used this technology. It is likely, therefore, that previous studies have underestimated the prevalence of mutations in MDS.

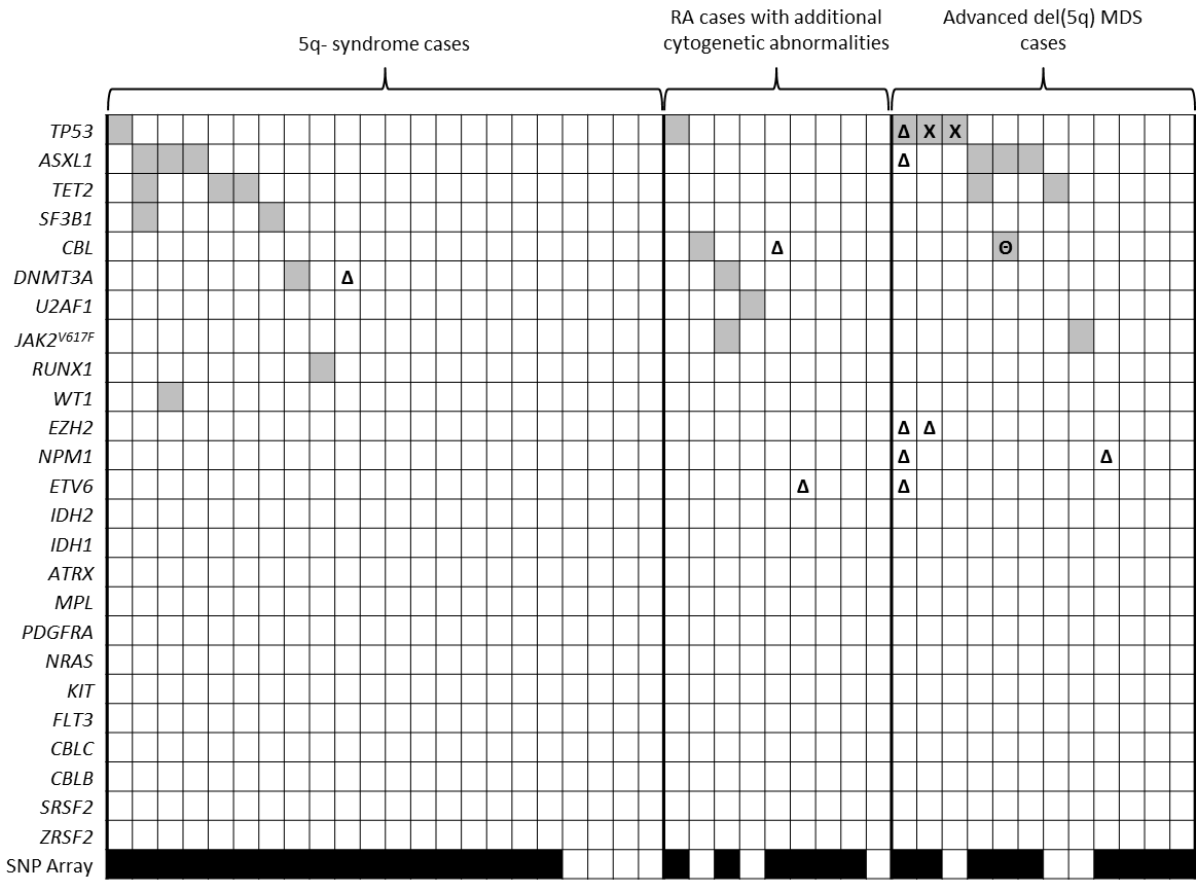

Figure 3.5 Mutations, deletions and loss of heterozygosity in 25 genes analysed in 43 del(5q) MDS samples.

Each column represents an individual sample with each row representing one gene in the sequencing panel. Grey boxes indicate mutated cases. Black boxes mark samples for which SNP-array data were available. Δ: gene encompassed within a region of cytogenetic loss. Θ: gene encompassed within a region of uniparental disomy (UPD) (Fernandez-Mercado *et al* 2013). X: two mutations in one gene in a single patient. Abbreviations: RA, refractory anaemia.

3.6 Diagnostic Application of the Myeloid Sequencing Panel

In addition to the explorative study into the molecular landscape of del(5q) MDS, the clinical utility of the panel has been demonstrated by applying it to a cohort of patients with confirmed or suspected myeloid disorders. Decisions regarding patients for whom the additional mutation data either altered or confirmed a diagnosis, impacted upon the treatment choice or determined prognosis, were made during weekly multi-disciplinary team (MDT) meetings involving relevant clinicians and clinical scientists.

In order to ensure that only clinically relevant mutations are returned to the clinicians, a reporting algorithm was designed, in consultation with both clinicians and clinical scientists, which would filter the mutation data for each patient. Only those variants with a VAF greater than 10%, a read depth of 50x and that are described as acquired in publically available databases were reported. Additionally, novel mutations predicted to produce a truncated protein were also reported.

The clinical utility of the assay was determined by considering changes in diagnosis, treatment or follow-up in response to the mutation data and the implications for prognosis as suggested by published disease-specific prognostic algorithms for AML (Patel *et al* 2012), MDS (Bejar *et al* 2011), myelofibrosis (Vannucchi *et al* 2013) and CMML (Itzykson *et al* 2013). In the year from April 2015, 270 clinical cases were submitted for sequencing

with the myeloid sequencing panel, comprising 66 AML cases, 74 MDS cases, 40 MPN cases and 90 with other pre-leukaemic malignancies, including primary myelofibrosis, pancytopenia and neutropenia. Mutations were identified in 60% of patients (n=164) of which 55% (n=91) were classified as clinically relevant (Figure 3.6, Appendix B).

Comment [a7]: Check this

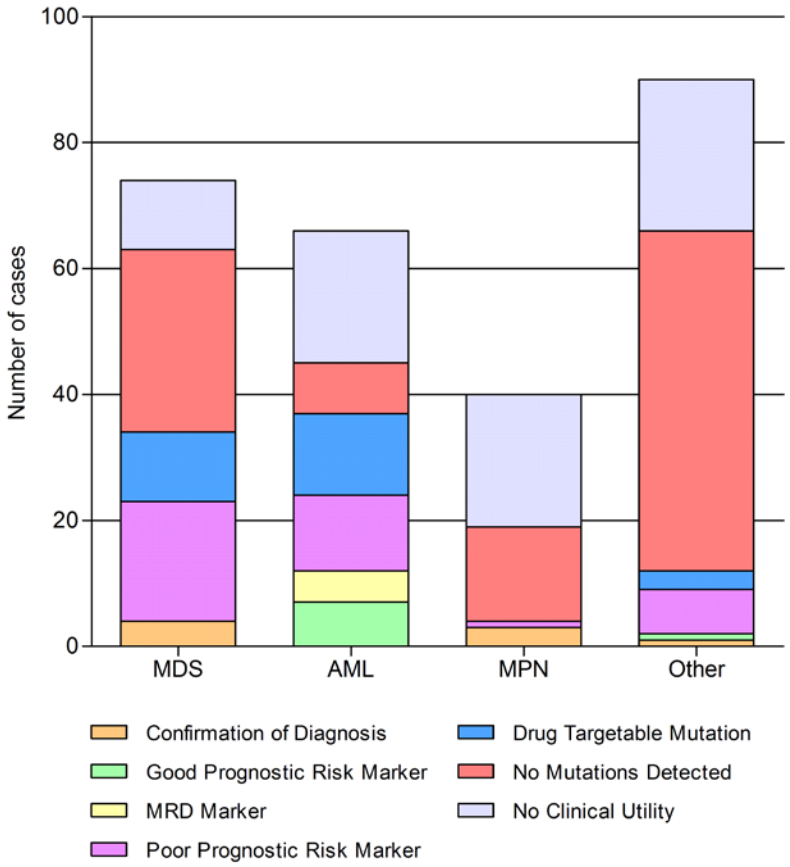

Figure 3.6 Clinical utility of the myeloid sequencing panel

Summary of the clinical implications of mutations detected in MDS, MPN, AML or other pre-leukaemic patients using the myeloid sequencing panel. Abbreviations: MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; AML, acute myeloid leukaemia; MRD, minimal residual disease.

Twenty-seven cases harboured mutations targetable by currently available treatments. Forty-three (61%) of the clinically important mutations indicated poor risk disease, with a further eight (9%) indicating good risk. Five AML patients (7%) harboured insertions in *NPM1*, which can be used to monitor minimal residual disease (MRD) (Ivey *et al* 2016). The final eight mutations (9%) confirmed a suspected diagnosis, including one case in which the presence of both *RUNX1* and *TET2* mutations confirmed a CMML diagnosis without the need for a bone marrow extraction. Furthermore, one case was found to have no mutations in any of the targeted genes, lending weight to the patient having something other than a myeloid malignancy.

3.7 Discussion

This study demonstrates the efficacy of a targeted next-generation sequencing panel for use in routine diagnostics. All 17 mutations present in the validation cohort were all ultimately identified in the MiSeq data, demonstrating both the robustness of the NGS chemistry, and concordance with the current gold-standard methods of mutation detection.

After single base substitutions, indels are the second most frequent form of variation found in the human genome, and are often linked to the pathogenesis of a disease where they alter the amino acid sequence of a coding gene. The majority of indels are between two and 16bp in size (Mullaney *et al* 2010), but there still remains a significant number of events

ranging up to 1,000bp (Mills *et al* 2011). These larger indels are difficult to detect in NGS data, due to the possibility of them being larger than the sequencing read length. A number of tools are available for indel detection, which vary in their methodologies. A recent comparison of different aligners demonstrated the strengths and weaknesses of each, with Pindel being able to detect indels up to 2,000bp in whole exome data, whilst GATK is better suited to 10-100bp indels (Ghoneim *et al* 2014). This study reinforces the importance of using the appropriate tools, in combination, for variant calling. In particular, the standard BWA/GATK pipeline was only able to correctly identify the single base substitutions and smaller indels, with Pindel and Platypus identifying both *FLT3*-ITDs and the *TP53* deletion.

The high sensitivity and low background noise level of the assay enabled the detection of mutations as low as 3% VAF. Recent studies have shown that low level, sub-clonal, mutations present at levels below the detection limit of Sanger sequencing, have the same impact on disease progression as high frequency clonal mutations. CLL patients with sub-clonal coding mutations in *TP53*, for example, demonstrate the same reduction in 5-year overall survival rates as those with clonal mutations (Nadeu *et al* 2016). Similarly, both clonal and sub-clonal *NOTCH1* mutations show a significant reduction in time to treatment (Nadeu *et al* 2016). A recent study in juvenile myelomonocytic leukaemia (JMML), showed that sub-clonal mutations in *SETBP1* confer reduced event free survival rates compared to wild-type *SETBP1* (Stieglitz *et al* 2015).

The finding that mutations were only detected in 10 of the 25 genes (40%) is not unsurprising. Whilst the genes selected for inclusion on the panel are recurrently mutated in myeloid disorders and, in some cases, have prognostic value, the recurrence of each varies between disease types. Genome-wide studies of AML have shown that despite a relatively low overall coding mutation count compared to other cancer genomes (Alexandrov *et al* 2013, Cancer Genome Atlas Research Network 2013), these mutations tend to be found in a small subset of genes, in particular *FLT3* (mutated in 20-30% of AML cases) (Fröhling *et al* 2002, Kottaridis *et al* 2001, Christian Thiede *et al* 2006, Whitman *et al* 2001), *NPM1* (48-64% AML cases) (Döhner *et al* 2005, Schnittger *et al* 2005, Thiede *et al* 2006), *ASXL1* (17% AML cases) (Schnittger *et al* 2013) and *DNMT3α* (18-30% AML cases) (Dufour *et al* 2010, Helbig *et al* 2014, Pabst *et al* 2001, Renneville *et al* 2009). By contrast, the same genes are mutated at a much lower frequencies in MDS, representative of its status as a pre-leukaemic condition. Furthermore, the del(5q) MDS subtype is considered to be relatively genetically stable, with very few somatic mutations (Scharenberg *et al* 2016) or chromosomal abnormalities (Wang *et al* 2008), a fact reflected in its generally good prognosis. These factors, along with the small cohort size in this study, and given that the panel was designed to be applicable to a wide range of myeloid malignancies, offer an explanation for the unmutated genes.

This panel has since been adopted as a routine test in the molecular haematology laboratory at the John Radcliffe Hospital, Oxford. In the year

since going live with the test, in April 2015, 270 cases have been screened using this test, with the results being fed back to clinicians to aid in the on-going diagnosis and treatment of AML and MPDs.

Chapter 4 WHOLE GENOME SEQUENCING OF 42 CHRONIC LYMPHOCYTIC LEUKAEMIA PATIENTS

4.1 Introduction

Chronic lymphocytic leukaemia (CLL) is the most common form of adult leukaemia in the western world (Swerdlow *et al* 2008), accounting for one third of new leukaemia cases each year (Siegel *et al* 2012). CLL is characterised by significant clinical heterogeneity. While one third of patients never require any treatment, the others invariably develop disease progression and, ultimately, resistance to chemotherapy. This clinical heterogeneity is reflected, at least in part, by characteristic biological features. Amongst the most frequent genomic aberrations are chromosomal deletions and amplifications (Döhner *et al* 2000), which have been associated with diverse phenotypes and contrasting clinical behaviours (Seiler *et al* 2006).

More recently, explorative genome-wide (Puente *et al* 2011, Schuh *et al* 2012) and exome-wide (Ljungström *et al* 2016, Quesada *et al* 2012) sequencing efforts have identified a number of additional, low frequency, recurrently acquired mutations in the coding regions of genes. Mutations in genes including *TP53* (Zenz *et al* 2008, Zenz, Eichhorst, *et al* 2010, Zenz, Mohr, Edelmann, *et al* 2009), *SF3B1* (Quesada *et al* 2012, D Rossi,

Bruscaggin, *et al* 2011, Wang *et al* 2011) and *NOTCH1* (Fabbri *et al* 2011, Rossi *et al* 2012, 2011b) have been associated with resistance to treatment, advanced disease, poor prognosis and transformation to Richters syndrome. To date, the majority of studies have focussed primarily on somatic changes within the coding sequence of a gene, although a recent study, using a combination of WGS and WES, identified recurrent non-coding mutations in CLL (Puente *et al* 2015), and started to define the significance and impact of these mutations. In that study, the authors defined a region of localised hypermutation, or kataegis, affecting a promoter site 300kb upstream of the *PAX5* gene. Furthermore, both WGS and WES approaches have revealed the clinical significance of sub-clonal evolution in CLL as a marker for poor prognosis (Landau *et al* 2013).

Furthermore, global analyses of WGS and WES data revealed the presence of specific mutation signatures in leukaemia and cancers (Alexandrov *et al* 2013, Nik-Zainal *et al* 2012, 2016). These signatures can be indicative of the mechanisms underlying the mutagenesis of the disease, for example ultra-violet light exposure in malignant melanomas or tobacco smoke in lung adenocarcinomas. For CLL, three mutational signatures have been proposed, associated with (1) the ageing process, (2) the activity of the APOBEC cytidine-deaminase enzyme family, and (3) polymerase η involved in the somatic hypermutation mechanism (Alexandrov *et al*, 2013).

Studies such as this illustrate the potential of whole genome sequencing to reveal important driver mutations outside of the coding regions. However,

the nature and biological significance of these non-coding regions in CLL remains largely unknown.

4.2 Aims and Objectives

The aim of this work is to provide a comprehensive analysis of the mutational landscape of the CLL genome, with reference to both somatically acquired mutations and structural variants.

4.3 Clinical and Biological Characteristics of the Patients

Informed consent from 42 CLL patients diagnosed and treated at the John Radcliffe Hospital, Oxford, was obtained in line with both the Declaration of Helsinki and the Oxford IRB ethics approval 09/H0606/5. Tumour and germline DNA was extracted from peripheral blood and saliva samples, respectively, in accordance with section 2.1. The cohort comprised 14 mutated and 25 unmutated IgHV cases, with a further three cases being classified as unmutated due to the presence of the IgHV3-21 sub-type, despite >98% homology to the germline sequence. The median age was 69 (range 49-94). Sixty-two per cent (26/42) were male, giving a male to female ratio of 1.6:1. Twenty-one cases were classified as being resistant to chemotherapy, defined as either the patient having relapsed within 24 months of initial treatment, or due to the presence of a disruption on chromosome 17 (either deletion of the 17p-arm or a mutation involving *TP53*). Eighteen cases were treated prior to sequencing (Table 4.1).

Table 4.1 Clinical and biological details of the patient cohort

Age	**n=42**
<69 years	19
≥69 years	23
Sex	
Male	26
Female	16
IgHV status	
Mutated	14
Unmutated	25
IgHV3-21	3
Chemo-resistance	
Positive due to relapse <24 months	10
Positive due to 17p disruption	11
Negative	21
Treatment status	
Prior treatment	18
No prior treatment	24

4.4 Results

4.4.1 Somatic Mutations Identified by WGS

Whole-genome sequencing was performed on tumour DNA extracted from peripheral blood samples taken from 42 CLL cases, alongside matched germline DNA from saliva samples, according to section 2.5.9. Tumour samples were sequenced to a mean depth of 39x (range 35-54x) and germline to a mean depth of 36x (range 28-54x). The removal of variants with either a low quality score, sequencing read depth below 10x or that were present in the matched germline sample, ensured that only high confidence, somatically acquired mutations were used for downstream

analysis. In total, 96,305 somatic mutations were identified across the 42 genomes, ranging from 153 to 4567 per sample (median = 2132) (Figure 4.1, Appendix C), corresponding to a mean mutation rate of 0.69 ±0.27 per megabase (range 0.05-1.52). A total of 7,944 genes harboured at least one mutation. Of these, 553 missense, 38 nonsense and 76 splice site mutations occurred in protein-coding regions, including one resulting in the loss of an initiator codon and one stop-loss event. Of the 6,055 insertions and deletions detected across the genomes, 54 affected protein-coding regions. One case, CLL134, appeared to harbour only 153 somatic mutations in total with no coding mutations at all (Figure 4.1). This is potentially due to the presence of tumour DNA in the germline sample, leading to removal of the majority of variants at the sequence data filtering stage.

Comment [a8]: Check still correct

Excluding mutations at the immunoglobulin loci on chromosomes 2, 14 and 22, patients with unmutated IgHV status exhibited significantly lower counts of both coding (P=0.0058) and total mutations (P<0.0001) (Figure 4.2). Conversely, there was no difference in either total (P=0.7312) or coding (P=0.1590) mutation rates between the chemo-sensitive and chemo-refractory groups. There was also no significant difference when comparing the number of mutations affecting the known CLL driver genes between the subgroups (Figure 4.2).

Comment [a9]: Landau refs

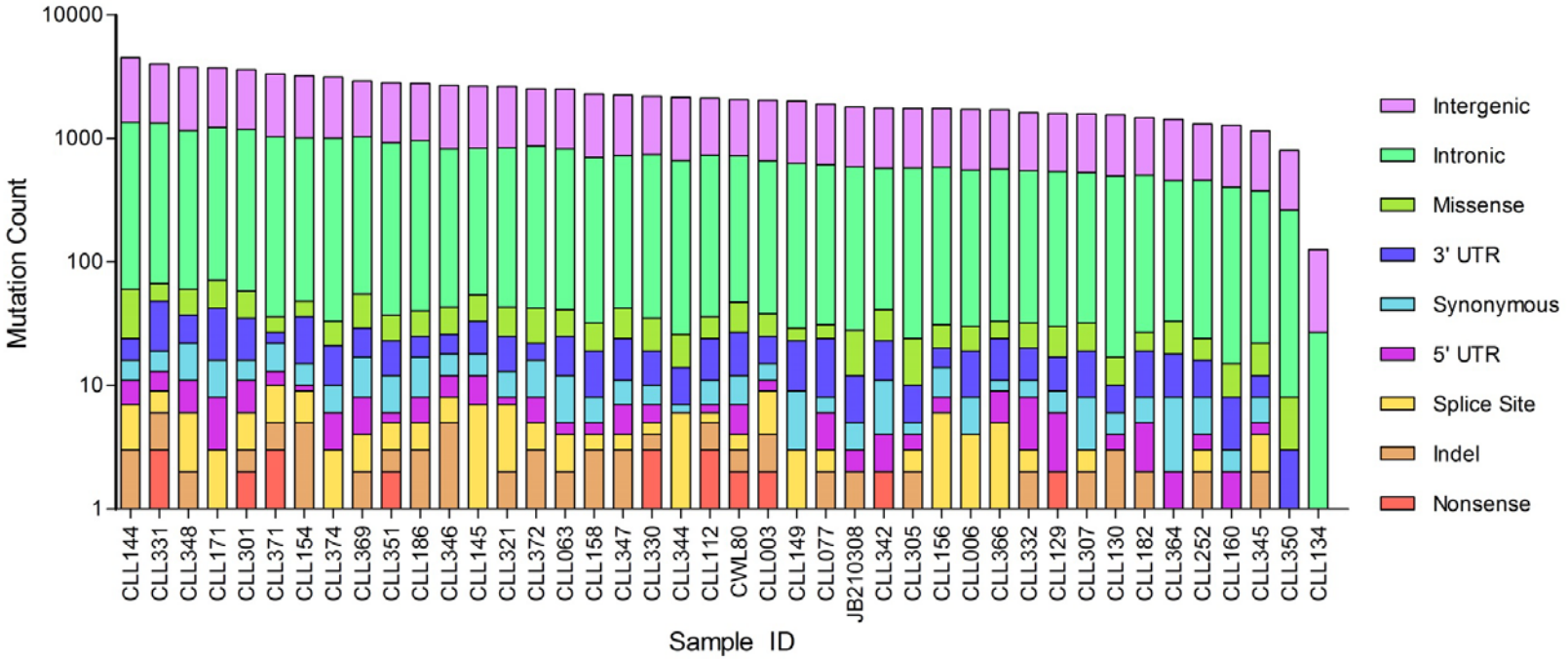

Figure 4.1 Distribution of mutations within gene boundaries across 42 CLL genomes.

Each column represents a single patient and shows the composition of mutation types identified within each. Patients are arranged according to total mutational load. The Y-axis is plotted on a log10 scale and shows total number of mutations for each patient.

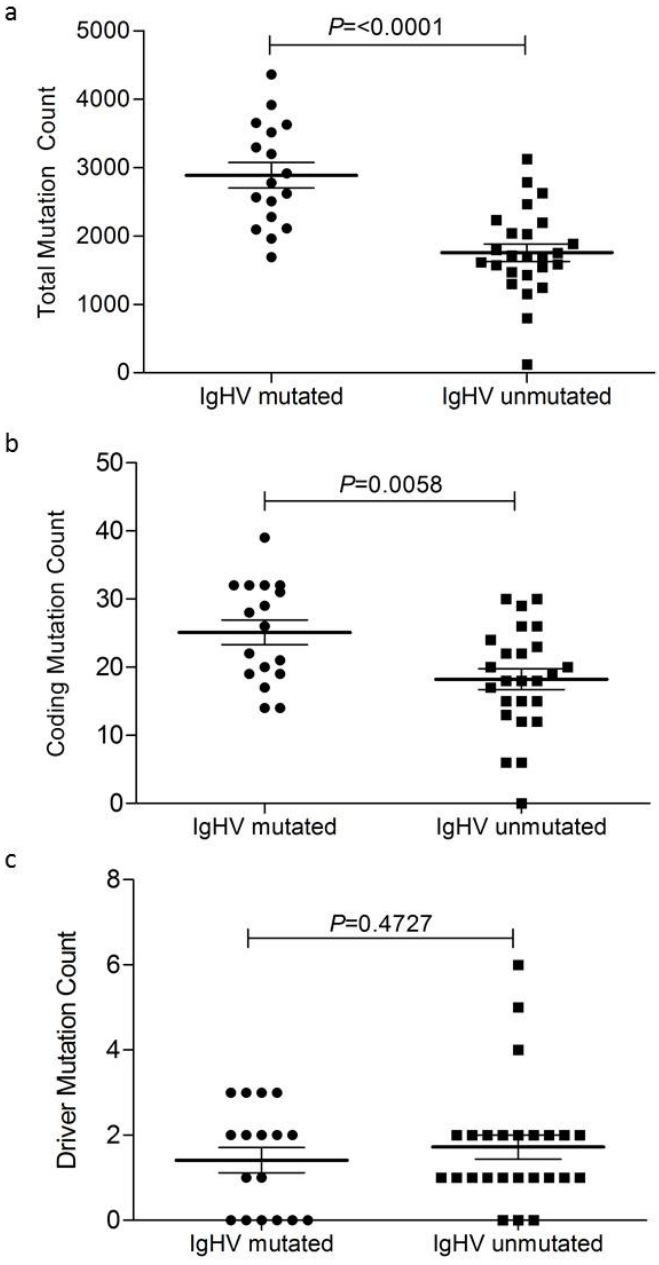

Figure 4.2 Total and coding mutation rates in IgHV mutated and unmutated CLL.

Dot plots comparing the (**a**) total, (**b**) coding and (**c**) driver mutation counts between IgHV mutated and unmutated cases. Each data point represents an individual patient.

4.4.2 Validation of Coding Mutations

To validate the presence of the somatically acquired mutations detected in the WGS data, a TSCA deep sequencing panel (Illumina, San Diego, CA, USA) was used. The panel was defined using the online DesignStudio pipeline (https://designstudio.illumina.com Illumina, San Diego, CA, USA) and was designed to target some, or all, of the coding region of 28 commonly mutated genes in CLL (Table 4.2). For the 40 cases for which tumour DNA was available, dual-barcoded sequencing libraries were prepared as described in section 2.5.8. The libraries were subjected to 2x151bp paired-end sequencing on a MiSeq instrument (Illumina, San Diego, CA, USA), to a mean depth of 2429x.

Table 4.2 Details of genomic regions targeted by the CLL TSCA sequencing panel.

Gene	Location	Chromosomal Coordinates (hg19)	Targeted Exons
ATM	11q22.3	108,093,559 - 108,239,826	All 63 exons
BCOR	Xp11.4	39,910,499 - 39,956,719	5, 10, 13
BIRC3	11q22.2	102,188,181 - 102,210,135	2-4, 9 & 10
BRAF	7q34	140,433,813 - 140,624,564	14
CDKN2A	9p21.3	21,967,751 - 21,975,132	All three exons
CHD2	15q26.1	93,443,551 - 93,571,237	20
DDX3X	Xp11.4	41,192,651 - 41,209,524	2, 10-12
EGR2	10q21.3	64,571,756 - 64,576,126	4
FBXW7	4q31.3	153,242,410 - 153,456,185	5, 9-11
HIST1H1E	6p22.2	26,156,559 - 26,157,343	1
IRF4	6p25.3	391,739 - 411,443	2 & 3
ITPKB	1q42.12	226,819,391 - 226,926,876	2 & 8
KLHL6	3q27.1	183,205,319 - 183,273,499	1 & 5
KRAS	12p12.1	25,358,180 - 25,403,854	2 & 3
MED12	Xq13.1	70,338,406 - 70,362,304	All 45 exons
MYD88	3p22.2	38,179,969 - 38,184,512	5
NOTCH1	9q34.3	139,388,896 - 139,440,238	34
NRAS	1p13.2	115,247,085 - 115,259,515	2 & 3
POT1	7q31.33	124,462,440 - 124,570,037	6-9
PTEN	10q23.31	89,623,195 - 89,728,532	All nine exons
RIPK1	6p25.2	3,076,998 - 3,115,421	8 & 10
SAMHD1	20q11.23	35,520,227 - 35,580,246	All 16 exons
SF3B1	2q33.1	198,256,698 - 198,299,771	14-16, 18
TGIF1	18p11.31	3,451,591 - 3,458,406	All three exons
TP53	17p13.1	7,571,720 - 7,590,868	All 11 exons
XPO1	2p15	61,705,069 - 61,765,418	16
ZFPM2	8q23.1	106,331,147 - 106,816,767	All eight exons
ZMYM3	Xq13.1	70,459,474 - 70,474,038	2, 6, 8, 15, 21, 34, 34

The majority of genes contained mutational hotspots that are well defined in the literature. In these cases, only those regions were targeted. In contrast, for genes where mutations are distributed throughout the coding sequence, in *ATM* for example, all exons were targeted. All co-ordinates refer to hg19 genome build.

4.4.2.1 Correlation with WGS Data

Of the 40 samples sequenced with the TSCA panel, 30 harboured a combined total of 57 mutations identified by the WGS data (Appendix D), and which were expected to be covered by the targeted sequencing panel. These mutations, located in 16 genes, comprised 48 non-synonymous substitutions and nine frameshift deletions, ranging in size from one to 19bp. The MiSeq successfully identified all 57 mutations. Furthermore, there was a high correlation of variant allele frequencies between both the WGS and MiSeq data (r=0.812, *P*<0.001, Spearman's rank correlation) (Figure 4.3).

Comment [a10]: Check still correct

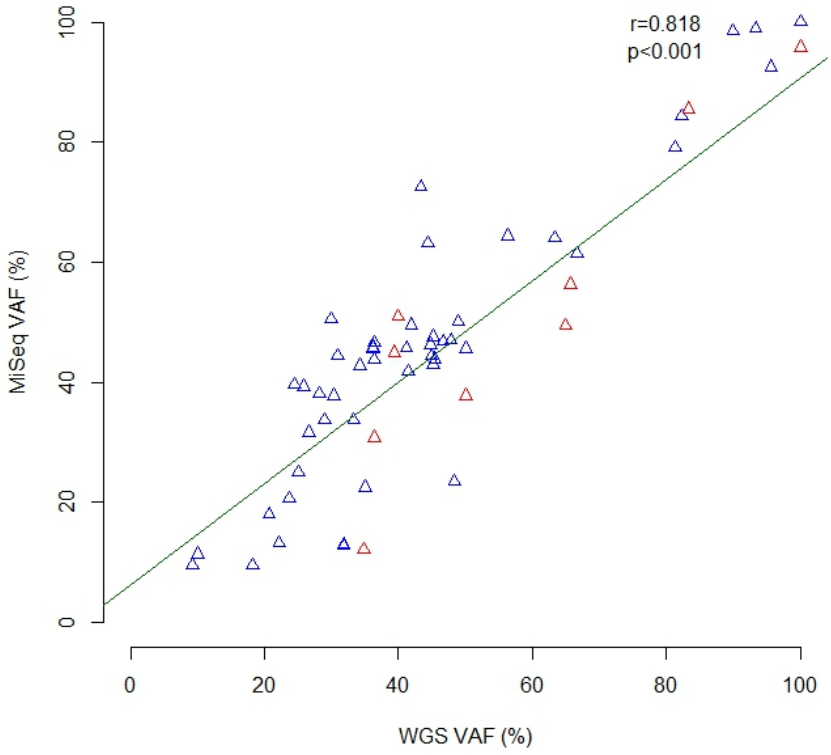

Figure 4.3 Correlation of variant allele frequencies between WGS and targeted deep sequencing.

Each data point represents an individual mutation. X-axis shows the VAF according to the WGS data. Y-axis shows the VAF according to the MiSeq targeted sequencing validation runs. Red data points indicate indels. Blue data points represent SNVs. Abbreviations: VAF, variant allele frequency.

4.4.2.2 Additional Mutations

In addition to the 57 known variants, four further mutations were identified in the MiSeq data (Table 4.3). These comprised three missense substitutions in *TP53* (Arg158His, Ser241Phe and Arg282Trp) in two cases (CLL307 and CLL374), and a single missense substitution in *MED12* (Glu33Gln) in CLL186.

Table 4.3 Additional mutations identified during WGS validation.

Sample	Gene	Variant	Chr	Coordinate (hg19)	Mutation	WGS		MiSeq	
						VAF	RD	VAF	RD
CLL186	*MED12*	G>C/C	X	70,338,701	Glu33Gln	100	5	100	22
CLL307	*TP53*	C>C/T	17	7,578,457	Arg158His	18	11	21	4187
CLL307	*TP53*	G>G/A	17	7,577,094	Arg282Trp	12	17	30	344
CLL374	*TP53*	G>G/A	17	7,577,559	Ser241Phe	82	11	79	2117

All co-ordinates refer to hg19 genome build. Abbreviations: Chr, chromosome; WGS, whole genome sequencing; VAF, variant allele frequency; RD, sequencing read depth.

Further investigation of the WGS data, in particular visualisation of the aligned reads using IGV (Robinson *et al* 2011) revealed that all four variants were present in the raw data, but were not listed in the variant call file. There was also no evidence of any variant reads in the respective germline sequencing data. This removes the possibility of the analysis pipeline identifying these mutations as germline, and thereby removing them from the analysis.

4.4.3 Mutations in CLL WGS Data

4.4.3.1 Recurrent Coding Mutations

As expected, the most frequently mutated genes, in terms of non-synonymous variants within the coding region, were *SF3B1*, *TP53*, *ATM* and *NOTCH1*, occurring in 31% (13/42), 21% (9/42), 14% (6/42) and 12% (5/42), respectively (Figure 4.4, Figure 4.5).

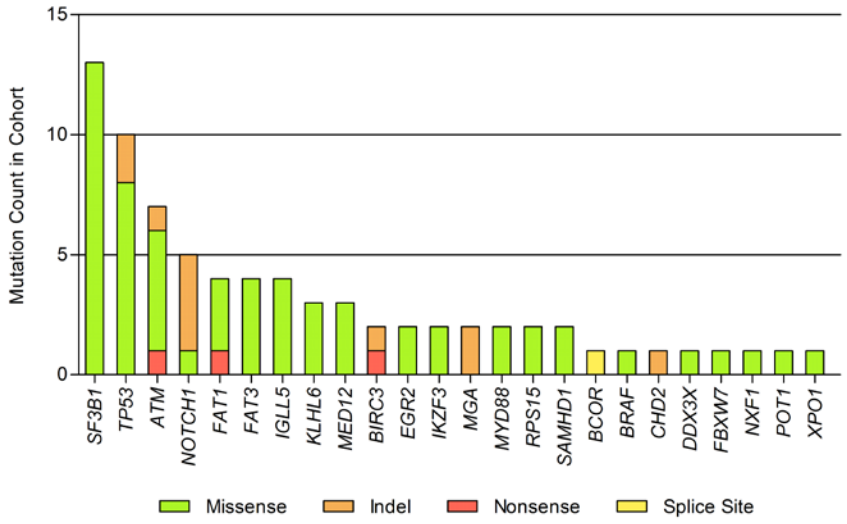

Figure 4.4 Recurrent coding mutations in CLL driver genes.

Bar chart showing the occurrence of coding mutations in genes understood to act as drivers of pathogenesis in CLL. Y-axis shows absolute mutation count for each gene.

Mutations in *SF3B1* were divided across 13 individuals, and tended to affect males more than females (*P*=0.0835). The most common mutation in *SF3B1* was the Lys700Glu substitution, identified in 31% (4/13) of mutated cases. The Glu622Asp substitution was seen in two patients, with one each of Tyr623Cys, Arg625Cys, Lys666Glu, Ile704Asn, Ile704Phe, Gly740Glu and Gly742Asp. All *SF3B1* mutations occurred between the third and fifth HEAT domains.

Missense mutations were the most frequently seen variant type in *TP53*, accounting for 80% (8/10) variants detected. All occurred within the DNA binding domain, and were recorded previously in the COSMIC database (http://cancer.sanger.ac.uk/cosmic). Furthermore, two cases harboured indels within the *TP53* coding region, one a 4bp deletion in exon 4, and the other a larger 19bp deletion in exon five.

Six cases contained a total of seven coding mutations in *ATM*, comprising five missense mutations, a 5bp deletion in exon 53 and a nonsense variant in exon 25, resulting in the loss of 60% of the coding sequence. Mutations in *ATM* co-occurred with del(11q) (4/6, *P*=0.008) and were associated with increased coding mutation rates (*P*=0.029).

Mutations in *NOTCH1* were confined to exon 34, with 83% (5/6) comprising the c.7541_7542delCT variant that results in the early termination of the PEST domain. The tumour suppressor gene *FAT1* harboured protein coding mutations in 9% of samples (4/42), encompassing both chemo-sensitive and refractory patients (Messina *et al* 2014).

Most *FAT1* mutations occurred within exon two (75%, 3/4), and included a nonsense mutation at the Lys125 residue, resulting in the loss of >97% of the coding region. Other previously described candidate driver genes harboured low frequency mutations in the cohort, including three in *MED12*, two each in *BIRC3*, *EGR2*, *IKZF3*, *MGA*, *MYD88*, *RPS15*, *SAMHD1*, and one each in *BCOR*, *BRAF*, *CHD2*, *DDX3X*, *FBXW7*, *NXF1*, *POT1* and *XPO1* (Clifford *et al* 2014, Fabbri *et al* 2011, Landau *et al* 2013, 2015, Ljungström *et al* 2016, Puente *et al* 2011, Schuh *et al* 2012, Wang *et al* 2011).

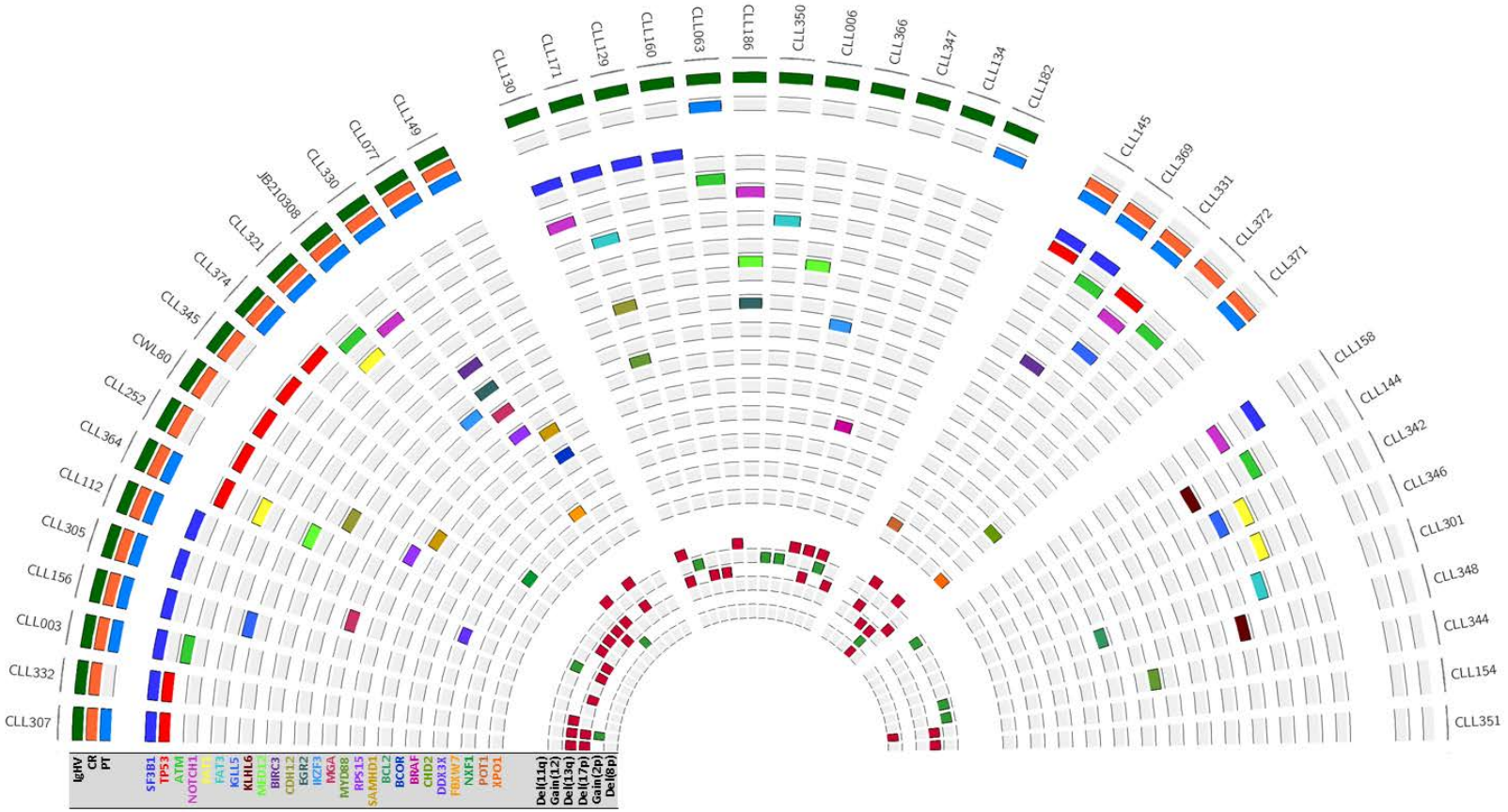

Figure 4.5 Co-occurrence of mutations in CLL driver genes and clinical risk factors in 42 CLL genomes.

Filled blocks indicate unmutated IgHV genes, chemo-refractoriness, prior treatment, the presence of a coding gene mutation or copy number aberration characteristic. Columns represent individual patients. The patients are arranged initially according to IgHV mutational status, then chemo-refractoriness. Abbreviations; CR, chemo-refractoriness; PT, prior treatment.

4.4.3.2 Mutations within Gene Boundaries

Mutations in non-coding regions of genes have been identified in a number of cancer types, including CLL (Kasar *et al* 2015), and have been linked to increased cancer susceptibility. To determine the genes most affected by somatic mutations, the mutation rate per bp was calculated for all genes. This included variants located in 5' UTRs, exons, introns and 3' UTRs of all genes with ≥ 2 mutations. The mutation rates were normalised for the length of the canonical transcript. A total of 4,075 genes were recurrently mutated.

Of the genes in the 99th percentile, a number have been previously identified with mutations in CLL; including *IGLL5* (Kasar *et al* 2015), *BTG2* (Landau *et al* 2013), *RPS15* (Ljungström *et al* 2016, Messina *et al* 2014, Quesada *et al* 2012) and *TP53* (Dufour *et al* 2013, Landau *et al* 2013, el Rouby *et al* 1993, Zenz, Eichhorst, *et al* 2010) (Figure 4.6).

The immunoglobulin Lambda Like Polypeptide 5 (*IGLL5*) gene harboured 107 mutations across 40% of cases (17/42), the majority of which were clustered in intron one (81%, 87/107) close to the boundary with exon one. Exon one harboured 6.5% (7/107) and 12% extended into intron two (13/107) (Figure 4.7). Mutations in exon one comprised three variants in the 5' UTR, with a further four missense mutations in the coding region, distributed across four patients. Patient CLL144 harboured two mutations in the 5' untranslated region; CLL331 contained one 5' UTR and one missense (Arg23His) mutation, CLL342 contained two missense (Ala44Pro

and Arg69Gly) mutations, whilst CLL156 contained a single missense mutation (Glu15Asp) that has been recorded previously in the COSMIC database (COSM1161415).

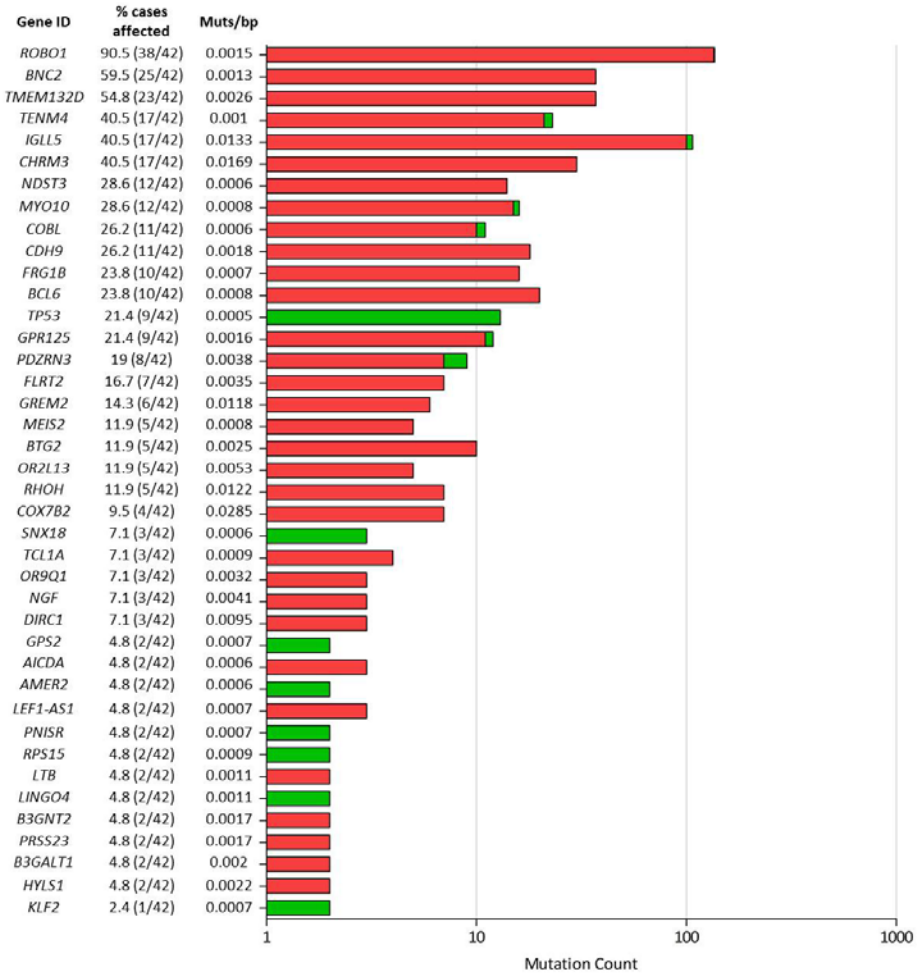

Figure 4.6 Genes with greatest mutation frequencies per bp.

Bar chart detailing the genes in the 99th percentile according to mutation rate per bp and their prevalence within the cohort. Mutation rates are calculated as the number of mutations, including coding and non-coding, divided by the length in bp of the canonical gene transcript. Red bars indicate intronic variants. Green bars indicate mutations in coding regions or UTRs. The X-axis represents the total number of mutations identified in each gene. Y-axis shows the gene IDs, the percentage and number of patients harbouring mutations, and the mutation rate per bp. Genes are arranged according to the percentage of cases affected.

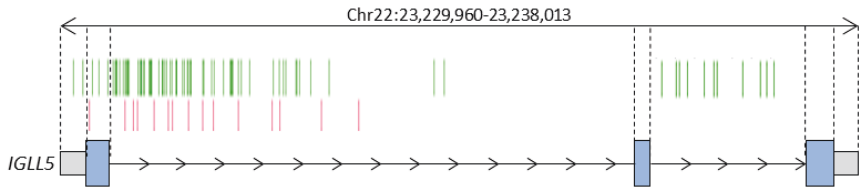

Figure 4.7 Somatic mutations in *IGLL5*.

Graphical representation of the IGLL5 gene showing the distribution of individual mutations in both IgHV mutated (green) and unmutated (red) CLL cases.

Of the 107 *IGLL5* mutations, 39% (42/107) have been recorded in the dbSNP database, including rs185781522 an Arg69Gly missense variant. This SNP is recorded in the ExAC database as being covered in <80% of individuals in the database, which may be indicative of a low-quality site and could, therefore, be a recurrent sequencing error. The presence of *IGLL5* mutations was closely associated with mutated IgHV status (P<0.001), although there was no relationship with either del(11q) (P=0.271), gain(12) (P=1), del(13q) (P=0.339) or del(17p) (P=1). ENCODE data for the GM12878 cell-line showed all *IGLL5* mutations were located in regions of low heterochromatin signal. The precise function of *IGLL5* remains largely unknown, other than it is a homolog of *IGLL1*, which is critical for the development of B-cells. Mutations in *IGLL5* have been described previously in a cohort of low-risk CLL, in which mutated cases showed a trend towards reduced expression compared to wild-type cases (Kasar *et al*, 2015). Furthermore, mutations in the 5' UTR and first exon were subclonal, whereas intronic mutations were clonal, suggesting that these two groups of mutation were acquired at different times in disease

development (Kasar *et al*, 2015). Analysis of RNA-Seq data, kindly provided by Dr Basile Stamatopoulos, showed no significant difference in *IGLL5* transcript level in this cohort when comparing either wild-type and coding mutations (*P*=0.439), wild-type and intronic mutations (*P*=0.417) or coding and intronic mutations (*P*=0.378) (Figure 4.8).

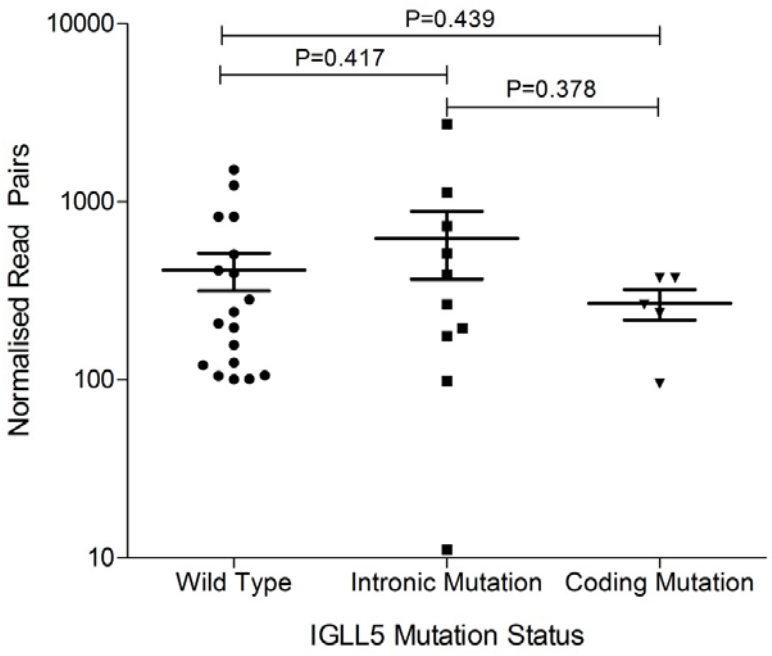

Figure 4.8 *IGLL5* expression levels in CLL.

Dot plots showing the gene expression levels associated with wild type IGLL5 transcripts, transcripts with intronic mutations and transcripts with coding mutations. Error bars show standard error of the mean (± SEM.)

Eleven intronic mutations across six patients in the tumour suppressor gene *BTG2* (five mutated and one unmutated IgHV cases) were also identified. All mutations clustered in intron 1, in an active promoter region,

however, RNA-Seq data showed no significant difference in transcript levels between patients with mutated *BTG2* and those without (*P*=0.1850).

Two IgHV mutated cases harboured three intronic mutations in *AICDA* (two in CLL366 and one in CLL158), which encodes the activation-induced cytidine deaminase (AID) enzyme responsible for somatic hypermutation in B-cells. All three mutations were located in a strong promoter region in the first intron. RNA-Seq data for these samples was unavailable.

4.4.4 Structural Variants

4.4.4.1 Copy Number Aberrations

Dr Sam Knight, of the Wellcome Trust Centre for Human Genetics, kindly provided copy-number aberration (CNA) analysis of the 42 CLL genomes. The analysis focussed initially on the most frequently reported CNAs in CLL, del(11q), gain(12), del(13q), del(17p), gain(2p) and del(8p). In this analysis, 33 of the 42 CLL cases (78%) harboured at least one CNA, with a median count of one. The most frequent CNA was del(13q), occurring in 43% of cases (18/42), followed by del(11q) (9/42, 21%) and del(17p) (8/42, 19%) (Table 4.4, Figure 4.5, Figure 4.9). Trisomy 12 was identified in 19% of CLL cases (8/42) and was found to associate with the chemo-sensitive cases (*P*=0.009).

Table 4.4 Recurrent copy number aberrations in 42 CLL genomes

Abnormality	No. cases (%)
Del(13q)	18/42 (43)
Del(11q)	9/42 (21)
Trisomy 12	8/42 (19)
Del(17p)	8/42 (19)
Gain(2p) (Chapiro *et al* 2010)	3/42 (7)
Del(8p) (Brown *et al* 2012)	2/42 (5)

In all 18 patients with del(13q), the minimally deleted region included *DLEU2*, mir3613 and both miR15a and miR16-1, with 89% (16/18) also covering the D13S319 locus and 50% (9/18) covering the *RB1* locus. All eight 17p deletions encompassed the *TP53* locus. Seven *TP53* mutations were shared between five del(17p) cases, including two frameshift (Tyr103Argfs*19 and Arg158Serfs*6), causing early termination with the loss of 69% and 59% of the coding region respectively. In addition to *TP53*, the MDR on chromosome 17 covered mir4520a, mir4520b, mir195, mir497 and mir324. The del(11q) MDR encompassed the *ATM* locus in all nine del(11q) cases, with four patients also harbouring *ATM* mutations, including one frameshift resulting in early termination (Ile2629Serfs*25). Partial or full gain of 2p (Chapiro *et al* 2010) was found in three of 42 cases (7%), ranging from 70Mb to 90.5Mb in length. The minimally amplified region contained the oncogenes *REL*, *ALK* and *MYCN*; however RNA-Seq data showed no increase in the expression levels of these targets between mutated and unmutated samples. Deletions of 8p were identified in three

patients; CLL364, CLL154 and CLL372, with an MDR of 18.6Mb stretching from 8p23.1 to 8p12. Patient CLL364 had concomitant del(11q) gain(12) and del(17p) events, CLL372 harboured both del(11q) and del(17p), whilst CLL154 harboured a 13q deletion. This concurs with previous studies that 8p deletions often co-exist with other copy number changes (Brown *et al* 2012, Forconi *et al* 2008). CLL364 was also the only case to display chromothripsis, defined as the presence of 10 or more copy number changes in a single chromosome.

4.4.4.2 Translocations

A total of 79 high confidence, somatic, inter-chromosomal translocations were identified across 30 patients (Figure 4.9, Table 4.5), with an average of 3.1 per case (range 1-13). With the exception of chromosome 22, all others were involved in at least one translocation. The most frequently affected region was the q-arm of chromosome 13, which accounted for 15% (23/158) of the translocation partners, and was affected in 11 cases. Of note, in five of those 11 cases, the translocation breakpoint centred on 13q14 and was concurrent with del(13q) at the same position.

The next most frequently affected regions were 1p (5.7%), 3q (5.7%), 14q (4.4%) and 17p (4.4%). One of the seven cases with 17p translocations, only one (CLL350) was centred on 17p13.1, and there was no evidence of a concurrent del(17p) event in this patient.

Seventeen of the 79 events (22%) are predicted to result in gene fusion proteins (tier one translocations), with a further 44% (35/79) involving a

single gene (tier two). The remaining 27 translocations involved only intergenic regions (tier three). The number of tier one and tier two events in refractory patients was significantly higher than in untreated cases (P<0.001), reflecting the increased genomic instability of this molecular subgroup.

Table 4.5 Details of translocations detected in 42 CLL genomes.

Sample ID	Chr	Coordinates (hg19)	Gene	Chr	Coordinates (hg19)	Gene	Tier
CLL372	7	8,663,231	*NXPH1*	5	37,709,720	*WDR70*	One
CLL006	8	70,602,300	*SLCO5A1*	1	91,853,200	*HFM1*	One
CLL345	9	33,130,549	*B4GALT1*	6	43,655,549	*MRPS18A*	One
CLL006	9	131,457,166	*SET*	2	116,376,668	*DPP10*	One
CLL006	13	21,746,650	*SKA3*	11	108,585,748	*DDX10*	One
CLL130	13	42,750,778	*DGKH*	2	214,996,194	*SPAG16*	One
CW_L80	15	41,853,322	*TYRO3*	13	44,069,827	*ENOX1*	One
CLL374	16	11,874,783	*ZC3H7A*	1	97,549,042	*DPYD*	One
CLL364	16	48,434,657	*MIR5095*	10	70,695,546	*DDX50*	One
CLL331	17	20,625,713	*NR_073509*	15	42,581,488	*GANC*	One
CLL372	17	42,116,511	*LSM12*	3	185,532,743	*LGF2BP2*	One
CLL252	18	9,386,313	*TWSG1*	17	27,269,460	*PHF12*	One
CLL252	18	29,085,459	*DSG2*	17	27,124,612	*FAM222B*	One
CLL301	18	60,793,528	*BCL2*	14	86,373,086	*IGH*	One
CLL348	18	60,793,528	*BCL2*	14	86,373,086	*IGH*	One
CLL186	19	45,210,662	*BCL3*	14	24,996,194	*IGH*	One
CLL372	X	96,772,745	*DIAPH2*	15	63,807,663	*USP3*	One
CLL130	1	117,690,373	*VTCN1*	10	52,548,106	Intergenic	Two
CW_L80	1	156,186,624	*PMF1 / PMF1-BGLAP*	20	26,190,323	Intergenic	Two
CLL301	2	11,929,913	*LPIN1*	4	162,053,779	Intergenic	Two
CLL130	2	214,929,829	*SPAG16*	1	89,072,486	Intergenic	Two
CLL374	3	12,526,084	*TSEN2*	19	30,862,678	Intergenic	Two
CLL307	3	31,620,715	*STT3B*	4	44,130,189	Intergenic	Two
CLL372	4	37,977,988	*TBC1D1*	18	52,407,341	Intergenic	Two
CLL332	4	66,413,946	*EPHA5*	2	42,051,404	Intergenic	Two
CLL006	4	103,872,694	*SLC9B1*	2	91,925,936	Intergenic	Two
CLL130	6	13,611,391	*SIRT5*	15	40,406,783	Intergenic	Two
CLL331	8	94,666,356	*LINC00535*	19	9,277,839	Intergenic	Two
CLL331	9	15,449,336	*SNAPC3*	5	52,041,533	Intergenic	Two
CLL364	9	18,951,669	*SAXO1*	4	24,713,600	Intergenic	Two
CLL364	9	27,511,928	*MOB3B*	12	64,573,814	Intergenic	Two

CLL331	9	28,174,273	*LINGO2*	5	101,286,989	Intergenic	Two
CLL342	9	115,804,583	*ZFP37*	4	114,699,773	Intergenic	Two
CLL331	10	74,767,587	*P4HA1*	17	32,673,513	Intergenic	Two
CLL371	10	74,767,587	*P4HA1*	17	32,673,499	Intergenic	Two
CLL364	12	46,277,056	*ARID2*	4	34,711,038	Intergenic	Two
CLL130	13	41,823,533	*MTRF1*	1	89,313,108	Intergenic	Two
CLL160	13	49,585,503	*FNDC3A*	3	106,401,370	Intergenic	Two
CLL160	13	51,495,444	*RNASEH2B*	3	98,355,667	Intergenic	Two
CLL160	13	51,624,971	*GUCY1B2*	3	145,020,376	Intergenic	Two
CLL344	13	74,708,039	*KLF12*	9	118,726,343	Intergenic	Two
CLL129	13	74,742,546	Intergenic	5	159,707,167	*CCNJL*	Two
CLL374	13	99,957,834	*UBAC2 / MIR548AN / GPR183*	17	21,345,246	Intergenic	Two
CLL364	14	36,186,681	*RALGAPA1*	12	11,652,111	Intergenic	Two
CLL144	14	101,113,952	Intergenic	6	41,927,166	*CCND3*	Two
CLL372	14	101,521,935	Intergenic	1	225,498,498	*DNAH14*	Two
CLL145	15	90,328,126	*ANPEP*	13	105,798,330	*Intergenic*	Two
CLL006	17	21,303,192	*KCNJ12*	20	29,517,483	Intergenic	Two
CLL364	17	21,303,192	*KCNJ12*	20	29,517,483	Intergenic	Two
CLL374	19	4,453,370	*UBXN6*	3	74,751,470	Intergenic	Two
JB 210308	19	11,421,943	*TSPAN16*	1	38,241,383	Intergenic	Two
CLL372	21	19,655,584	*TMPRSS15*	13	79,548,645	Intergenic	Two
CLL374	3	74,756,594	Intergenic	1	121,485,433	Intergenic	Three
CLL364	7	57,754,536	Intergenic	6	154,216,103	Intergenic	Three
CLL130	10	51,740,169	Intergenic	1	89,309,996	Intergenic	Three
CLL364	12	46,007,143	Intergenic	4	33,814,499	Intergenic	Three
CLL182	12	68,185,130	Intergenic	6	157,613,848	Intergenic	Three
CLL374	12	127,940,972	Intergenic	8	92,457,820	Intergenic	Three
CLL129	13	46,899,835	Intergenic	5	163,791,833	Intergenic	Three
CLL160	13	47,335,753	Intergenic	3	106,401,127	Intergenic	Three
CLL330	13	48,627,074	Intergenic	8	57,735,922	Intergenic	Three
CLL160	13	48,710,887	Intergenic	3	104,937,978	Intergenic	Three
CLL160	13	49,179,166	Intergenic	3	145,020,088	Intergenic	Three
CLL160	13	50,464,198	Intergenic	3	98,213,117	Intergenic	Three
CLL160	13	50,464,300	Intergenic	3	98,358,948	Intergenic	Three
CLL130	13	52,735,038	Intergenic	10	50,643,371	Intergenic	Three
CLL372	13	79,548,569	Intergenic	11	81,957,761	Intergenic	Three
CLL364	14	39,662,187	Intergenic	12	11,667,959	Intergenic	Three
CLL130	15	40,405,166	Intergenic	12	123,135,798	Intergenic	Three
CLL372	17	6,048,351	Intergenic	6	77,287,806	Intergenic	Three
CLL350	17	10,138,719	Intergenic	11	37,429,056	Intergenic	Three
CLL156	17	21,666,669	Intergenic	13	63,621,280	Intergenic	Three
CLL345	17	41,486,914	Intergenic	11	83,028,992	Intergenic	Three

CLL077	19	27,738,439	Intergenic	1	121,485,249	Intergenic	Three
CLL374	19	27,872,204	Intergenic	3	42,747,598	Intergenic	Three
CLL351	19	36,066,634	Intergenic	5	71,146,742	Intergenic	Three
CW_L80	20	26,148,193	Intergenic	13	63,637,910	Intergenic	Three
CLL149	20	29,438,015	Intergenic	6	154,215,663	Intergenic	Three
CLL154	M	16,571	Intergenic	11	49,883,563	Intergenic	Three

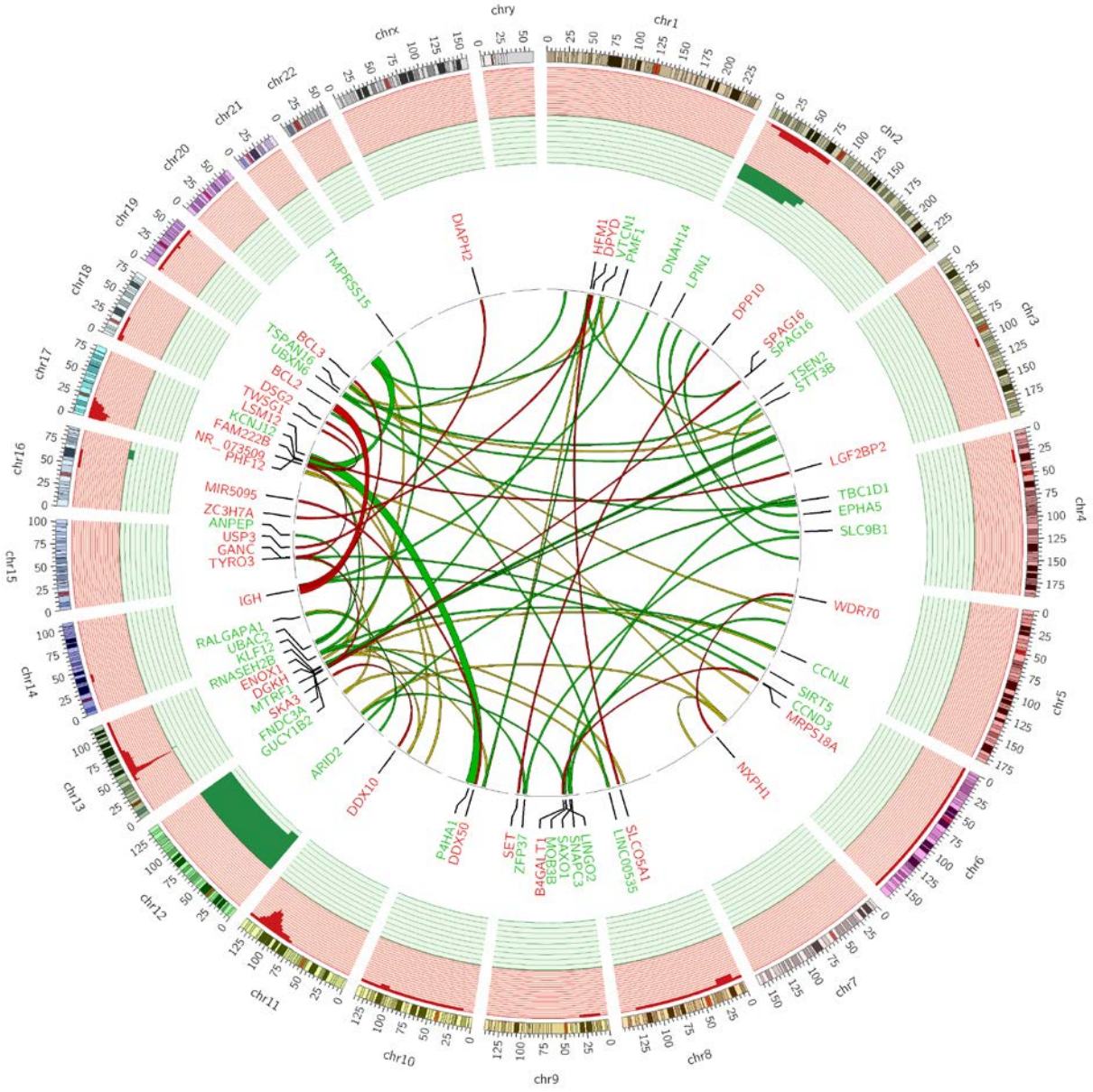

Figure 4.9 Structural rearrangements in CLL.

Circos plot depicting the frequency and location of the CNAs and translocations detected in 42 CLL genomes. The two outermost tracks represent CNAs; CN loss (red) and CN gain (green), with each track scaled appropriately for each MDR (CN loss n=18, CN gain n=8). The centre plot shows translocations identified in the cohort. Red links indicate tier one events (gene:gene), green links represent tier two events (gene:intergenic) and grey indicate tier three (intergenic:intergenic). The widths of the bands are indicative of the number of each individual event detected in the cohort.

Two untreated cases (CLL301 and CLL348) harboured *IGH/BCL2* (14;18) translocations, one each anchored in the 3' and 5' UTR of *BCL2,* respectively. This was the only copy number or structural variant present in CLL301, however CLL348 also harboured trisomy 12. A single case with an *IGH/BCL3* translocation (14;19) was also identified. Interestingly, a translocation between the first exon of the *RNASEH2B* gene on chromosome 13, and an intergenic region on chromosome 3 was found in CLL160. The protein encoded by *RNASEH2B* forms one subunit of RNase H2, which plays an important role in the maintenance of genome stability. Loss-of-function mutations in *RNASEH2B* are frequently seen in the inflammatory disease Aicardi-Goutieres syndrome (AGS) (Crow *et al* 2006). Interestingly, mutations in *SAMHD1* are also present in both AGS (Rice *et al* 2009) and CLL (Clifford *et al* 2014).

4.4.5 Identification of Regions of Kataegis

Localised regions with increased mutation rates, or kataegis, have been described previously in other types of cancer. The presence of kataegic regions was detected via the construction of rainfall plots as previously described (Nik-Zainal *et al* 2012). Each region was defined as containing a minimum of six substitutions, insertions or deletions whose intermutational distances are more than two standard deviations below the median intermutational distance for that sample (Lawrence *et al* 2013).

4.4.5.1 Kataegis within Individual Patients

Seventy-four instances of kataegis were identified in 31 samples affecting 14 chromosomes (Appendix E). The region size ranged from 219bp to 4.3Mb, with a median of 93,854bp. The total number of mutations across all regions was 1,591, with SNVs comprising the majority (1553/1591, 97%), followed by small deletions (30/1591, 2%) and insertions (17/1591, 1%). The number of mutations within each region ranged from six to 99, with a median of 14.

Comment [a11]: Check still correct

The spectrum of mutations contained within kataegis regions was significantly different to that found in the whole genome sequencing overall (Figure 4.10). The contribution of A>C, A>T and C>A transversions was significantly lower in the kataegic regions (P<0.001), whereas C>G transversions were enriched (P<0.001).

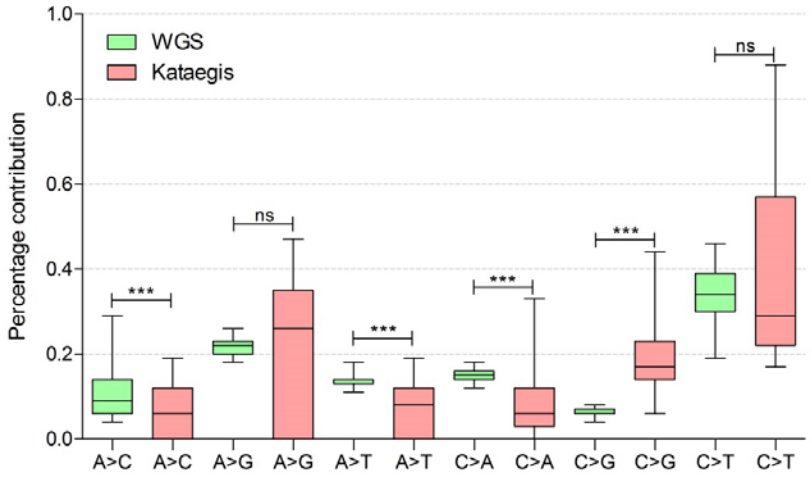

Figure 4.10 Comparison of mutation composition between whole genome sequencing data and kataegic regions.

Each box-and-whisker plot displays the contribution of each substitution type to the overall mutation burden in both the whole genome (green) and kataegic (red) regions. Whiskers represent minimum and maximum values.

Some chromosomal regions were recurrently affected, most notably, and as expected, those encoding the immunoglobulin light and heavy chains on chromosomes 2 (11/74, 15%), 14 (33/74, 45%) and 22 (13/74, 18%). These regions accounted for 88% (1,397/1,591) of all kataegic mutations (Figure 4.11).

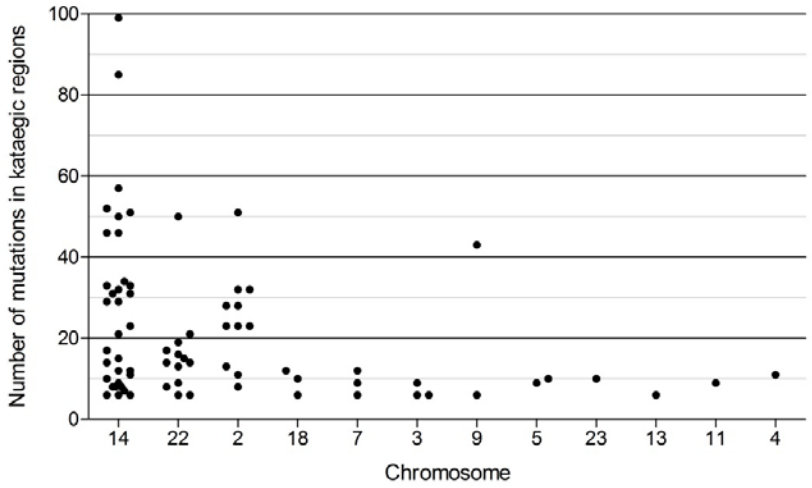

Figure 4.11 Individual kataegic regions in CLL.

Each data point represents a single kataegis event in one patient. Chromosomes are arranged according to the number of kataegic regions present. Y-axis indicates the number of mutations within each region.

Eighteen regions across 11 patients did not occur at sites of immunoglobulin rearrangement (Table 4.6), ranging in length from 219bp to 159kb. Thirteen of these regions were located within or across gene boundaries. Importantly, almost half of these (6/13 46%) affected non-coding regions of genes known to carry coding mutations in CLL including *KLHL6* (Puente *et al* 2011), *MEGF9* (Quesada *et al* 2013), *CDH12* (Landau *et al* 2013), *CADPS2* (Landau *et al* 2013, Puente *et al* 2011), *LRP1B* (Landau *et al* 2013, Quesada *et al* 2013) and *ATM* (Schaffner *et al* 1999, Stankovic *et al* 1999).

Table 4.6 Kataegis regions within individual CLL patients, excluding immunoglobulin gene loci.

Sample	Mutations in region	Chr	Region Start (hg19)	Region End (hg19)	Genes in Region
CLL156	8	2	140,885,911	141,045,820	*LRP1B*
CLL154	6	3	157,290,339	157,295,704	*C3orf55*
CLL348	6	3	183,273,058	183,273,364	*KLHL6*
CLL063	9	5	21,810,369	21,843,504	*CDH12*
CLL301	12	7	122,433,699	122,517,296	*CADPS2*
CLL301	6	7	122,622,586	122,638,447	*TAS2R16*
CLL351	43	9	123,416,699	123,479,606	*MEGF9*
CLL144	9	11	108,121,624	108,129,499	*ATM*
CLL307	6	13	51,664,141	51,665,954	*LINC00371*
CLL156	6	14	21,835,327	21,835,546	*SUPT16H*
CLL252	6	18	9,284,149	9,374,417	*ANKRD12, TWSG1*
CLL301	12	18	60,873,525	60,988,029	*BCL2*
CLL348	10	18	60,906,440	60,988,117	*BCL2*
CLL348	9	3	187,660,014	187,660,463	-
CLL371	11	4	59,097,561	59,114,524	-
CLL301	10	5	26,070,177	26,074,185	-
CLL371	9	7	49,672,829	49,682,118	-
CLL342	6	9	76,656,503	76,656,760	-

All co-ordinates refer to hg19 genome build.

One recurrent region of kataegis centred on the 5' region of *BCL2* on chromosome 18 occurred in two patients (Figure 4.12). These regions were 81,677bp and 114,504bp in length and harboured 10 and 12 mutations respectively, starting 1.5kb upstream of the start codon and terminating within the second intron. Both cases affected by kataegis in *BCL2* harboured concurrent t(14;18) translocations in the same gene. The breakpoint in CLL301 was located much further upstream, in the 3' UTR, whereas the CLL348 breakpoint was centred on the kataegic region in the 5' UTR. A further three mutations were identified in the same region in CLL154. Correlation of the position of these mutations against the ENCODE

database revealed overlapping regions of both mono- and tri-methylated histone 3 lysine 4 residues (H3K4me1 and H3K4me3, respectively), which are known markers of active gene enhancers and promoters, respectively. Analysis of the RNA-Seq data showed that cases with mutations in these regulatory regions displayed increased levels of *BCL2* transcript expression, compared to both cases with intronic mutations and wild-type cases (Figure 4.12). *BCL2* is over-expressed in CLL (Kitada *et al* 1998), but mutations in its coding region have not been described.

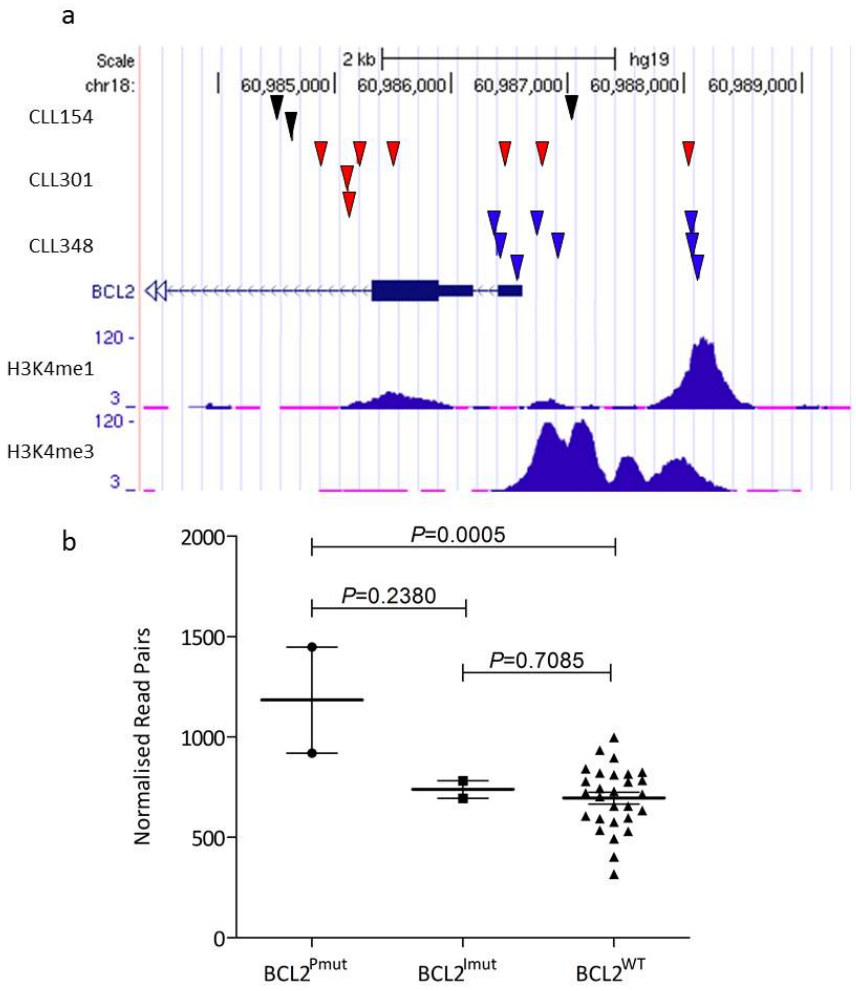

Figure 4.12 Regions of kataegis and gene expression levels in *BCL2*.

(**a**) Location of somatic mutations in *BCL2* in cases CLL154 (black), CLL301 (red) and CLL348 (blue) overlaying H3K4me1 and H3K4me3 data for the K562 cell line from ENCODE. (**b**) Dot plot showing the expression levels of the *BCL2* transcript comparing cases with no *BCL2* mutations (*BCL2*^{WT}), cases with intronic mutations downstream of exon one (*BCL2*^{Imut}) and cases with mutations in the promoter region (*BCL2*^{Pmut}). Error bars show ± SEM.

One case, CLL144, harbored a kataegic region within *ATM*, comprising nine mutations in a 7.8kb span between exon 10 and intron 15. Seventy-eight percent (7/9) of the mutations occurred in intronic regions, with two more in exon 10. The two coding mutations affected adjacent bases, resulting in

a single serine to valine residue substitution (Ser478Val), predicted as tolerated by SIFT. RNA-Seq data suggests a similar level of reduced *ATM* expression as those cases harboring del(11q), which itself was significantly lower than the wild-type cases (*P*=0.0005) (Figure 4.13), as expected (Stankovic *et al* 1999).

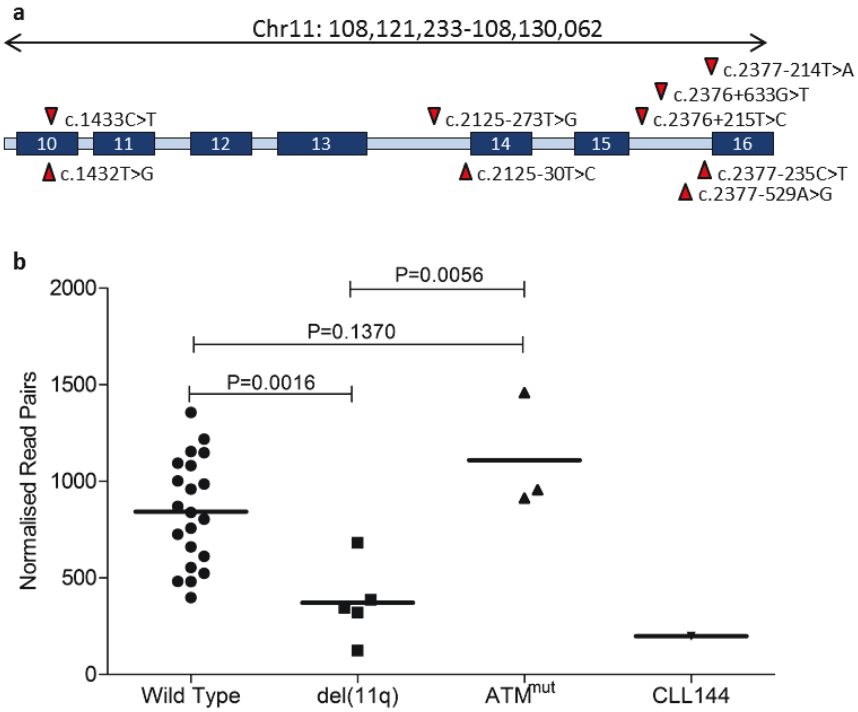

Figure 4.13 Kataegic *ATM* mutations in CLL144.

a) Graphical representation of *ATM* showing the distribution of kataegic mutations in CLL144. Dark and light blue represent exonic and intronic regions of the *ATM* transcript, respectively. **b**) Dot plot comparing the ATM transcript expression levels of CLL cases with no 11q disruption (wild type), those with del(11q), mutations in ATM (ATM^{mut}) and CLL144. Expression levels are shown as reads per million aligned reads. Error bars show ± SEM.

Kataegis was also observed in the non-coding region of one gene affected

by coding mutations in other haematological malignancies: *SUPT16H* (Abaan *et al* 2013, Patel *et al* 2012), however this region also overlaps an AluY repeat, and so could be a result of low sequence quality. The remaining four genes affected by kataegis; *ANKRD12*, *C3orf55*, *LINC00371* and *TAS2R16* have not been described in the context of either cancer or haematological disease and include a microRNA and an open reading frame. Overall, this means that eight of 14 (57%) of kataegis regions in genes were related to genes implicated in the pathogenesis of haematological disease.

Sites of kataegis have been shown to co-localise with structural rearrangements (Nik-Zainal *et al* 2012, Walker *et al* 2015). As such, the 19 kataegic regions were correlated with the CNA data. Of the 18 kataegis regions not at sites of somatic hypermutation, only three were found to be in close proximity to deletion events. The kataegis site in CLL307 was located 482bp downstream of a 13q deletion. CLL301 harboured a region located 19kb into a 141kb gain on chromosome 7. Sample CLL252 contained a region sitting between two deletions on chromosome 18, 21.8kb downstream of one and 765bp upstream of the other. In line with both the *BCL*2 data and previous observations (Nik-Zainal *et al* 2012, Walker *et al* 2015), this suggests that kataegis may also be a consequence of structural rearrangements.

4.4.5.2 Kataegis across the Cohort

By integrating the somatic variants from all 42 patients, an additional 66

regions with high mutation counts were identified. Regions in centromeric or telomeric areas, along with those in areas of segmental duplication were excluded. Seventeen contained genes affected by multiple mutations (Table 4.7).

Table 4.7 Kataegis regions across the cohort

No. of Patients Affected	Mutations in region	Chr	Start (hg19)	Stop (hg19)	Genes in Region
12	18	1	187,201,043	187,292,419	*LINC01036*
14	16	2	79,902,807	79,995,799	*CTNNA2*
6	19	3	60,528,213	60,588,662	*FHIT*
11	15	3	174,110,612	174,189,396	*NAALADL2*
12	24	3	187,418,868	187,490,720	*BCL6*
12	16	4	62,913,619	62,987,350	*LPHN3*
11	17	4	93,701,552	93,798,834	*GRID2*
14	19	4	136,800,043	136,899,101	*LINC00613*
12	18	4	158,200,175	158,289,441	*GRIA2*
14	15	4	162,406,955	162,494,646	*FSTL5*
10	15	4	162,502,368	162,596,645	*FSTL5*
12	16	5	19,801,833	19,892,529	*CDH18*
13	17	7	82,400,133	82,495,632	*PCLO*
14	25	13	70,404,177	70,498,648	*KLHL1*
14	18	18	64,102,223	64,199,022	*CDH19*

All co-ordinates refer to hg19 genome build. Abbreviations; Chr, chromosome.

A cluster of 20 mutations, in 24% of patients (10/42), were localised to a 1kb region in the first intron of *BCL6* (Figure 4.14a). No other *BCL6* mutations were detected in any other patients. All ten cases were of mutated IgHV status, and were predominantly treatment naïve (8/10, 80%). All 20 mutations were heterozygous single base substitutions. There were no recurrent mutations. ENCODE data revealed these mutations lie in a region of strong tri-methylation of the H3K4 residue, suggestive of an

active promoter (Figure 4.14a). Translocations involving the *BCL6* locus have been described in B-cell lymphomas; however there was no evidence of such a translocation in any of these samples. Mutations in the non-coding region of *BCL6*, in IgHV mutated CLL, have been described previously in similar frequencies to that found in this cohort, although no link between *BCL6* mutational status and the protein expression level has been identified (Capello *et al* 2000, L Pasqualucci *et al* 2000). Indeed, no significant difference in *BCL6* expression was observed between those with intronic mutations and wild-type cases (*P*=0.2261) (Figure 4.14b).

Other kataegic areas of interest include a 16 mutation region spanning 93kb across exon three of the tumour suppressor gene *CTNNA2* on chromosome 2, with mutations present in 33% of cases (14/42). The 16 mutations were evenly distributed between FR and treatment naïve patients, the majority of which were intronic (15/16, 94%), with only one missense variant in exon three (His552Pro). There is no record of a variant at this residue in the COSMIC database. Analysis of the RNA-Seq data showed no expression of *CTNNA2* in any sample, regardless of mutational status.

Another kataegic region of interest was identified in Lysine-specific methyltransferase 2C (*KMT2C*) on chromosome 7. Sixteen patients (38%) contained a combined 23 mutations in this region, with an enrichment in treatment naïve cases (10 vs 4, *P*=0.050). RNA-Seq data was available for 11 mutated and 21 unmutated cases, however no significant difference

was seen in transcript levels between the groups.

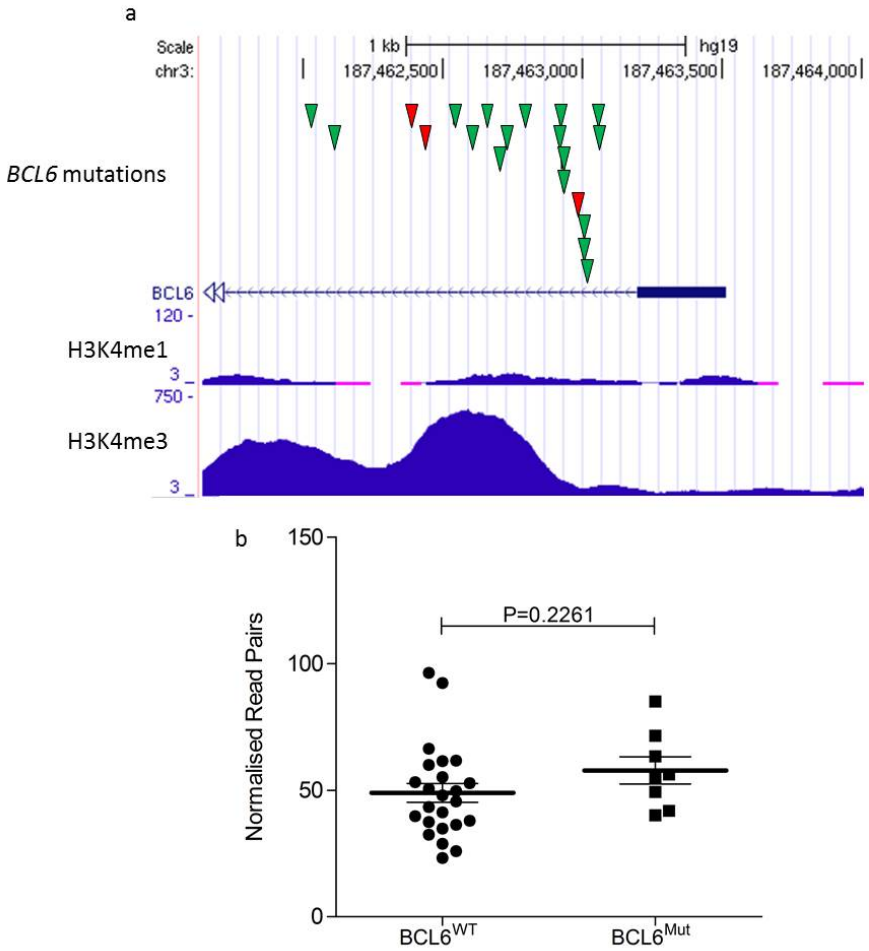

Figure 4.14 *BCL6* kataegic mutations and transcript expression levels.

(**a**) Schematic representation of the first exon and intron of the *BCL6* gene, showing the position of the 20 kataegic mutations. Green and red triangles represent mutations in untreated and refractory patients, respectively. Lower tracks display ENCODE ChIP-Seq data for the K562 cell line from ENCODE. High H3K4me3 signals are indicative of promoter regions. (**b**) Dot plot comparing *BCL6* transcript expression levels between patients with and without *BCL6* mutations. Error bars show ± SEM.

4.4.6 Extraction of Mutational Signatures

WGS and WES data facilitates the detection of distinct mutational profiles, or signatures, within many cancer types, including in CLL. Using methods

previously described (Alexandrov *et al* 2013, Brunet *et al* 2004) four distinct signatures were isolated from the tumour sequencing data, and two in the germline samples (Figure 4.15).

4.4.6.1 CLL Tumour Mutational Signatures

Tumour signature one (Tsig1) was primarily composed of substitutions involving cytosine residues, with C>T transitions and C>A transversions being the most common. Tsig1 was previously only identified in 2% of cancer cases, including breast and medulloblastoma (Alexandrov *et al* 2013).

Tsig2 had a relatively even contribution from all mutation types, with T>G transversions contributing the most. This signature was previously seen in nine different cancer types, including myeloma and B-cell lymphomas, but not in CLL, and has not been linked to any particular cause (Alexandrov *et al* 2013).

Like Tsig1, Tsig3 is also characterised by the presence of C>T transitions. However, in contrast to those of Tsig1, these dominate in an NpCpG context. Interestingly, Tsig3 is identical to the signature previously ascribed to the increased deamination activity of 5-methyl-cytosine, resulting in C>T substitutions at NpCpG motifs.

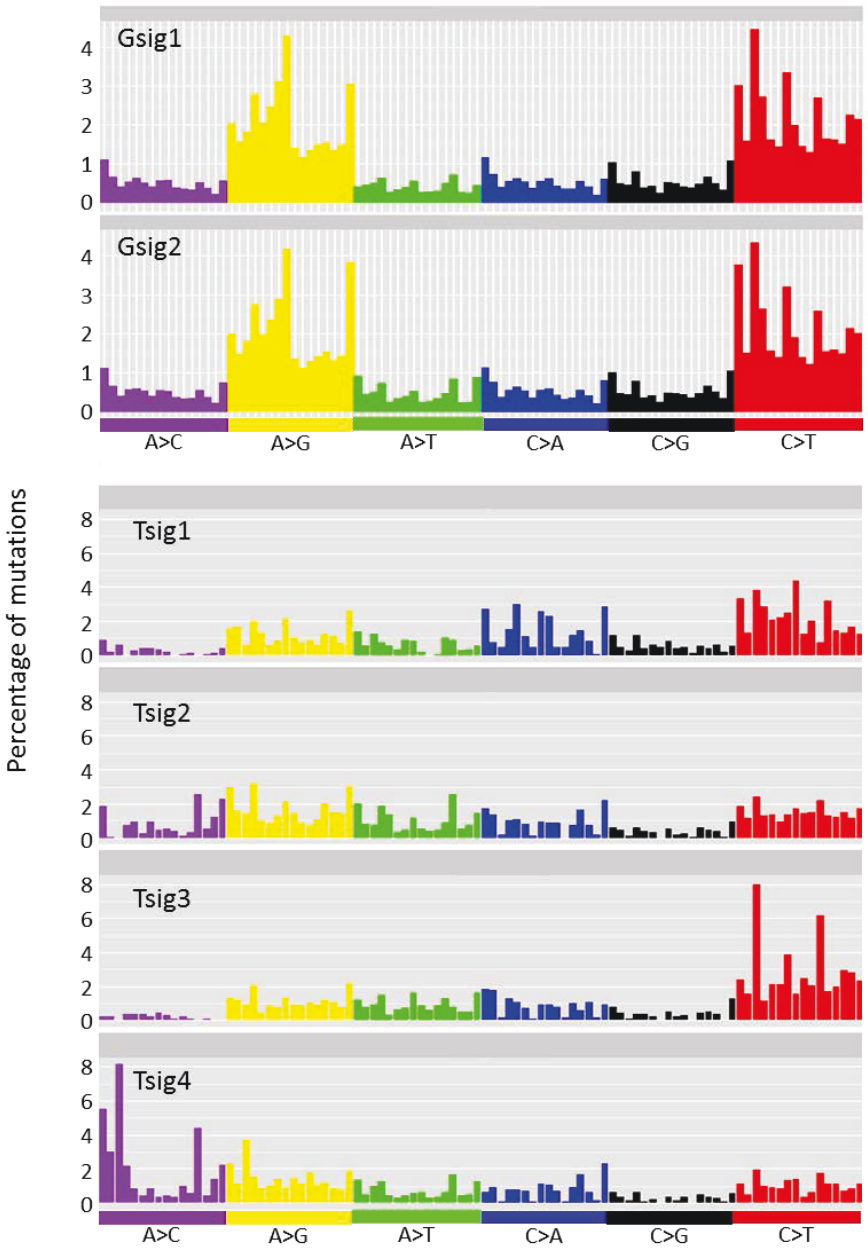

Figure 4.15 Germline and tumour mutational signatures detected in CLL patients.

Each signature is arranged according to the 96 possible substitution classifications, as defined by both the nucleotide substitution and the sequence context immediately 5' and 3' to the mutated position. Each substitution type is represented on the x-axis in a different colour, whereas the y-axis displays the percentage contribution of each mutation type to that particular signature.

Tsig4 is characterised by T>G transversions at both CpTpT and TpTpT trinucleotides. Tsig4 is the non-canonical AID signature (ncAID) associated with the activity of polymerase η during AID induced somatic hypermutation of the IgHV locus in CLL (Alexandrov *et al* 2013, Kasar *et al* 2015).

A high level of inter-patient diversity in the distribution of mutation signatures was also observed. For example, seven out of 22 (33%) chemo-sensitive cases only harboured three of the four signatures compared to 0 out of 21 chemo-refractory cases (Figure 4.16).

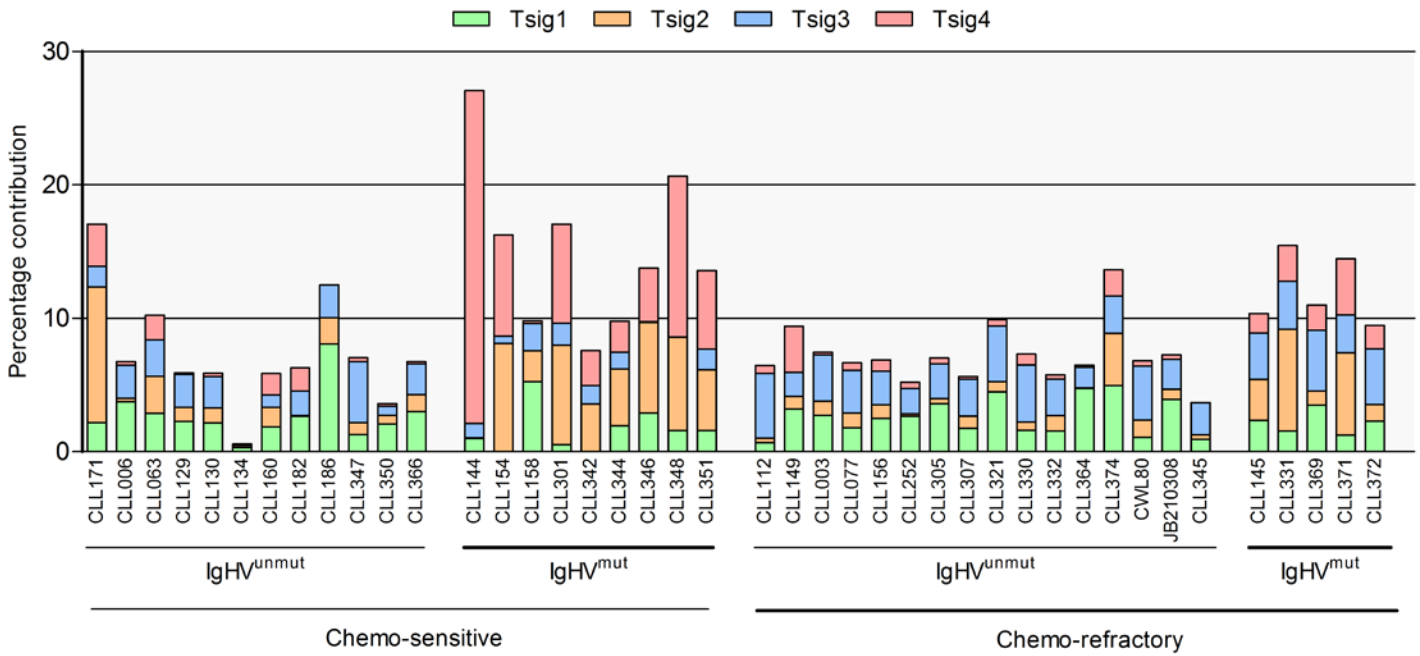

Figure 4.16 Contribution of different mutational signatures to the overall mutation burden in individual CLL patients.

Each bar represents a single patient, while the vertical axis shows the percentage of mutations contributed by each signature. Patients are divided first by sensitivity to chemotherapy, then by immunoglobulin mutation status.

4.4.6.2 CLL Germline Mutation Signatures

In contrast to the tumour samples, the matched germline samples harboured only two mutational signatures, Gsig1 and Gsig2 (Figure 4.15), present at similar levels in all 42 cases. Both signatures are closely matched in composition, being defined by a high incidence of C>T transitions in an NpCpG context, and, as expected, are related to the ageing signature found in the tumour genomes. Indeed, Gsig1 displays a strong correlation with the age of the patient (*P*=0.0015) reinforcing the idea that mutations in the germline genomes are acquired sporadically throughout the lifetime of the patient.

4.4.6.3 Correlation with Clinical and Biological Characteristics

Tsig1 showed no significant correlation with any clinical or biological characteristic. Tsig2 was found to correlate with higher patient age at the time of diagnosis in this cohort (*P*<0.001). The mutation pattern corresponding to Tsig3 has been found in more than 60% of cancer types, and has been also shown to correlate with the age of the patient at diagnosis (Alexandrov *et al* 2013). In this cohort, not only did it correlate with age (*P*=0.02), it was also more prevalent in the chemo-refractory patients compared to the treatment naïve cases (*P*<0.001) despite there being no significant difference in the ages of the two subgroups (*P*=0.9469).

The mutations ascribed to Tsig4 are a result of the underlying mechanism behind somatic hypermutation in B-cells, a relationship confirmed by the

fact that cases with a high Tsig4 signature also had hypermutated IgHV genes (*P*=0.0005) and an increased total mutation burden (*P*=0.0097) compared to those with low TSig4 (Figure 4.17). Interestingly, compared to chemo-refractory cases, chemo-sensitive patients showed a higher variability of Tsig4 mutational load, likely due to both the higher number of IgHV mutated patients in the untreated cohort (n=10) compared to the refractory (n=7) and the level of variance from the IgHV sequence, which is greater in the untreated cases (*P*<0.0001, F test).

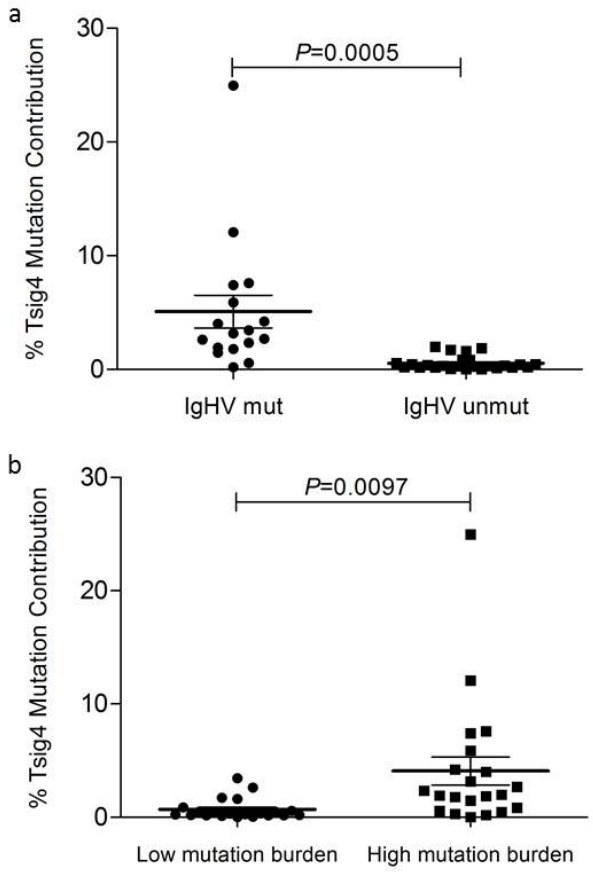

Figure 4.17 Tsig4 correlates with mutated IgHV genes and higher somatic mutation burden in CLL.

Dot plots describing the relationship between Tsig4 and (**a**) IgHV mutation status and (**b**) total mutation count. Each dot represents a single patient. Low mutation burden cases are those with a total mutation count lower than the median for the cohort (2,095). Error bars show ± SEM.

4.4.6.4 Canonical-AID Induced Mutations

The introduction of mutations by activation-induced cytidine deaminase (AID) is an ongoing process in CLL (Huemer *et al* 2014, Kasar *et al* 2015). In total, 3,524 canonical AID (cAID) mutations (C>T and G>A transitions in WRCY and RGYW motifs, respectively) were identified, accounting for 3.6%

of all mutations in the cohort. Forty-two per cent (1,476/3,524) of these occurred within the boundaries of a gene. As expected, there was a particular clustering of cAID mutations at the IgHV (n=178) locus, on chromosome 14 (Figure 4.18).

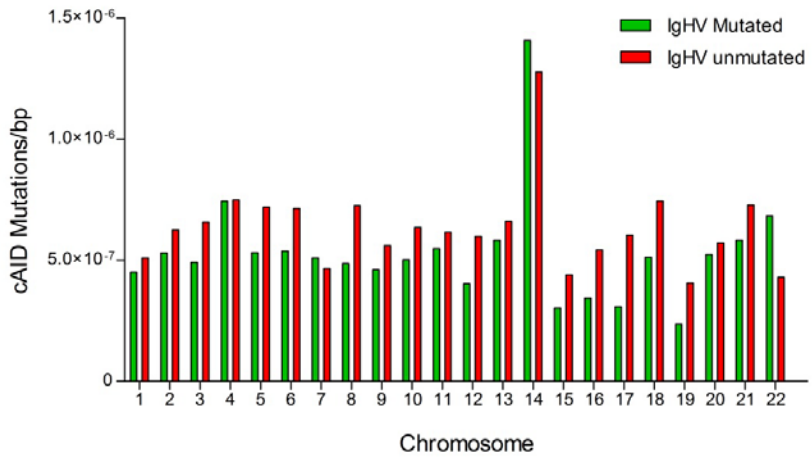

Figure 4.18 Occurrence of cAID mutations across the CLL genome.

Bar chart showing the enrichment of cAID mutations on chromosome 14 in both IgHV mutated and unmutated patients. Mutation counts are normalised to account for differing chromosome lengths.

Although cAID mutations associated with mutated IgHV status (*P*=0.0328), they were nevertheless also present across the genome and in unmutated patients and even patients with IgHV with 100% homology to the germline (Figure 4.19). There was no significant difference in the occurrence of cAID mutations between refractory and treatment naïve cases (*P*=0.6292). Interestingly, *IGLL5* contained 11 cAID mutations, accounting for 10% of all *IGLL5* variants.

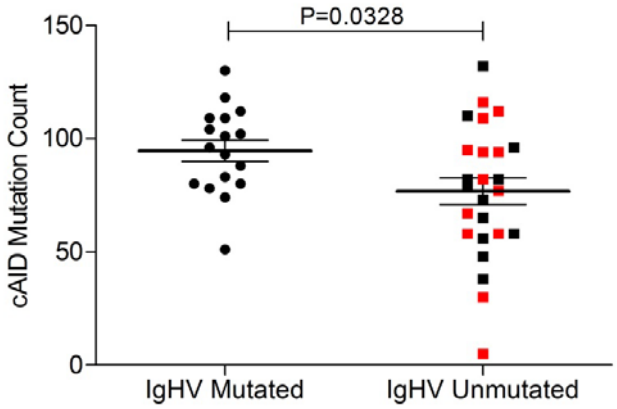

Figure 4.19 Distribution of cAID mutations between IgHV mutated and unmutated CLL patients.

Dot plot showing the absolute count of cAID mutations in the IgHV mutated and unmutated subgroups. Red data points represent patients with 100% homology to the IgHV germline sequence. Error bars show ± SEM.

4.5 Discussion

CLL is a disease that displays a high level of biological and clinical heterogeneity. The application of whole genome sequencing in cancer studies has enabled researchers to obtain reliable and accurate information on translocations and copy-number aberrations, as well as identifying mutations in non-coding regions of genes. It also facilitates the extraction of tumour-specific mutation signatures, which are frequently linked to particular causative agents. This study has sought to utilise whole-genome sequencing to provide a wide-ranging description of the CLL genome.

This study demonstrates the validity of whole-genome sequencing as an approach to characterise the mutational landscape of a cancer genome. The presence of coding mutations in the WGS data was confirmed by the targeted deep sequencing panel, which previous studies have shown is just as accurate and even more sensitive than Sanger sequencing (Fernandez-Mercado *et al* 2013). Additionally, the variant allele frequencies correlate well between these two sequencing protocols. It is important to note, however, that even if the sequencing chemistry works well, accurate variant detection also depends upon using the most appropriate downstream analysis pipeline. For example, the targeted deep sequencing panel highlighted four missense mutations, one in *MED12* and three in *TP53*, which were not initially identified by the WGS. Visual inspection of

the aligned WGS reads revealed that the mutations were present in the raw data but had not been picked up by the variant calling software.

In concordance with other studies, these data show that CLL has a relatively low global mutation rate compared to other cancer types (Kasar *et al* 2015, Landau *et al* 2013, Lawrence *et al* 2013, Schuh *et al* 2012). Furthermore, whilst there is a significant difference in both the coding and non-coding mutation count between the $IgHV^{mut}$ and $IgHV^{unmut}$ subgroups, there was no difference in the occurrence of mutations in known CLL driver genes. This suggests that oncogenesis and disease progression in CLL are linked to the presence of mutations in a small number of key genes. Alternatively, it might be that patients with high absolute mutation burden are protected from disease progression by the immune system. It has been shown that checkpoint inhibition is more effective in cancers with high mutational load (Hugo *et al* 2016).

Coding mutations in *ATM* are well characterised in CLL and, along with deletions of 11q, are known to confer shorter overall survival and increased risk of disease progression (Austen *et al* 2005, 2007, Nadeu, Delgado, Royo, Baumann, Stankovic, Pinyol, Jares, Navarro, Martin-Garcia, Bea, Salaverría, *et al* 2016, Skowronska *et al* 2012, Stankovic *et al* 1999). By using whole-genome sequencing, this study identified a region of kataegis, comprised of predominantly intronic mutations, in a single CLL patient, spanning introns 10 to 15 of the *ATM* gene. Although there was no evidence of an 11q deletion in this patient, the level of *ATM* transcript

expression was significantly reduced, and was comparable to that of the del(11q) subgroup. This evidence suggests that non-coding mutations may represent a new mechanism for *ATM* disruption in CLL.

By using complex mathematical modelling, it has become possible to isolate a number of mutational patterns, or signatures, within the molecular landscape of cancer (Alexandrov *et al* 2013, Morganella *et al* 2016, Nik-Zainal *et al* 2012, 2016). In this study, four such signatures were identified, each of which correlated well with those previously seen in CLL (Alexandrov *et al* 2013). In particular, Tsig3 correlates with both the age of the patient at diagnosis and chemo-refractoriness. A recent study also identified an on-going ageing signature in CLL that is associated with advanced disease and, interestingly, with sub-clonal mutations (Kasar *et al* 2015). Sub-clonal mutations in CLL have been shown to expand over time, often due to external selection pressures such as chemotherapy, and as a result contribute towards disease progression (Landau *et al* 2013). This supports the findings of this study that the ageing signature is more prevalent in refractory patients than in treatment naïve patients. Future work, including sub-clonal analysis on sequential patient samples, could offer additional insights to this.

An ncAID signature was also isolated from the tumour sequencing data, which is caused by the inefficient activity of DNA polymerase eta during mismatch repair following AID activity. The ncAID signature is correlated with both increased mutational burden and mutated IgHV status in this

cohort, although when comparing only IgHV^{mut} patients, there is enrichment for this signature in the treatment naïve samples. Kasar *et al* (2015) found that mutations attributed to the ncAID signature tended to be clonal, hypothesising that they occur early in tumour development, perhaps even prior to tumour initiation.

As part of the maturation process, healthy B-cells undergo SHM of the immunoglobulin (Ig) loci in the germinal centre. The activation-induced cytidine deaminase enzyme (AID) is responsible for inducing SHM. AID deaminates cytidine bases preferentially in a WRCY motif, converting them into uracils. If these changes are allowed to proceed to replication unrepaired, DNA polymerases insert an adenosine residue opposite the uracil, ultimately resulting in a C>T transition. These mutations are clustered at the Ig loci, although evidence shows a significant level of mutation (Kato *et al* 2012, Laura Pasqualucci *et al* 2008) and chromosomal disruption (Chiarle *et al* 2011, Hakim *et al* 2012, Robbiani *et al* 2008) in non-Ig genes. The present study reveals significant levels of AID activity in regions other than the Ig loci and shows that it also occurs in cases with 100% homology to the germline at the IgHV locus.

The majority of cancer sequencing studies have focussed on somatic mutations within the coding region of the genome. However, this only comprises 3% of the human genome. This study has shown that not only do the majority of somatically acquired mutations in CLL occur in non-coding regions, but that mutational hotspots occur in regulatory regions of genes.

Mutations in regulatory regions have been described in cancer, perhaps most notably in the promoter region of the *TERT* gene. These variants result in increased gene expression levels by introducing *de novo* ETS binding sites in a number of cancer types (Allory *et al* 2014, Brennan *et al* 2013, Horn *et al* 2013, Huang *et al* 2013, Liu *et al* 2013, Rachakonda *et al* 2013, Vinagre *et al* 2013, Wu *et al* 2014). Many other mutational hotspots, however, show no correlation between mutational status and transcript expression level (Fredriksson *et al* 2014). This finding is supported by these data in which, with the exception of *BCL2* and *ATM*, there was no correlation between regulatory region mutations and increased or decreased gene expression when compared to wild-type cases. However, the fact that a number of these regions occur in genes frequently affected by coding mutations in CLL suggests that these may not be random events.

There are a number of possible reasons for this lack of association. It may be that some mutations have a more subtle effect on gene function than can be seen at whole-transcriptome level. Alternatively it may be that these regions are simply more susceptible to mutations, perhaps due to inaccessibility of the DNA repair mechanisms and that these are events without functional consequences. On the other hand, it may be that the small sample size of this study is insufficient to generate a statistically significant value. It is also worth noting that much of the data generated by the ENCODE project is based on cell-line DNA, which may not be representative of genuine CLL DNA, potentially rendering the annotation of mutated regions as promoters or enhancers as incorrect, and thus making

direct comparisons difficult. Most likely, the lack of difference in gene expression between mutant and wild-type cases is due to the fact that in wild-type cases indirect mechanisms lead to changes in gene expression similar to the mutations identified. Future, large-scale sequencing efforts, such as that provided by the 100,000 genomes project, may provide a better insight into these non-coding variants and mechanistic studies need to be performed to understand the functional impact of these non-coding mutations.

In summary, this work demonstrates that global mutation signatures are present in CLL patients, with contrasting levels according to both IgHV mutation status and chemo-refractoriness. Further work is required to fully define the functional significance of these findings; however this study serves to highlight the heterogeneous nature of CLL, particularly with regards to the number of non-coding mutations present and, importantly, demonstrates that, with the exception of low VAF mutations, whole-genome sequencing provides SNV calls that are comparable to clinical grade targeted sequencing.

Chapter 5 DISCUSSION

5.1 Introduction

The advent of NGS has provided clinicians and researchers alike with a powerful new tool that promises to further our understanding of the molecular alterations underpinning the development and progression of cancer. Information regarding the mutational status of a number of key genes can be used to aid in diagnosis, predict a patient's response to treatment, or give an indication as to the likelihood of remission or relapse. Furthermore, whole genome sequencing can provide information on novel therapeutic targets, opening up new options for patients and clinicians in the future. With the work presented in this thesis, I demonstrate firstly that targeted NGS is a reliable, more sensitive alternative to established mutation screening techniques and that it offers genuine clinical value when offered as a routine service in a diagnostic laboratory. Furthermore, targeted NGS can be used as a research tool to characterise a small section of the mutational landscape of a particular cohort of patients. Secondly, I show that whole-genome sequencing can provide a comprehensive analysis of the CLL genome, revealing the heterogeneous nature of the molecular landscape, including the presence of specific mutational signatures, structural alterations and localised regions of increased mutational frequency.

5.2 Targeted Next-Generation Sequencing as a Diagnostic Tool

AML is the most common myeloid malignancy in Europe (Visser *et al* 2012). The molecular landscape of AML is somewhat complex, with both structural alterations and gene mutations providing important clinical information at the time of diagnosis. Indeed, at the time of writing, the WHO guidelines for classifying AML and related neoplasms are predominantly concerned with the presence of specific translocations (Grimwade 2012, Swerdlow *et al* 2008), despite the fact that 40% of cases possess a normal karyotype (Grimwade *et al* 2010). Recurrently mutated genes have been identified within the wider group of myeloid disorders (AML, MDS and MPN), a number of which provide important clinical information at the time of diagnosis. For example, mutations in *NPM1* increases the efficacy of treatment with either Bortezomib or arsenic trioxide in AML patients (M Huang *et al* 2013), leading to higher complete remission rates and longer overall survival, but only in the absence of concomitant mutations in the *FLT3* gene (K Döhner *et al* 2005, Hollink *et al* 2009, Schlenk *et al* 2008, Christian Thiede *et al* 2006), information that requires screening of both genes. The importance of these recent findings is reflected in the fact that the WHO classification for AML and related neoplasms includes cases with either *NPM1* or *CEBP*α mutations as distinct groups (Swerdlow *et al* 2008). There is, therefore, a case to be made for a diagnostic test targeting several genes simultaneously, in order to provide

as clear a characterisation as possible of the mutational landscape of each individual patient. The work presented in chapter 3 demonstrates that a targeted sequencing assay can reliably detect mutations with clinical implications with a greater degree of sensitivity than is obtainable via existing methods, including Sanger sequencing.

5.2.1 A Higher Degree of Sensitivity Offers Additional Clinical Insights

As chapter three shows, this TSCA assay is able to detect mutations present in as few as 3% of the sequencing reads, this being an marked improvement on the 10-20% limit of detection offered by Sanger sequencing (Tsiatis *et al* 2010). As a result, targeted NGS is better suited to detecting sub-clonal mutations within a cancer cell population. A number of recent studies in CLL, for example, have shown that low level, sub-clonal mutations can have the same impact on disease progression as high frequency clonal mutations, predicting for shorter progression free survival (Landau *et al* 2013, Nadeu, Delgado, Royo, Baumann, Stankovic, Pinyol, Jares, Navarro, Martin-Garcia, Bea, Salaverria, *et al* 2016) and the development of chemo-refractoriness (Rossi *et al* 2014).

Furthermore, the ongoing risk of relapse following initial treatment can be routinely assessed via the continuous monitoring of disease specific, low-level mutations (Grimwade and Freeman 2014, Ivey *et al* 2016). This

minimal residual disease (MRD) monitoring currently employs modified quantitative real-time PCR assays to accurately quantify the level of mutation present in a post-treatment sample (Gorello *et al* 2006, Mancini *et al* 2015, Warren *et al* 2012) with sensitivity as low as 2.5% (Shivarov *et al* 2014). Given that NGS generates a large number of clonal DNA molecules, making it possible to quantify the allelic ratio of a given variant, combined with the sensitivity and low background noise levels of the targeted panel described in chapter 3, it is possible that this is another clinical application. Indeed, in the time since I developed this panel, a number of studies have applied NGS to the question of MRD (Bibault *et al* 2015, Spencer *et al* 2013, Thol *et al* 2012), showing good concordance with existing methods (Bibault *et al* 2015).

5.2.2 Alternative Target Enrichment Strategies

The TruSeq Custom amplicon system used in chapter three is only one of a number of target enrichment strategies currently on the market, which differ in both enrichment method and sequencing performance. There are presently two main methods for target enrichment, those that employ hybridisation of DNA probes to isolate the region(s) of interest from a whole genome preparation, such as the SureSelect^{XT} and Haloplex HS assays (both Agilent Technologies), and those that use PCR primers to amplify the chosen regions, including the TSCA assay (Illumina, San Diego, CA, USA), the Ion AmpliSeq assay (Thermo Fisher Scientific, Waltham, MA, USA) and the SeqCap EZ system (Roche Diagnostics Ltd., Basel,

Switzerland). Comparisons of the different methods, in the context of whole exome capture, show that hybridisation-based sequencing methods produce libraries with greater complexity than amplicon enrichment (Samorodnitsky, Jewell, *et al* 2015). Amplicon enrichment produces greater average sequencing coverage across the exome, along with greater standard deviation of coverage, suggesting that hybridisation techniques result in better sequencing uniformity (Samorodnitsky, Jewell, *et al* 2015). The level of G/C content in a given target region can also impact upon the efficiency of the enrichment method used as both high (<60%) and low (<40%) levels cause reduced amplification and, therefore, sequencing depth (Asan *et al* 2011, Clark *et al* 2011, Samorodnitsky, Datta, *et al* 2015). This is an important consideration when designing diagnostic panels in order to generate sufficient sequencing depth across clinically important genes with a high GC content, including *CEBPa* (Ahn *et al* 2016, Wouters *et al* 2009), which is one of the markers included in the WHO classification guidelines for AML cases (Swerdlow *et al* 2008).

5.2.3 World Health Organisation 2016 Classification

With the publication of the 2008 World Health Organisation guide for the classification of AML and related neoplasms, came the introduction of two provisional entries regarding the presence of specific mutations, namely those with mutated *NPM1* and those with mutated *CEBPa*. In the time since these guidelines were published, studies have shed new light on the

diagnostic and prognostic information provided by existing and new molecular markers. To reflect this, the 2016 update to the WHO classification guidelines has refined and confirmed these groupings (Arber *et al* 2016). AML with mutated *NPM1* has been upgraded from provisional to a full category. The finding that bi-allelic, but not mono-allelic, *CEBPa* mutations offer improved rates of OS, EFS and reduced risk of relapse (Ahn *et al* 2016, Wouters *et al* 2009), has resulted in a refinement of this category to only include bi-allelic mutated patients. A new provisional entity of AML with mutated *RUNX1* has also been added to the system, since AML cases with mutated *RUNX1* has been shown to have a worse prognosis (Gaidzik *et al* 2011, Mendler *et al* 2012, Schnittger *et al* 2011, Tang *et al* 2009).

5.2.4 Implementation of Next-Generation Sequencing in a Diagnostic Setting

In the years since the chemistries behind next-generation sequencing were first described (Bentley *et al* 2008, Margulies *et al* 2005), NGS has been used to characterise the molecular landscape of a number of diseases, including cancer. This has led to the discovery of new diagnostic and prognostic markers, and new therapeutic targets, with the ultimate goal being the provision of personalised, or precision, treatment on a case-by-case basis. With the work described in chapter three, I demonstrated that a next-generation targeted sequencing panel provides the means to generate accurate and sensitive information regarding the mutational

status of multiple genes simultaneously, and that this same information is influencing the diagnosis, treatment and follow-up response decisions in a diagnostic laboratory.

Targeted sequencing panels offer a number of advantages over a whole genome sequencing approach, particularly in the specific context of a diagnostic laboratory. The very nature of targeted sequencing means that the region(s) included in the panel can be very tightly regulated by the laboratory, with only those with a confirmed relationship between the presence of a mutation and pathology being included and reported on. Other advantages of targeted include the ability to multiplex samples together, a smaller DNA input requirement and a faster turnaround time than WGS. Targeted sequencing is also much cheaper than WGS, costing £300 for the myeloid panel described in chapter three, compared to £4,000 for a single whole genome.

It is also worth mentioning the question of ethics. As the targets of a sequencing panel are very tightly defined, the question of incidental findings is greatly reduced and, given the tightly controlled scope of the test, consent is usually easily obtained from the patient in clinic. In contrast, suppose a sample from a patient with AML is sent to a diagnostic laboratory for whole genome sequencing, with the goal of screening for mutations in *NPM1*. In this instance, it is possible, however unlikely, that a variant linked to an increased risk of developing another disease may be

identified. Alternatively, a heterozygous mutation in a recessive condition may be detected, which may impact upon the patient's reproductive choices. Indeed, the European Society of Human Genetics (ESHG) has recommended the use of targeted sequencing panels over WGS in order to avoid these incidental findings (van El *et al* 2013).

Of course, the absence of gene mutations can also offer important diagnostic information to a clinician. Of the 270 diagnostic cases screened with the myeloid sequencing panel in chapter three, 64 were either sent to the laboratory as an unexplained anaemia or as a suspected case of AML, MDS or MPD. In these instances, the panel is being used when a clinician is unsure of the condition, and wishes to obtain additional evidence before making any clinical decision.

5.3 Characterisation of the CLL Genome

Chronic lymphocytic leukaemia is characterised by a high level of both clinical and biological heterogeneity, with some patients requiring no treatment at all, while others develop resistance to chemotherapy and, therefore, disease progression. This heterogeneity may be explained by the diversity of mutations and structural variants present within the CLL genome. The work presented in chapter four sought to leverage the power of next-generation whole genome sequencing in an attempt to provide a comprehensive characterisation of a representative cohort of clinical CLL

patients. Furthermore, it sought to demonstrate the utility of WGS in a clinical setting.

5.3.1 Appropriate Germline DNA Source Selection for Subtraction Analysis

Cancer is the result of de-regulated clonal expansion of a single cell. This de-regulation is generally linked to one or more somatically acquired genetic alteration, including gene mutations and/or chromosomal alterations. In order to better understand the methods by which cancer develops, and to be able to predict the future disease course, including how a patient will respond to a particular therapy, it is important to be able to identify these mutations. Complicating the matter is the presence of germline mutations in the patient's genome. These inherited changes are natural variations from the reference human genome and, as such, will appear as false positives in the cancer sample following DNA sequencing. One method of differentiating these from somatically acquired mutations is to sequence both germline and tumour DNA samples from a single patient, removing any mutations present in the germline DNA from those in the tumour DNA. Those that remain are considered somatically acquired. This approach is dependent on the germline DNA source containing no tumour cells, since this would cause the removal of tumour specific mutations during the subtraction analysis. Therefore the choice of germline

material is critical in obtaining an accurate picture of the tumour mutational landscape.

In chapter four, WGS was used to isolate somatically acquired mutations in 42 CLL genomes. The DNA used as a germline sample for each patient was extracted from buccal swabs, a method which has been used in a number of previous whole genome and exome studies in CLL (Kasar *et al* 2015, Ljungström *et al* 2016, Schuh *et al* 2012). However, buccal swabs and saliva samples are known to contain high levels of haematopoietic cells (Chaudhary *et al* 2015, Endler *et al* 1999, C Thiede *et al* 2000), raising the issue of tumour contamination in the germline DNA. Indeed, in the present study, one patient was shown to harbour only 153 somatic mutations, well below the median for this cohort, none of which affecting protein coding sites. It is possible that the germline DNA sample contained a high level of tumour cell contamination in this instance.

It has also been shown that multiple buccal swabs from the same patient can vary in content. One study investigating *DNMT3a* mutations in AML analysed two swabs from one individual, one of which was found to contain the mutation with the other being wild-type (Ewalt *et al* 2011). Other recent whole genome studies in CLL isolated B-cells from T-cells based on specific cell surface markers using flow cytometry (Landau *et al* 2013, Puente *et al* 2011). The T-cell population, being from a different cell lineage, are then used as a source of germline DNA for the CLL B-cells.

However, this approach also has limitations. It is based upon the assumption that the cancer initiating mutation arises once the cell lineage fate has been decided. A number of studies have demonstrated the presence of tumorigenic mutations in haematopoietic stem cell populations (Mian *et al* 2015, Shlush *et al* 2014), a fact that indicates mutations may be present in all haematopoietic cell lineages.

5.3.2 Determination of the Function and Clinical Impact of Non-Coding Mutations in Cancer

To date, the majority of cancer sequencing studies have focussed on somatic mutations within the coding region of the genome. However, this only comprises 3% of the human genome. The data presented in chapter four demonstrates that not only do the majority of somatically acquired mutations in CLL occur in non-coding regions, but that mutational hotspots occur in regions that regulate the expression of genes. The likely function of such regions can be determined using methods including DNase I hypersensitivity site sequencing (DNase-Seq) or chromatin immunoprecipitation sequencing (ChIP-Seq). The presence of DNase I sites are understood to be an indicator of open and accessible chromatin and, therefore, active genes. ChIP-Seq is used to identify chromatin modifications, specifically the methylation and acetylation status of key histones, which are hallmarks of promoter and enhancer regions (Hon *et al* 2009). The genomic targets of such regions can be determined using a

chromosome conformation capture-on-chip sequencing (4C-seq) approach (van de Werken *et al* 2012) to characterise the physical interactions between regulatory regions and genomic targets. Indeed, in one recent large-scale sequencing study in CLL, the authors identified a densely mutated region on chromosome 9p13 (Puente *et al* 2015). The presence of a DNase I hypersensitive site, along with nucleosomes containing both H3K4me1 and H3K27ac, indicated this was the site of an active enhancer. Using a combination of gene expression arrays and 4C-seq, they identified *PAX5* as the target of the enhancer and found that expression levels of *PAX5* were altered in mutated patients (Puente *et al* 2015). Furthermore, the study showed that the introduction of the same mutations in B-cell cell lines resulted in the same reduction in *PAX5* expression levels, further confirming the enhancer function of this region. The work presented in chapter four identified clusters of mutations in non-coding regions of genes, in particular within the first intron of *BCL6*. It would be interesting to further investigate the functional and clinical consequence of these mutations using 4C-Seq and RNA-Seq in a larger patient cohort.

5.4 Future Work and Directions

The work presented in this thesis demonstrates that both targeted and whole genome sequencing could be introduced into the diagnostic setting in order to aid with the clinical decision making process. However, the

findings here, and recent developments in the field of DNA sequencing, have raised further avenues for potential investigation.

DNA sequencing using modified nanopores is able to generate sequencing reads of up to 50,000bp (Ip *et al* 2015) in comparison to the 300bp reads obtainable through current Illumina SBS methods. This increase in read length brings with it opportunities to detect genomic alterations larger than single nucleotide changes or short insertion/deletion events.

It has been shown that CLL patients with hypermutated IgHV (<98% homology to the germline) have significantly increased OS rates compared to un-mutated patients (Damle *et al* 1999, Hamblin *et al* 1999) and, as such, screening for IgHV mutation status is an important part of the diagnostic work-up in new CLL patients. This screening is predominantly performed using a multiplex PCR followed by Sanger sequencing, an approach which, with its limited sensitivity, results in only the dominant IgHV clone being detected. Furthermore, the failure rate for this assay ranges from 8-18% (Austen *et al* 2005, Burger *et al* 2014, Kröber *et al* 2002). Recent work using NGS methods have shown that 24% of CLL patients harbour at least 2 different IgHV rearrangements, the specific profile of which can improve risk stratification (Stamatopoulos *et al* 2016).

It would be useful, therefore, to be able to integrate both IgHV mutation testing with screening for coding mutations in clinically relevant genes. The long read lengths, and PCR-free protocols afforded by nanopore sequencing make this an ideal approach for this.

REFERENCES

Abaan OD, Polley EC, Davis SR, Zhu YJ, Bilke S, Walker RL, Pineda M, Gindin Y, Jiang Y, Reinhold WC, Holbeck SL, Simon RM, Doroshow JH, Pommier Y and Meltzer PS (2013) The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer research* 73(14): 4372–82.

Abbas S, Erpelinck-Verschueren CAJ, Goudswaard CS, Löwenberg B and Valk PJM (2010) Mutant Wilms' tumor 1 (WT1) mRNA with premature termination codons in acute myeloid leukemia (AML) is sensitive to nonsense-mediated RNA decay (NMD). *Leukemia* 24(3): 660–3.

Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A, Beran M, Gilliland DG, Levine RL and Tefferi A (2011) Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia* 25(7): 1200–2.

Abu-Duhier FM, Goodeve AC, Wilson GA, Care RS, Peake IR and Reilly JT (2001) Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *British journal of haematology* 113(4): 983–8.

Adès L, Guerci A, Raffoux E, Sanz M, Chevallier P, Lapusan S, Recher C, Thomas X, Rayon C, Castaigne S, Tournilhac O, de Botton S, Ifrah N, Cahn J-Y, Solary E, Gardin C, Fegeux N, Bordessoule D, Ferrant A, Meyer-Monard S, Vey N, Dombret H, Degos L, Chevret S and Fenaux P (2010) Very long-term outcome of acute promyelocytic leukemia after treatment with all-trans retinoic acid and chemotherapy: the European APL Group experience. *Blood* 115(9): 1690–6.

Ahn J-S, Kim J-Y, Kim H-J, Kim Y-K, Lee S-S, Jung S-H, Yang D-H, Lee J-J, Kim NY, Choi SH, Minden MD, Jung CW, Jang J-H, Kim HJ, Moon JH, Sohn SK, Won J-H, Kim S-H and Kim DDH (2016) Normal karyotype acute myeloid leukemia patients with CEBPA double mutation have a favorable prognosis but no survival benefit from allogeneic stem cell transplant. *Annals of hematology* 95(2): 301–10.

Alcalay M, Meani N, Gelmetti V, Fantozzi A, Fagioli M, Orleth A, Riganelli D, Sebastiani C, Cappelli E, Casciari C, Sciurpi MT, Mariano AR, Minardi SP, Luzi L, Muller H, Di Fiore PP, Frosina G and Pelicci PG (2003) Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *The Journal of clinical investigation* 112(11): 1751–61.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk

M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson J V, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ and Stratton MR (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463): 415–21.

Allory Y, Beukers W, Sagrera A, Flández M, Marqués M, Márquez M, van der Keur KA, Dyrskjot L, Lurkin I, Vermeij M, Carrato A, Lloreta J, Lorente JA, Carrillo-de Santa Pau E, Masius RG, Kogevinas M, Steyerberg EW, van Tilborg AAG, Abas C, Orntoft TF, Zuiverloon TCM, Malats N, Zwarthoff EC and Real FX (2014) Telomerase reverse transcriptase promoter mutations in bladder cancer: high frequency across stages, detection in urine, and lack of association with outcome. *European urology* 65(2): 360–6.

Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD and Chang DD (2008) Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26(10): 1626–34.

Aranaz P, Hurtado C, Erquiaga I, Migueliz I, Ormazabal C, Cristobal I, Garcia-Delgado M, Novo FJ and Vizmanos JL (2012) CBL mutations in myeloproliferative neoplasms are also found in its proline-rich domain and in patients with the V617FJAK2 (2012/02/09.). *Haematologica* 97(8): 1234–1241.

Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M and Vardiman JW (2016) *The 2016 revision to the World Health Organization (WHO) classification of lymphoid neoplasms. Blood*.

Asan, Xu Y, Jiang H, Tyler-Smith C, Xue Y, Jiang T, Wang JJJ, Wu M, Liu X, Tian G, Yang H, Zhang X, Wang JJJ, Wang JJJ, Yang H and Zhang X (2011) Comprehensive comparison of three commercial human whole-exome capture platforms (2011/10/01.). *Genome biology* 12(9): R95.

Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, Taylor AMR, Fegan C, Moss P and Stankovic T (2005) Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood*. American Society of Hematology 106(9): 3175–82.

Austen B, Skowronska A, Baker C, Powell JE, Gardiner A, Oscier D, Majid A, Dyer M, Siebert R, Taylor AM, Moss PA and Stankovic T (2007) Mutation status of the residual ATM allele is an important determinant of the cellular

response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 25(34): 5448–57.

Balatti V, Bottoni A, Palamarchuk A, Alder H, Rassenti LZ, Kipps TJ, Pekarsky Y and Croce CM (2012) NOTCH1 mutations in CLL associated with trisomy 12. *Blood*. American Society of Hematology 119(2): 329–31.

Baliakas P, Hadzidimitriou A, Sutton L-A, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, Cortese D, Navrkalova V, Rose-Zerilli MJJ, Smedby KE, Juliusson G, Anagnostopoulos A, Makris AM, Navarro A, Delgado J, Oscier D, Belessi C, Stilgenbauer S, Ghia P, Pospisilova S, Gaidano G, Campo E, Strefford JC, Stamatopoulos K and Rosenquist R (2014) Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*.

Baumann T, Delgado J, Santacruz R, Martínez-Trillos A, Rozman M, Aymerich M, López C, Costa D, Carrió A, Villamor N and Montserrat E (2016) CD49d (ITGA4) expression is a predictor of time to first treatment in patients with chronic lymphocytic leukaemia and mutated IGHV status. *British journal of haematology* 172(1): 48–55.

Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, Scott MA, Erber WN, Green AR and Cancer Genome Project (2005) Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *The Lancet* 365(9464): 1054–1061.

Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D, Whitman SP, Paschka P, Holland KB, Schwind S, Wu Y-Z, Powell BL, Carter TH, Kolitz JE, Wetzler M, Carroll AJ, Baer MR, Moore JO, Caligiuri MA, Larson RA and Bloomfield CD (2010) Mutations of the Wilms tumor 1 gene (WT1) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 116(5): 788–92.

Beer PA, Campbell PJ, Scott LM, Bench AJ, Erber WN, Bareford D, Wilkins BS, Reilly JT, Hasselbalch HC, Bowman R, Wheatley K, Buck G, Harrison CN and Green AR (2008) MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. *Blood* 112(1): 141–9.

Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, Kantarjian H, Raza A, Levine RL, Neuberg D and Ebert BL (2011) Clinical effect of point mutations in myelodysplastic syndromes. *The New England journal of medicine* 364(26): 2496–506.

Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR and Sultan C (1976) Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *British journal of haematology* 33(4): 451–8.

Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR and Sultan C (1985) Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Annals of internal medicine*. American College of Physicians 103(4): 620.

Bentley DR et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry (2008/11/07.). *Nature* 456(7218): 53–59.

Bibault J-E, Figeac M, Hélevaut N, Rodriguez C, Quief S, Sebda S, Renneville A, Nibourel O, Rousselot P, Gruson B, Dombret H, Castaigne S and Preudhomme C (2015) Next-generation sequencing of FLT3 internal tandem duplications for minimal residual disease monitoring in acute myeloid leukemia. *Oncotarget*. Impact Journals, LLC 6(26): 22812–21.

Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Jacquillat C, Boivin P, Lesty C, Duault MT, Monconduit M, Belabbes S and Gremy F (1981) A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* 48(1): 198–206.

Binet JL, Lepoprier M, Dighiero G, Charron D, D'Athis P, Vaugier G, Beral HM, Natali JC, Raphael M, Nizet B, Follezou JY, Leporrier M, Dighiero G, Charron D, Vaugier G, Beral HM, Natali JC, Raphael M, Nizet B and Follezou JY (1977) A clinical staging system for chronic lymphocytic leukemia.Prognostic significance. *Cancer* 40(2): 855–64.

Bloomfield CD, Lawrence D, Byrd JC, Carroll A, Pettenati MJ, Tantravahi R, Patil SR, Davey FR, Berg DT, Schiffer CA, Arthur DC and Mayer RJ (1998) Frequency of prolonged remission duration after high-dose cytarabine intensification in acute myeloid leukemia varies by cytogenetic subtype. *Cancer research* 58(18): 4173–9.

Bo MD, Del Principe MI, Pozzo F, Ragusa D, Bulian P, Rossi D, Capelli G, Rossi FM, Niscola P, Buccisano F, Bomben R, Zucchetto A, Maurillo L, de Fabritiis P, Amadori S, Gaidano G, Gattei V and Del Poeta G (2014) NOTCH1 mutations identify a chronic lymphocytic leukemia patient subset with worse prognosis in the setting of a rituximab-based induction and consolidation treatment. *Annals of hematology* 93(10): 1765–74.

Bolli N, Nicoletti I, De Marco MF, Bigerna B, Pucciarini A, Mannucci R, Martelli MP, Liso A, Mecucci C, Fabbiano F, Martelli MF, Henderson BR and Falini B (2007) Born to be exported: COOH-terminal nuclear export signals of different strength ensure cytoplasmic accumulation of nucleophosmin leukemic mutants. *Cancer research* 67(13): 6230–7.

Bolton L, Reiman A, Lucas K, Timms J and Cree IA (2015) KRAS mutation analysis by PCR: a comparison of two methods. *PloS one* 10(1): e0115672.

Borthakur G, Kantarjian H, Wang X, Plunkett WK, Gandhi V V, Faderl S, Garcia-Manero G, Ravandi F, Pierce S and Estey EH (2008) Treatment of

core-binding-factor in acute myelogenous leukemia with fludarabine, cytarabine, and granulocyte colony-stimulating factor results in improved event-free survival. *Cancer* 113(11): 3181–5.

Bosch F, Abrisqueta P, Villamor N, Terol MJ, González-Barca E, Ferra C, González Diaz M, Abella E, Delgado J, Carbonell F, García Marco JA, Escoda L, Ferrer S, Monzó E, González Y, Estany C, Jarque I, Salamero O, Muntañola A and Montserrat E (2009) Rituximab, fludarabine, cyclophosphamide, and mitoxantrone: a new, highly active chemoimmunotherapy regimen for chronic lymphocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 27(27): 4578–84.

Bosch F, Ferrer A, Villamor N, González M, Briones J, González-Barca E, Abella E, Gardella S, Escoda L, Pérez-Ceballos E, Asensi A, Sayas MJ, Font L, Altés A, Muntañola A, Bertazzoni P, Rozman M, Aymerich M, Giné E and Montserrat E (2008) Fludarabine, cyclophosphamide, and mitoxantrone as initial therapy of chronic lymphocytic leukemia: high response rate and disease eradication. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(1): 155–61.

Boultwood J, Perry J, Pellagatti A, Fernandez-Mercado M, Fernandez-Santamaria C, Calasanz MJ, Larrayoz MJ, Garcia-Delgado M, Giagounidis A, Malcovati L, Della Porta MG, Jädersten M, Killick S, Hellström-Lindberg E, Cazzola M and Wainscoat JS (2010) Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia* 24(5): 1062–5.

Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu C-J, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner DD, Van Meir EG, Prados M, Sloan A, Black KL, Eschbacher J, Finocchiaro G, Friedman W, Andrews DW, Guha A, Iacocca M, O'Neill BP, Foltz G, Myers J, Weisenberger DJ, Penny R, Kucherlapati R, Perou CM, Hayes DN, Gibbs R, Marra M, Mills GB, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird PW, Haussler D, Getz G, Chin L and TCGA Research Network (2013) The somatic genomic landscape of glioblastoma. *Cell* 155(2): 462–77.

Brown JR, Hanna M, Tesar B, Werner L, Pochet N, Asara JM, Wang YE, Dal Cin P, Fernandes SM, Thompson C, Macconaill L, Wu CJ, Van de Peer Y, Correll M, Regev A, Neuberg D and Freedman AS (2012) Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18(14): 3791–802.

Brunet J-P, Tamayo P, Golub TR and Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences 101(12): 4164–9.

Burger JA, Keating MJ, Wierda WG, Hartmann E, Hoellenriegel J, Rosin NY, de Weerdt I, Jeyakumar G, Ferrajoli A, Cardenas-Turanzas M, Lerner S, Jorgensen JL, Nogueras-González GM, Zacharian G, Huang X, Kantarjian H, Garg N, Rosenwald A and O'Brien S (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): 1090–9.

Byrd JC, Furman RR, Coutre SE, Burger JA, Blum KA, Coleman M, Wierda WG, Jones JA, Zhao W, Heerema NA, Johnson AJ, Shaw Y, Bilotti E, Zhou C, James DF and O'Brien S (2015) Three-year follow-up of treatment-naïve and previously treated patients with CLL and SLL receiving single-agent ibrutinib. *Blood*. American Society of Hematology 125(16): 2497–506.

Byrd JC, Gribben JG, Peterson BL, Grever MR, Lozanski G, Lucas DM, Lampson B, Larson RA, Caligiuri MA and Heerema NA (2006) Select high-risk genetic features predict earlier progression following chemoimmunotherapy with fludarabine and rituximab in chronic lymphocytic leukemia: justification for risk-adapted therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 24(3): 437–43.

Byrd JC, Ruppert AS, Mrózek K, Carroll AJ, Edwards CG, Arthur DC, Pettenati MJ, Stamberg J, Koduru PRK, Moore JO, Mayer RJ, Davey FR, Larson RA and Bloomfield CD (2004) Repetitive cycles of high-dose cytarabine benefit patients with acute myeloid leukemia and inv(16)(p13q22) or t(16;16)(p13;q22): results from CALGB 8461. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 22(6): 1087–94.

Calpe E, Purroy N, Carpio C, Abrisqueta P, Carabia J, Palacio C, Castellví J, Crespo M and Bosch F (2013) ZAP-70 promotes the infiltration of malignant B-lymphocytes into the bone marrow by enhancing signaling and migration after CXCR4 stimulation. *PloS one* 8(12): e81221.

Cancer Genome Atlas Research Network TCGAR (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine*. NIH Public Access 368(22): 2059–74.

Capello D, Fais F, Vivenza D, Migliaretti G, Chiorazzi N, Gaidano G and Ferrarini M (2000) Identification of three subgroups of B cell chronic lymphocytic leukemia based upon mutations of BCL-6 and IgV genes. *Leukemia*. Nature Publishing Group 14(5): 811–815.

Castaigne S, Chomienne C, Daniel MT, Ballerini P, Berger R, Fenaux P and Degos L (1990) All-trans retinoic acid as a differentiation therapy for acute promyelocytic leukemia. I. Clinical results. *Blood* 76(9): 1704–9.

Chandra P, Luthra R, Zuo Z, Yao H, Ravandi F, Reddy N, Garcia-Manero G, Kantarjian H and Jones D (2010) Acute myeloid leukemia with t(9;11)(p21-22;q23): common properties of dysregulated ras pathway signaling and genomic progression characterize de novo and therapy-related cases. *American journal of clinical pathology* 133(5): 686–93.

Chapiro E, Leporrier N, Radford-Weiss I, Bastard C, Mossafa H, Leroux D, Tigaud I, De Braekeleer M, Terré C, Brizard F, Callet-Bauchu E, Struski S, Veronese L, Fert-Ferrer S, Taviaux S, Lesty C, Davi F, Merle-Béral H, Bernard OA, Sutton L, Raynaud SD and Nguyen-Khac F (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): 63–8.

Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet J-P, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, Onofrio R, Pugh TJ, Rajkumar SV, Ramos AH, Siegel DS, Sivachenko A, Stewart AK, Trudel S, Vij R, Voet D, Winckler W, Zimmerman T, Carpten J, Trent J, Hahn WC, Garraway LA, Meyerson M, Lander ES, Getz G and Golub TR (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* 471(7339): 467–72.

Chaudhary G, Dogra TD and Raina A (2015) Evaluation of blood, buccal swabs, and hair follicles for DNA profiling technique using STR markers. *Croatian medical journal*. Medicinska Naklada 56(3): 239–45.

Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho Y-J, Myers DR, Choi VW, Compagno M, Malkin DJ, Neuberg D, Monti S, Giallourakis CC, Gostissa M and Alt FW (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147(1): 107–19.

Chigrinova E, Rinaldi A, Kwee I, Rossi D, Rancoita PM V, Strefford JC, Oscier D, Stamatopoulos K, Papadaki T, Berger F, Young KH, Murray F, Rosenquist R, Greiner TC, Chan WC, Orlandi EM, Lucioni M, Marasca R, Inghirami G, Ladetto M, Forconi F, Cogliatti S, Votavova H, Swerdlow SH, Stilgenbauer S, Piris MA, Matolcsy A, Spagnolo D, Nikitin E, Zamò A, Gattei V, Bhagat G, Ott G, Zucca E, Gaidano G and Bertoni F (2013) Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood*. American Society of Hematology 122(15): 2673–82.

Chou W-C, Hou HA, Chen C-Y, Tang J-L, Yao M, Tsay W, Ko B-S, Wu S-J, Huang S-Y, Hsu S-C, Chen Y-C, Huang Y-N, Chang Y-C, Lee F-Y, Liu M-C, Liu C-W, Tseng M-H, Huang C-F and Tien H-F (2010) Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood* 115(14): 2749–54.

Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, Butte AJ and Snyder M (2011) Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology* 29(10): 908–914.

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 Y-L, Pasero P, Maelfait

J, Titsias M, Cohen DR, Henderson SJ, Ross MT, Bentley D, Hillmen P, Pettitt A, Rehwinkel J, Knight SJL, Taylor JC, Crow YJ, Benkirane M and Schuh A (2014) SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* 123(7): 1021–31.

Colombo E, Marine J-C, Danovi D, Falini B and Pelicci PG (2002) Nucleophosmin regulates the stability and transcriptional activity of p53. *Nature cell biology* 4(7): 529–33.

Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, Marcé S, López-Guillermo A, Campo E and Montserrat E (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *The New England journal of medicine*. Massachusetts Medical Society 348(18): 1764–75.

Crow YJ, Leitch A, Hayward BE, Garner A, Parmar R, Griffith E, Ali M, Semple C, Aicardi J, Babul-Hirji R, Baumann C, Baxter P, Bertini E, Chandler KE, Chitayat D, Cau D, Déry C, Fazzi E, Goizet C, King MD, Klepper J, Lacombe D, Lanzi G, Lyall H, Martínez-Frías ML, Mathieu M, McKeown C, Monier A, Oade Y, Quarrell OW, Rittey CD, Rogers RC, Sanchis A, Stephenson JBP, Tacke U, Till M, Tolmie JL, Tomlin P, Voit T, Weschke B, Woods CG, Lebon P, Bonthron DT, Ponting CP and Jackson AP (2006) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nature genetics* 38(8): 910–6.

Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M and Chiorazzi N (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. American Society of Hematology 94(6): 1840–7.

Deaglio S, Vaisitti T, Aydin S, Bergui L, D'Arena G, Bonello L, Omedé P, Scatolini M, Jaksic O, Chiorino G, Efremov D and Malavasi F (2007) CD38 and ZAP-70 are functionally linked and mark CLL cells with high migratory potential. *Blood* 110(12): 4012–21.

Deaton AM and Bird A (2011) CpG islands and the regulation of transcription. *Genes & development* 25(10): 1010–22.

Delgado J, Pratt G, Phillips N, Briones J, Fegan C, Nomdedeu J, Pepper C, Aventin A, Ayats R, Brunet S, Martino R, Valcarcel D, Milligan D and Sierra J (2009) Beta2-microglobulin is a better predictor of treatment-free survival in patients with chronic lymphocytic leukaemia if adjusted according to glomerular filtration rate. *British journal of haematology* 145(6): 801–5.

Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, Kosmider O, Le Couedic J-PP, Robert F, Alberdi A, Lecluse Y, Plo I, Dreyfus FJ, Marzac C, Casadevall N, Lacombe C, Romana SP, Dessen P, Soulier J, Viguie F, Fontenay M, Vainchenker W, Bernard OA, Massé A, Kosmider O, Le Couedic J-PP, Robert F, Alberdi A, Lécluse Y, Plo I, Dreyfus FJ, Marzac C,

Casadevall N, Lacombe C, Romana SP, Dessen P, Soulier J, Viguié F, Fontenay M, Vainchenker W and Bernard OA (2009) Mutation in TET2 in myeloid cancers (2009/05/29.). *The New England journal of medicine* 360(22): 2289–2301.

Dielschneider RF, Xiao W, Yoon J-Y, Noh E, Banerji V, Li H, Marshall AJ, Johnston JB and Gibson SB (2014) Gefitinib targets ZAP-70-expressing chronic lymphocytic leukemia cells and inhibits B-cell receptor signaling. *Cell death & disease* 5: e1439.

Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz MA, Sierra J, Tallman MS, Löwenberg B and Bloomfield CD (2010) Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 115(3): 453–74.

Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, Döhner K, Bentz M and Lichter P (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *The New England journal of medicine*. Massachusetts Medical Society 343(26): 1910–6.

Döhner H, Stilgenbauer S, James MR, Benner A, Weilguni T, Bentz M, Fischer K, Hunstein W and Lichter P (1997) 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. *Blood*. American Society of Hematology 89(7): 2516–22.

Döhner K, Schlenk RF, Habdank M, Scholl C, Rücker FG, Corbacioglu A, Bullinger L, Fröhling S and Döhner H (2005) Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* 106(12): 3740–6.

Doneda L, Montillo M, Intropido L, Tedeschi A, Morra E and Larizza L (2003) Interphase fluorescence in situ hybridization analysis of del(11)(q23) and del(17)(p13) in chronic lymphocytic leukemia. *Cancer Genetics and Cytogenetics* 140(1): 31–36.

Dufort S, Richard M-J and de Fraipont F (2009) Pyrosequencing method to detect KRAS mutation in formalin-fixed and paraffin-embedded tumor tissues. *Analytical biochemistry* 391(2): 166–8.

Dufour A, Palermo G, Zellmeier E, Mellert G, Duchateau-Nguyen G, Schneider S, Benthaus T, Kakadia PM, Spiekermann K, Hiddemann W, Braess J, Truong S, Patten N, Wu L, Lohmann S, Dornan D, GuhaThakurta D, Yeh R-F, Salogub G, Solal-Celigny P, Dmoszynska A, Robak T, Montillo M, Catalano J, Geisler CH, Weisser M and Bohlander SK (2013) Inactivation of TP53 correlates with disease progression and low miR-34a expression in previously treated chronic lymphocytic leukemia patients. *Blood* 121(18):

3650–7.

Dufour A, Schneider F, Metzeler KH, Hoster E, Schneider S, Zellmeier E, Benthaus T, Sauerland M-C, Berdel WE, Büchner T, Wörmann B, Braess J, Hiddemann W, Bohlander SK and Spiekermann K (2010) Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28(4): 570–7.

Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, McKenna A, Carter SL, Cibulskis K, Sivachenko A, Saksena G, Voet D, Ramos AH, Auclair D, Thompson K, Sougnez C, Onofrio RC, Guiducci C, Beroukhim R, Zhou Z, Lin L, Lin J, Reddy R, Chang A, Landrenau R, Pennathur A, Ogino S, Luketich JD, Golub TR, Gabriel SB, Lander ES, Beer DG, Godfrey TE, Getz G and Bass AJ (2013) Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. (2013/03/26.). *Nature genetics* 45(5): 478–86.

Dürig J, Nückel H, Cremer M, Führer A, Halfmeyer K, Fandrey J, Möröy T, Klein-Hitpass L and Dührsen U (2003) ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia. *Leukemia*. Nature Publishing Group 17(12): 2426–34.

Van Dyke DL, Werner L, Rassenti LZ, Neuberg D, Ghia E, Heerema NA, Dal Cin P, Dell Aquila M, Sreekantaiah C, Greaves AW, Kipps T and Kay NE (2016) The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): the CLL Research Consortium experience. *British journal of haematology* 173(1): 105–13.

Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, Bullinger L, Kühn MWM, Gerhardinger A, Bloehdorn J, Radtke I, Su X, Ma J, Pounds S, Hallek M, Lichter P, Korbel J, Busch R, Mertens D, Downing JR, Stilgenbauer S and Döhner H (2012) High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood*. American Society of Hematology 120(24): 4783–94.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J and Turner S (2009) Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)* 323(5910): 133–8.

van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson S V, Howard HC, Cambon-Thomsen A, Knoppers BM, Meijers-Heijboer H,

Scheffer H, Tranebjaerg L, Dondorp W, de Wert GMWR and ESHG Public and Professional Policy Committee (2013) Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. *European journal of human genetics : EJHG* 21 Suppl 1: S1-5.

Endler G, Greinix H, Winkler K, Mitterbauer G and Mannhalter C (1999) Genetic fingerprinting in mouthwashes of patients after allogeneic bone marrow transplantation. *Bone Marrow Transplantation* 24(1): 95–98.

Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones A V, Waghorn K, Zoi K, Ross FM, Reiter A, Hochhaus A, Drexler HG, Duncombe A, Cervantes F, Oscier D, Boultwood J, Grand FH and Cross NCP (2010) Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders (2010/07/06.). *Nature genetics* 42(8): 722–726.

Ewalt M, Galili NG, Mumtaz M, Churchill M, Rivera S, Borot F, Raza A and Mukherjee S (2011) DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia. *Blood Cancer Journal*. Nature Publishing Group 1(3): e9-3.

Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, Ma J, Grunn A, Fangazio M, Capello D, Monti S, Cresta S, Gargiulo E, Forconi F, Guarini A, Arcaini L, Paulli M, Laurenti L, Larocca LM, Marasca R, Gattei V, Oscier D, Bertoni F, Mullighan CG, Foá R, Pasqualucci L, Rabadan R, Dalla-Favera R and Gaidano G (2011) Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *The Journal of experimental medicine* 208(7): 1389–401.

Falini B, Bolli N, Shan J, Martelli MP, Liso A, Pucciarini A, Bigerna B, Pasqualucci L, Mannucci R, Rosati R, Gorello P, Diverio D, Roti G, Tiacci E, Cazzaniga G, Biondi A, Schnittger S, Haferlach T, Hiddemann W, Martelli MF, Gu W, Mecucci C and Nicoletti I (2006) Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood*. American Society of Hematology 107(11): 4514–23.

Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L, La Starza R, Diverio D, Colombo E, Santucci A, Bigerna B, Pacini R, Pucciarini A, Liso A, Vignetti M, Fazi P, Meani N, Pettirossi V, Saglio G, Mandelli F, Lo-Coco F, Pelicci P-G and Martelli MF (2005) Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *The New England journal of medicine*. Massachusetts Medical Society 352(3): 254–66.

Fasan A, Haferlach C, Alpermann T, Jeromin S, Grossmann V, Eder C, Weissmann S, Dicker F, Kohlmann A, Schindela S, Kern W, Haferlach T and Schnittger S (2014) The role of different genetic subtypes of CEBPA mutated AML. *Leukemia*. Nature Publishing Group 28(4): 794–803.

Fernandez-Mercado M, Burns A, Pellagatti A, Giagounidis A, Germing U, Agirre X, Prosper F, Aul C, Killick S, Wainscoat JS, Schuh A and Boultwood J (2013) Targeted re-sequencing analysis of 25 genes commonly mutated in

myeloid disorders in del(5q) myelodysplastic syndromes. *Haematologica* 98(12): 1856–64.

Fiedler W, Kayser S, Kebenko M, Janning M, Krauter J, Schittenhelm M, Götze K, Weber D, Göhring G, Teleanu V, Thol F, Heuser M, Döhner K, Ganser A, Döhner H and Schlenk RF (2015) A phase I/II study of sunitinib and intensive chemotherapy in patients over 60 years of age with acute myeloid leukaemia and activating FLT3 mutations. *British journal of haematology* 169(5): 694–700.

Fischer K, Cramer P, Busch R, Böttcher S, Bahlo J, Schubert J, Pflüger KH, Schott S, Goede V, Isfort S, von Tresckow J, Fink A-M, Bühler A, Winkler D, Kreuzer K-A, Staib P, Ritgen M, Kneba M, Döhner H, Eichhorst BF, Hallek M, Stilgenbauer S and Wendtner C-M (2012) Bendamustine in combination with rituximab for previously untreated patients with chronic lymphocytic leukemia: a multicenter phase II trial of the German Chronic Lymphocytic Leukemia Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30(26): 3209–16.

Fischer K, Cramer P, Busch R, Stilgenbauer S, Bahlo J, Schweighofer CD, Böttcher S, Staib P, Kiehl M, Eckart MJ, Kranz G, Goede V, Elter T, Bühler A, Winkler D, Kneba M, Döhner H, Eichhorst BF, Hallek M and Wendtner C-M (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 : official journal of the American Society of Clinical Oncology* 29(26): 3559–66.

Foà R, Del Giudice I, Cuneo A, Del Poeta G, Ciolli S, Di Raimondo F, Lauria F, Cencini E, Rigolin GM, Cortelezzi A, Nobile F, Callea V, Brugiatelli M, Massaia M, Molica S, Trentin L, Rizzi R, Specchia G, Di Serio F, Orsucci L, Ambrosetti A, Montillo M, Zinzani PL, Ferrara F, Morabito F, Mura MA, Soriani S, Peragine N, Tavolaro S, Bonina S, Marinelli M, De Propris MS, Starza I Della, Piciocchi A, Alietti A, Runggaldier EJ, Gamba E, Mauro FR, Chiaretti S and Guarini A (2014) Chlorambucil plus rituximab with or without maintenance rituximab as first-line treatment for elderly chronic lymphocytic leukemia patients. *American Journal of Hematology* 89(5): 480–486.

Fonatsch C, Gudat H, Lengfelder E, Wandt H, Silling-Engelhardt G, Ludwig WD, Thiel E, Freund M, Bodenstein H and Schwieder G (1994) Correlation of cytogenetic findings with clinical features in 18 patients with inv(3)(q21q26) or t(3;3)(q21;q26). *Leukemia* 8(8): 1318–26.

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U and Campbell PJ (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research*. Oxford University Press 43(Database issue): D805-11.

Forconi F, Rinaldi A, Kwee I, Sozzi E, Raspadori D, Rancoita PM V., Scandurra M, Rossi D, Deambrogi C, Capello D, Zucca E, Marconi D, Bomben R, Gattei V, Lauria F, Gaidano G and Bertoni F (2008) Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. *British Journal of Haematology*.

Fredriksson NJ, Ny L, Nilsson JA and Larsson E (2014) Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature Genetics*. Nature Research 46(12): 1258–1263.

Fröhling S, Schlenk RF, Breitruck J, Benner A, Kreitmeier S, Tobis K, Döhner H and Döhner K (2002) Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. *Blood* 100(13): 4372–80.

Gaidano G, Foà R and Dalla-Favera R (2012) Molecular pathogenesis of chronic lymphocytic leukemia. *The Journal of clinical investigation* 122(10): 3432–8.

Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Röck J, Paschka P, Corbacioglu A, Krauter J, Schlegelberger B, Ganser A, Späth D, Kündgen A, Schmidt-Wolf IGH, Götze K, Nachbaur D, Pfreundschuh M, Horst HA, Döhner H and Döhner K (2011) RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29(10): 1364–72.

Gaidzik VI, Schlenk RF, Moschny S, Becker A, Bullinger L, Corbacioglu A, Krauter J, Schlegelberger B, Ganser A, Döhner H and Döhner K (2009) Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood* 113(19): 4505–11.

Gaidzik VI, Schlenk RF, Paschka P, Stölzle A, Späth D, Kuendgen A, von Lilienfeld-Toal M, Brugger W, Derigs HG, Kremers S, Greil R, Raghavachar A, Ringhoffer M, Salih HR, Wattad M, Kirchen HG, Runde V, Heil G, Petzer AL, Girschikofsky M, Heuser M, Kayser S, Goehring G, Teleanu M-V, Schlegelberger B, Ganser A, Krauter J, Bullinger L, Döhner H and Döhner K (2013) Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). *Blood* 121(23): 4769–77.

Gale RE, Green C, Allen C, Mead AJ, Burnett AK, Hills RK and Linch DC (2008) The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood* 111(5): 2776–84.

Gattei V, Bulian P, Del Principe MI, Zucchetto A, Maurillo L, Buccisano F, Bomben R, Dal-Bo M, Luciano F, Rossi FM, Degan M, Amadori S and Del

Poeta G (2008) Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia. *Blood* 111(2): 865–73.

Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, Prebet T, Nezri M, Sainty D, Olschwang S, Xerri L, Chaffanet M, Mozziconacci M-JJ, Vey N, Birnbaum D, Adélaïde J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, Prebet T, Nezri M, Sainty D, Olschwang S, Xerri L, Chaffanet M, Mozziconacci M-JJ, Vey N and Birnbaum D (2009) Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia (2009/04/25.). *British journal of haematology* 145(6): 788–800.

Gharizadeh B, Nordström T, Ahmadian A, Ronaghi M and Nyrén P (2002) Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Analytical biochemistry* 301(1): 82–90.

Ghoneim DH, Myers JR, Tuttle E and Paciorkowski AR (2014) Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. *BMC research notes*. BioMed Central 7(1): 864.

Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, Steensma DP, Fryer A, Goudie DR, Krantz ID and Traeger-Synodinos J (2008) Mutations in the chromatin-associated protein ATRX (2008/04/15.). *Human mutation* 29(6): 796–802.

Di Giovanni S, Valentini G, Carducci P and Giallonardo P (1989) Beta-2-microglobulin is a reliable tumor marker in chronic lymphocytic leukemia. *Acta haematologica* 81(4): 181–5.

Del Giudice I, Rossi D, Chiaretti S, Marinelli M, Tavolaro S, Gabrielli S, Laurenti L, Marasca R, Rasi S, Fangazio M, Guarini A, Gaidano G and Foà R (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 (2011/12/31.). *Haematologica* 97(3): 437–441.

González-Gascón Y Marín I, Hernández-Sánchez M, Rodríguez-Vicente A-E, Sanzo C, Aventín A, Puiggros A, Collado R, Heras C, Muñoz C, Delgado J, Ortega M, González M-T, Marugán I, de la Fuente I, Recio I, Bosch F, Espinet B, González M, Hernández-Rivas J-M and Hernández J-Á (2015) A high proportion of cells carrying trisomy 12 is associated with a worse outcome in patients with chronic lymphocytic leukemia. *Hematological oncology*.

Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E, Dearden CE, Richards SM, Catovsky D and Morgan GJ (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): 2223–9.

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC and McCombie WR (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome research* 25(11): 1750–6.

Gorello P, Cazzaniga G, Alberti F, Dell'Oro MG, Gottardi E, Specchia G, Roti G, Rosati R, Martelli MF, Diverio D, Coco F Lo, Biondi A, Saglio G, Mecucci C and Falini B (2006) Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (NPM1) gene mutations. *Leukemia*. Nature Publishing Group 20(6): 1103–1108.

Grand FH, Hidalgo-Curtis CE, Ernst T, Zoi K, Zoi C, McGuire C, Kreil S, Jones A, Score J, Metzgeroth G, Oscier D, Hall A, Brandts C, Serve H, Reiter A, Chase AJ and Cross NCP (2009) Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood* 113(24): 6182–92.

Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, Krysiak K, Harris CC, Koboldt DC, Larson DE, McLellan MD, Dooling DJ, Abbott R, Fulton RS, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Grillot M, Baty J, Heath S, Frater JL, Nasim T, Link DC, Tomasson MH, Westervelt P, DiPersio JF, Mardis ER, Ley TJ, Wilson RK and Walter MJ (2011) Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nature Genetics* 44(1): 53–57.

Grimwade D (2012) The changing paradigm of prognostic factors in acute myeloid leukaemia. *Best practice & research. Clinical haematology* 25(4): 419–25.

Grimwade D and Freeman SD (2014) Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for 'prime time'? *Blood* 124(23): 3345–3355.

Grimwade D, Hills RK, Moorman A V, Walker H, Chatters S, Goldstone AH, Wheatley K, Harrison CJ and Burnett AK (2010) Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* 116(3): 354–65.

Grimwade D, Walker H, Oliver F, Wheatley K, Harrison C, Harrison G, Rees J, Hann I, Stevens R, Burnett A and Goldstone A (1998) The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* 92(7): 2322–33.

Gruber M and Wu CJ (2014) Evolving understanding of the CLL genome. *Seminars in hematology* 51(3): 177–87.

Guièze R, Robbe P, Clifford R, de Guibert S, Pereira B, Timbs A, Dilhuydy M-S, 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 SJL, Tournilhac O and Schuh A (2015) Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood* 126(18): 2110–7.

Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon K-R, Jankovic M, Oliveira T, Bothmer A, Voss TC, Ansarah-Sobrinho C, Mathe E, Liang G, Cobell J, Nakahashi H, Robbiani DF, Nussenzweig A, Hager GL, Nussenzweig MC and Casellas R (2012) DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* 484(7392): 69–74.

Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating MJ, Montserrat E, Rai KR and Kipps TJ (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*. American Society of Hematology 111(12): 5446–56.

Hallek M, Fischer K, Fingerle-Rowson G, Fink AM, Busch R, Mayer J, Hensel M, Hopfinger G, Hess G, von Grünhagen U, Bergmann M, Catalano J, Zinzani PL, Caligaris-Cappio F, Seymour JF, Berrebi A, Jäger U, Cazin B, Trneny M, Westermann A, Wendtner CM, Eichhorst BF, Staib P, Bühler A, Winkler D, Zenz T, Böttcher S, Ritgen M, Mendila M, Kneba M, Döhner H and Stilgenbauer S (2010) Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet (London, England)* 376(9747): 1164–74.

Hamblin TJ, Davis Z, Gardiner A, Oscier DG and Stevenson FK (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. American Society of Hematology 94(6): 1848–54.

Harada Y and Harada H (2009) Molecular pathways mediating MDS/AML with focus on AML1/RUNX1 point mutations. *Journal of Cellular Physiology*, 16–20.

Hayakawa F, Towatari M, Kiyoi H, Tanimoto M, Kitamura T, Saito H and Naoe T (2000) Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines. *Oncogene* 19(5): 624–31.

Helbig G, Wozniczka K, Wieclawek A, Soja A, Bartkowska-Chrobok A and Kyrcz-Krzemien S (2014) Clinical relevance of mutant NPM1 and CEBPA in patients with acute myeloid leukaemia - preliminary report. *Contemporary oncology (Poznań, Poland)* 18(4): 241–5.

Herbaux C, Duployez N, Badens C, Poret N, Gardin C, Decamp M, Eclache V, Daliphard S, Murati A, Cony-Makhoul P, Cheze S, Beve B, Lacoste C, Prebet T, Hunault-Berger M, Maloisel F, Renneville A, Figeac M, Stamatoullas-Bastard A, Bastard C, Fenaux P, Preudhomme C and Rose C (2015)

Incidence of ATRX mutations in myelodysplastic syndromes, the value of microcytosis. *American Journal of Hematology* 90(8): 737–738.

Hernández JÁ, Hernández-Sánchez M, Rodríguez-Vicente AE, Grossmann V, Collado R, Heras C, Puiggros A, Martín AÁ, Puig N, Benito R, Robledo C, Delgado J, González T, Queizán JA, Galende J, de la Fuente I, Martín-Núñez G, Alonso JM, Abrisqueta P, Luño E, Marugán I, González-Gascón I, Bosch F, Kohlmann A, González M, Espinet B and Hernández-Rivas JM (2015) A Low Frequency of Losses in 11q Chromosome Is Associated with Better Outcome and Lower Rate of Genomic Mutations in Patients with Chronic Lymphocytic Leukemia. *PloS one*. Public Library of Science 10(11): e0143073.

Hess JL (2004) MLL: a histone methyltransferase disrupted in leukemia. *Trends in molecular medicine* 10(10): 500–7.

Ho AS, Kannan K, Roy DM, Morris LGT, Ganly I, Katabi N, Ramaswami D, Walsh LA, Eng S, Huse JT, Zhang J, Dolgalev I, Huberman K, Heguy A, Viale A, Drobnjak M, Leversha MA, Rice CE, Singh B, Iyer NG, Leemans CR, Bloemena E, Ferris RL, Seethala RR, Gross BE, Liang Y, Sinha R, Peng L, Raphael BJ, Turcan S, Gong Y, Schultz N, Kim S, Chiosea S, Shah JP, Sander C, Lee W and Chan TA (2013) The mutational landscape of adenoid cystic carcinoma. *Nature genetics* 45(7): 791–8.

Hollink IHIM, van den Heuvel-Eibrink MM, Arentsen-Peters STCJM, Zimmermann M, Peeters JK, Valk PJM, Balgobind B V, Sonneveld E, Kaspers GJL, de Bont ESJM, Trka J, Baruchel A, Creutzig U, Pieters R, Reinhardt D and Zwaan CM (2011) Characterization of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia. *Haematologica* 96(3): 384–92.

Hollink IHIM, Zwaan CM, Zimmermann M, Arentsen-Peters TCJM, Pieters R, Cloos J, Kaspers GJL, de Graaf SSN, Harbott J, Creutzig U, Reinhardt D, van den Heuvel-Eibrink MM and Thiede C (2009) Favorable prognostic impact of NPM1 gene mutations in childhood acute myeloid leukemia, with emphasis on cytogenetically normal AML. *Leukemia* 23(2): 262–70.

Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, Payne-Turner D, Churchman M, Andersson A, Chen S-C, McCastlain K, Becksfort J, Ma J, Wu G, Patel SN, Heatley SL, Phillips LA, Song G, Easton J, Parker M, Chen X, Rusch M, Boggs K, Vadodaria B, Hedlund E, Drenberg C, Baker S, Pei D, Cheng C, Huether R, Lu C, Fulton RS, Fulton LL, Tabib Y, Dooling DJ, Ochoa K, Minden M, Lewis ID, To LB, Marlton P, Roberts AW, Raca G, Stock W, Neale G, Drexler HG, Dickins RA, Ellison DW, Shurtleff SA, Pui C-H, Ribeiro RC, Devidas M, Carroll AJ, Heerema NA, Wood B, Borowitz MJ, Gastier-Foster JM, Raimondi SC, Mardis ER, Wilson RK, Downing JR, Hunger SP, Loh ML and Mullighan CG (2013) The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature genetics* 45(3): 242–52.

Hon GC, Hawkins RD and Ren B (2009) Predictive chromatin signatures in

the mammalian genome. *Human Molecular Genetics*. Oxford University Press 18(R2): R195–R201.

Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schadendorf D and Kumar R (2013) TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)* 339(6122): 959–61.

Hou HA, Huang TC, Lin LI, Liu CY, Chen CY, Chou WC, Tang JL, Tseng MH, Huang CF, Chiang YC, Lee FY, Liu MC, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC and Tien HF (2010) WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system (2010/04/07.). *Blood* 115(25): 5222–5231.

Hou HA, Kuo Y-Y, Liu C-Y, Chou W-C, Lee MC, Chen C-Y, Lin L-I, Tseng M-H, Huang C-F, Chiang Y-C, Lee F-Y, Liu M-C, Liu C-W, Tang J-L, Yao M, Huang S-Y, Ko B-S, Hsu S-C, Wu S-J, Tsay W, Chen Y-C and Tien H-F (2012) DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood* 119(2): 559–68.

Huang DW, Raley C, Jiang MK, Zheng X, Liang D, Rehman MT, Highbarger HC, Jiao X, Sherman B, Ma L, Chen X, Skelly T, Troyer J, Stephens R, Imamichi T, Pau A, Lempicki RA, Tran B, Nissley D, Lane HC and Dewar RL (2016) Towards Better Precision Medicine: PacBio Single-Molecule Long Reads Resolve the Interpretation of HIV Drug Resistant Mutation Profiles at Explicit Quasispecies (Haplotype) Level. *Journal of data mining in genomics & proteomics*. NIH Public Access 7(1).

Huang FW, Hodis E, Xu MJ, Kryukov G V, Chin L, Garraway LA, Berger MF, Barretina J, Hodis E, Krauthammer M, Imielinski M, Bennett DC, Michaloglou C, Zhang A, Pirker C, Rutter JL, Horikawa I, Cable PL, Afshari C, Barrett JC and Babitt JL (2013) Highly recurrent TERT promoter mutations in human melanoma. *Science (New York, N.Y.)*. American Association for the Advancement of Science 339(6122): 957–9.

Huang M, Thomas D, Li MX, Feng W, Chan SM, Majeti R and Mitchell BS (2013) Role of cysteine 288 in nucleophosmin cytoplasmic mutations: sensitization to toxicity induced by arsenic trioxide and bortezomib. *Leukemia* 27(10): 1970–80.

Huang SJT, Gillan TL, Gerrie AS, Hrynchak M, Karsan A, Ramadan K, Smith AC, Toze CL and Bruyere H (2016) Influence of clone and deletion size on outcome in chronic lymphocytic leukemia patients with an isolated deletion 13q in a population-based analysis in British Columbia, Canada. *Genes, chromosomes & cancer* 55(1): 16–24.

Huang T, Zhuge J and Zhang WW (2013) Sensitive detection of BRAF V600E mutation by Amplification Refractory Mutation System (ARMS)-PCR. *Biomarker research* 1(1): 3.

Huemer M, Rebhandl S, Zaborsky N, Gassner F, Hainzl S, Weiss L, Hebenstreit D, Greil R and Geisberger R (2014) AID induces intraclonal diversity and genomic damage in CD86(+) chronic lymphocytic leukemia cells. *European Journal of Immunology* 44(12): 3747–3757.

Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Ribas A and Lo RS (2016) Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* 165(1): 35–44.

Di Ianni M, Baldoni S, Rosati E, Ciurnelli R, Cavalli L, Martelli MF, Marconi P, Screpanti I and Falzetti F (2009) A new genetic lesion in B-CLL: a NOTCH1 PEST domain mutation. *British journal of haematology* 146(6): 689–91.

Ibrahim S (2001) CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. *Blood*. American Society of Hematology 98(1): 181–186.

Ingram W, Lea NC, Cervera J, Germing U, Fenaux P, Cassinat B, Kiladjian JJ, Varkonyi J, Antunovic P, Westwood NB, Arno MJ, Mohamedali A, Gaken J, Kontou T, Czepulkowski BH, Twine NA, Tamaska J, Csomer J, Benedek S, Gattermann N, Zipperer E, Giagounidis A, Garcia-Casado Z, Sanz G and Mufti GJ (2006) The JAK2 V617F mutation identifies a subgroup of MDS patients with isolated deletion 5q and a proliferative bone marrow. *Leukemia* 20(7): 1319–1321.

Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, Leggett RM, Eccles DA, Zalunin V, Urban JM, Piazza P, Bowden RJ, Paten B, Mwaigwisya S, Batty EM, Simpson JT, Snutch TP, Birney E, Buck D, Goodwin S, Jansen HJ, O'Grady J and Olsen HE (2015) MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research* 4(1075): 1–35.

Itzykson R, Kosmider O, Renneville A, Gelsi-Boyer V, Meggendorfer M, Morabito M, Berthon C, Adès L, Fenaux P, Beyne-Rauzy O, Vey N, Braun T, Haferlach T, Dreyfus F, Cross NCP, Preudhomme C, Bernard OA, Fontenay M, Vainchenker W, Schnittger S, Birnbaum D, Droin N and Solary E (2013) Prognostic score including gene mutations in chronic myelomonocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 31(19): 2428–36.

Ivey A, Hills RK, Simpson MA, Jovanovic J V., Gilkes A, Grech A, Patel Y, Bhudia N, Farah H, Mason J, Wall K, Akiki S, Griffiths M, Solomon E, McCaughan F, Linch DC, Gale RE, Vyas P, Freeman SD, Russell N, Burnett AK and Grimwade D (2016) Assessment of Minimal Residual Disease in Standard-Risk AML. *New England Journal of Medicine*. Massachusetts Medical Society 374(5): 422–433.

James C, Ugo V, Le Couédic J-P, Staerk J, Delhommeau F, Lacout C, Garçon L, Raslova H, Berger R, Bennaceur-Griscelli A, Villeval JL, Constantinescu SN, Casadevall N and Vainchenker W (2005) A unique clonal JAK2 mutation

leading to constitutive signalling causes polycythaemia vera. *Nature* 434(7037): 1144–1148.

Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, Alpermann T, Roller A, Kohlmann A, Haferlach T, Kern W and Schnittger S (2014) SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* 28(1): 108–17.

Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, Simes RJ, Chalchal H, Shapiro JD, Robitaille S, Price TJ, Shepherd L, Au H-J, Langer C, Moore MJ and Zalcberg JR (2008) K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *The New England journal of medicine* 359(17): 1757–65.

Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, Sougnez C, Gabriel S, Lander ES, Kim HT, Getz G and Brown JR (2015) Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nature communications*. Nature Publishing Group 6: 8866.

Kaspers GJL, Zimmermann M, Reinhardt D, Gibson BES, Tamminga RYJ, Aleinikova O, Armendariz H, Dworzak M, Ha S-Y, Hasle H, Hovi L, Maschan A, Bertrand Y, Leverger GG, Razzouk BI, Rizzari C, Smisek P, Smith O, Stark B and Creutzig U (2013) Improved outcome in pediatric relapsed acute myeloid leukemia: results of a randomized trial on liposomal daunorubicin by the International BFM Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 31(5): 599–607.

Kato L, Begum NA, Burroughs AM, Doi T, Kawai J, Daub CO, Kawaguchi T, Matsuda F, Hayashizaki Y and Honjo T (2012) Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. *Proceedings of the National Academy of Sciences of the United States of America* 109(7): 2479–84.

Keating MJ, O'Brien S, Albitar M, Lerner S, Plunkett W, Giles F, Andreeff M, Cortes J, Faderl S, Thomas D, Koller C, Wierda W, Detry MA, Lynn A and Kantarjian H (2005) Early results of a chemoimmunotherapy regimen of fludarabine, cyclophosphamide, and rituximab as initial therapy for chronic lymphocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 23(18): 4079–88.

Kitada S, Andersen JW, Akar S, Zapata JM, Takayama S, Krajewski S, Wang HG, Zhang X, Bullrich F, Croce CM, Rai K, Hines J and Reed JC (1998) Expression of apoptosis-regulating proteins in chronic lymphocytic leukemia: correlations with In vitro and In vivo chemoresponses. *Blood*. American Society of Hematology 91(9): 3379–89.

Klein K, Kaspers G, Harrison CJ, Beverloo HB, Reedijk A, Bongers M, Cloos J, Pession A, Reinhardt D, Zimmerman M, Creutzig U, Dworzak M, Alonzo TA,

Johnston D, Hirsch B, Zapotocky M, De Moerloose B, Fynn A, Lee V, Taga T, Tawa A, Auvrignon A, Zeller B, Forestier E, Salgado C, Balwierz W, Popa A, Rubnitz J, Raimondi S and Gibson B (2015) Clinical Impact of Additional Cytogenetic Aberrations, *cKIT* and *RAS* Mutations, and Treatment Elements in Pediatric t(8;21)-AML: Results From an International Retrospective Study by the International Berlin-Frankfurt-Münster Study Group. *Journal of Clinical Oncology* 33(36): 4247–4258.

Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA, Walker H, Wheatley K, Bowen DT, Burnett AK, Goldstone AH and Linch DC (2001) The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United King. *Blood* 98(6): 1752–9.

Kralovics R, Passamonti F, Buser AS, Teo S-S, Tiedt R, Passweg JR, Tichelli A, Cazzola M and Skoda RC (2005) *A gain-of-function mutation of JAK2 in myeloproliferative disorders. The New England journal of medicine*, 1779–90. Available at: http://www.nejm.org/doi/full/10.1056/NEJMoa051113 (accessed 07/08/15).

Krauth M-T, Alpermann T, Bacher U, Eder C, Dicker F, Ulke M, Kuznia S, Nadarajah N, Kern W, Haferlach C, Haferlach T and Schnittger S (2015) WT1 mutations are secondary events in AML, show varying frequencies and impact on prognosis between genetic subgroups. *Leukemia* 29(3): 660–7.

Kröber A, Seiler T, Benner A, Bullinger L, Brückle E, Lichter P, Döhner H, Stilgenbauer S, Bru E and Do H (2002) V H mutation status , CD38 expression level , genomic aberrations , and survival in chronic lymphocytic leukemia. *Blood* 100(4): 1410–1416.

Krönke J, Schlenk RF, Jensen K-O, Tschürtz F, Corbacioglu A, Gaidzik VI, Paschka P, Onken S, Eiwen K, Habdank M, Späth D, Lübbert M, Wattad M, Kindler T, Salih HR, Held G, Nachbaur D, von Lilienfeld-Toal M, Germing U, Haase D, Mergenthaler H-G, Krauter J, Ganser A, Göhring G, Schlegelberger B, Döhner H and Döhner K (2011) Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29(19): 2709–16.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ and Marra MA (2009) Circos: An information aesthetic for comparative genomics. *Genome Research*. Cold Spring Harbor Laboratory Press 19(9): 1639–1645.

Kui JS, Espinal-Witter R and Wang YL (2013) Laboratory detection of JAK2V617F in human myeloproliferative neoplasms. *Methods in molecular biology (Clifton, N.J.)* 999: 41–57.

Kundu M and Liu PP (2001) Function of the inv(16) fusion gene CBFB-MYH11. *Current opinion in hematology* 8(4): 201–5.

Kurki S, Peltonen K, Latonen L, Kiviharju TM, Ojala PM, Meek D and Laiho M (2004) Nucleolar protein NPM interacts with HDM2 and protects tumor suppressor protein p53 from HDM2-mediated degradation. *Cancer cell* 5(5): 465–75.

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, Neuberg D, Brown JR, Getz G and Wu CJ (2013) Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* 152(4): 714–726.

Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, Carter SL, Cibulskis K, Mertens D, Sougnez CL, Rosenberg M, Hess JM, Edelmann J, Kless S, Kneba M, Ritgen M, Fink A, Fischer K, Gabriel S, Lander ES, Nowak MA, Döhner H, Hallek M, Neuberg D, Getz G, Stilgenbauer S and Wu CJ (2015) Mutations driving CLL and their evolution in progression and relapse. *Nature*. NIH Public Access 526(7574): 525–530.

Langemeijer SMC, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, Stevens-Linders E, van Hoogen P, van Kessel AG, Raymakers RAP, Kamping EJ, Verhoef GE, Verburgh E, Hagemeijer A, Vandenberghe P, de Witte T, van der Reijden BA and Jansen JH (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes (2009/06/02.). *Nature genetics* 41(7): 838–842.

LaRochelle O, Bertoli S, Vergez F, Sarry J-E, Mansat-De Mas V, Dobbelstein S, Dastugue N, Strzelecki A-C, Cavelier C, Creancier L, Pillon A, Kruczynski A, Demur C, Sarry A, Huguet F, Huynh A, Récher C and Delabesse E (2011) Do AML patients with DNMT3A exon 23 mutations benefit from idarubicin as compared to daunorubicin? A single center experience. *Oncotarget* 2(11): 850–61.

Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau D-A, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CWM, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES and Getz G (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457): 214–8.

Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P and Preudhomme C (2005) CEBPA point mutations in hematological malignancies. *Leukemia* 19(3): 329–34.

Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJP, Boggon TJ, Wlodarska I, Clark JJ, Moore S, Adelsperger J, Koo S, Lee JC, Gabriel S, Mercher T, D'Andrea A, Fröhling S, Döhner K, Marynen P, Vandenberghe P, Mesa RA, Tefferi A, Griffin JD, Eck MJ, Sellers WR, Meyerson M, Golub TR, Lee SJ and Gilliland DG (2005) Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7(4): 387–397.

Levis M (2013) FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* 2013: 220–6.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER and Wilson RK (2010) DNMT3A mutations in acute myeloid leukemia. (2010/11/12.). *The New England journal of medicine* 363(25): 2424–2433.

Lièvre A, Bachet J-B, Boige V, Cayre A, Le Corre D, Buc E, Ychou M, Bouché O, Landi B, Louvet C, André T, Bibeau F, Diebold M-D, Rougier P, Ducreux M, Tomasic G, Emile J-F, Penault-Llorca F and Laurent-Puig P (2008) KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26(3): 374–9.

Lin K, Farahani M, Yang Y, Johnson GG, Oates M, Atherton M, Douglas A, Kalakonda N and Pettitt AR (2014) Loss of MIR15A and MIR16-1 at 13q14 is associated with increased TP53 mRNA, de-repression of BCL2 and adverse outcome in chronic lymphocytic leukaemia. *British journal of haematology* 167(3): 346–55.

von Lindern M, Fornerod M, van Baal S, Jaegle M, de Wit T, Buijs A and Grosveld G (1992) The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. *Molecular and cellular biology* 12(4): 1687–97.

Link DC, Schuettpelz LG, Shen D, Wang J, Walter MJ, Kulkarni S, Payton JE, Ivanovich J, Goodfellow PJ, Le Beau M, Koboldt DC, Dooling DJ, Fulton RS, Bender RHF, Fulton LL, Delehaunty KD, Fronick CC, Appelbaum EL, Schmidt H, Abbott R, O'Laughlin M, Chen K, McLellan MD, Varghese N, Nagarajan R, Heath S, Graubert TA, Ding L, Ley TJ, Zambetti GP, Wilson RK and Mardis ER (2011) Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. *JAMA*

305(15): 1568–76.

Liu X, Bishop J, Shan Y, Pai S, Liu D, Murugan AK, Sun H, El-Naggar AK and Xing M (2013) Highly prevalent TERT promoter mutations in aggressive thyroid cancers. *Endocrine-related cancer* 20(4): 603–10.

Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5% nuclease assay. *Genetic Analysis: Biomolecular Engineering* 14: 143–149.

Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton L-A, 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 and Rosenquist R (2016) Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood*. American Society of Hematology 127(8): 1007–16.

Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, Cruz-Gordillo P, Francis JM, Zhang C-Z, Shalek AK, Satija R, Trombetta JJ, Lu D, Tallapragada N, Tahirova N, Kim S, Blumenstiel B, Sougnez C, Lowe A, Wong B, Auclair D, Van Allen EM, Nakabayashi M, Lis RT, Lee G-SM, Li T, Chabot MS, Ly A, Taplin M-E, Clancy TE, Loda M, Regev A, Meyerson M, Hahn WC, Kantoff PW, Golub TR, Getz G, Boehm JS and Love JC (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nature biotechnology* 32(5): 479–84.

Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, Novak AJ, Dogan A, Ansell SM, Link BK, Zou L, Gould J, Saksena G, Stransky N, Rangel-Escareno C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Hernandez-Lemus E, Schwarz-Cruz y Celis A, Imaz-Rosshandler I, Ojesina AI, Jung J, Pedamallu CS, Lander ES, Habermann TM, Cerhan JR, Shipp MA, Getz G, Golub TR, Rangel-Escareño C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Hernández-Lemus E, Schwarz-Cruz y Celis A, Imaz-Rosshandler I, Ojesina AI, Jung J, Pedamallu CS, Lander ES, Habermann TM, Cerhan JR, Shipp MA, Getz G and Golub TR (2012) Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. (2012/02/22.). *Proceedings of the National Academy of Sciences of the United States of America* 109(10): 3879–84.

Loupakis F, Ruzzo A, Cremolini C, Vincenzi B, Salvatore L, Santini D, Masi G, Stasi I, Canestrari E, Rulli E, Floriani I, Bencardino K, Galluccio N, Catalano V, Tonini G, Magnani M, Fontanini G, Basolo F, Falcone A and Graziano F (2009) KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *British journal of cancer* 101(4): 715–21.

Lu J, Gao J, Zhang J, Sun J, Wu H, Shi X, Teng L and Liang Z (2015) Association between BRAF V600E mutation and regional lymph node metastasis in papillary thyroid carcinoma. *International journal of clinical and experimental pathology* 8(1): 793–9.

Lunter G and Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads (2010/10/29.). *Genome research* 21(6): 936–939.

Luo C, Tsementzi D, Kyrpides N, Read T and Konstantinidis KT (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PloS one* 7(2): e30087.

Luskin MR, Lee J-W, Fernandez HF, Abdel-Wahab O, Bennett JM, Ketterling RP, Lazarus HM, Levine RL, Litzow MR, Paietta EM, Patel JP, Racevskis J, Rowe JM, Tallman MS, Sun Z and Luger SM (2016) Benefit of high-dose daunorubicin in AML induction extends across cytogenetic and molecular groups. *Blood* 127(12): 1551–8.

Madoui M-A, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A, Lemainque A, Wincker P and Aury J-M (2015) Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC genomics* 16: 327.

Majid A, Lin TT, Best G, Fishlock K, Hewamana S, Pratt G, Yallop D, Buggins AGS, Wagner S, Kennedy BJ, Miall F, Hills R, Devereux S, Oscier DG, Dyer MJS, Fegan C and Pepper C (2011) CD49d is an independent prognostic marker that is associated with CXCR4 expression in CLL. *Leukemia research* 35(6): 750–6.

Mancini M, Hasan SK, Ottone T, Lavorgna S, Ciardi C, Angelini DF, Agostini F, Venditti A and Lo-Coco F (2015) Two Novel Methods for Rapid Detection and Quantification of DNMT3A R882 Mutations in Acute Myeloid Leukemia. *The Journal of Molecular Diagnostics* 17(2): 179–184.

Mandoli A, Singh AA, Jansen PWTC, Wierenga ATJ, Riahi H, Franci G, Prange K, Saeed S, Vellenga E, Vermeulen M, Stunnenberg HG and Martens JHA (2014) CBFB-MYH11/RUNX1 together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. *Leukemia* 28(4): 770–8.

Mardis ER et al (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. Elsevier 24(3): 133–141.

Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott R, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy J V., Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson

MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK and Ley TJ (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine* 361(11): 1058–66.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-JJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes X V, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-BB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-BB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF and Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. (2005/08/02.). *Nature* 437(7057): 376–80.

Marková J, Michková P, Burčková K, Březinová J, Michalová K, Dohnalová A, Maaloufová JS, Soukup P, Vítek A, Cetkovský P and Schwarz J (2012) Prognostic impact of DNMT3A mutations in patients with intermediate cytogenetic risk profile acute myeloid leukemia. *European journal of haematology* 88(2): 128–35.

Mauro FR, Molica S, Laurenti L, Cortelezzi A, Carella AM, Zaja F, Chiarenza A, Angrilli F, Nobile F, Marasca R, Musolino C, Brugiatelli M, Piciocchi A, Vignetti M, Fazi P, Gentile G, De Propris MS, Della Starza I, Marinelli M, Chiaretti S, Del Giudice I, Nanni M, Albano F, Cuneo A, Guarini A and Foà R (2014) Fludarabine plus alemtuzumab (FA) front-line treatment in young patients with chronic lymphocytic leukemia (CLL) and an adverse biologic profile. *Leukemia research* 38(2): 198–203.

McPherson JD et al (2001) A physical map of the human genome (2001/03/10.). *Nature* 409(6822): 934–941.

Mendler JH, Maharry K, Radmacher MD, Mrózek K, Becker H, Metzeler KH, Schwind S, Whitman SP, Khalife J, Kohlschmidt J, Nicolet D, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G and Bloomfield CD (2012) RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30(25): 3109–18.

Mertens D, Philippen A, Ruppel M, Allegra D, Bhattacharya N, Tschuch C, Wolf S, Idler I, Zenz T and Stilgenbauer S (2009) Chronic lymphocytic

leukemia and 13q14: miRs and more. *Leukemia & lymphoma*. Taylor & Francis 50(3): 502–5.

Meshinchi S, Alonzo TA, Stirewalt DL, Zwaan M, Zimmerman M, Reinhardt D, Kaspers GJ, Heerema NA, Gerbing R, Lange BJ and Radich JP (2006) Clinical implications of FLT3 mutations in pediatric AML (2006/08/17.). *Blood* 108(12): 3654–3661.

Messina M, Del Giudice I, Khiabanian H, Rossi D, Chiaretti S, Rasi S, Spina V, Holmes AB, Marinelli M, Fabbri G, Piciocchi A, Mauro FR, Guarini A, Gaidano G, Dalla-Favera R, Pasqualucci L, Rabadan R and Foà R (2014) Genetic lesions associated with chronic lymphocytic leukemia chemo-refractoriness. *Blood* 123(15): 2378–88.

Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrozek K, Nicolet D, Whitman SP, Wu Y-Z, Schwind S, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G and Bloomfield CD (2011) ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood* 118(26): 6920–6929.

Mian SA, Rouault-Pierre K, Smith AE, Seidl T, Pizzitola I, Kizilors A, Kulasekararaj AG, Bonnet D and Mufti GJ (2015) SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. *Nature communications*. Nature Publishing Group 6: 10004.

Mikheyev AS and Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. *Molecular ecology resources* 14(6): 1097–102.

Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, Kemeza DM, Strassler DS, Ponting CP, Webber C and Devine SE (2011) Natural genetic variation caused by small insertions and deletions in the human genome. *Genome research* 21(6): 830–9.

Mizuki M, Fenski R, Halfter H, Matsumura I, Schmidt R, Müller C, Grüning W, Kratz-Albers K, Serve S, Steur C, Büchner T, Kienast J, Kanakura Y, Berdel WE and Serve H (2000) Flt3 mutations from patients with acute myeloid leukemia induce transformation of 32D cells mediated by the Ras and STAT5 pathways. *Blood*. American Society of Hematology 96(12): 3907–14.

Mizuki M, Schwable J, Steur C, Choudhary C, Agrawal S, Sargin B, Steffen B, Matsumura I, Kanakura Y, Böhmer FD, Müller-Tidow C, Berdel WE and Serve H (2003) Suppression of myeloid transcription factors and induction of STAT response genes by AML-specific Flt3 mutations. *Blood*. American Society of Hematology 101(8): 3164–73.

Molica S (2006) Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. *Leukemia & lymphoma*. Taylor & Francis 47(8): 1477–80.

Morganella S et al (2016) The topography of mutational processes in breast cancer genomes. *Nature Communications*. Nature Publishing Group 7:

11383.

Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, Trinh DL, Bolger-Munro M, Taylor G, Hadj Khodabakhshi A, Ben-Neriah S, Pon J, Meissner B, Woolcock B, Farnoud N, Rogic S, Lim EL, Johnson NA, Shah S, Jones S, Steidl C, Holt R, Birol I, Moore R, Connors JM, Gascoyne RD and Marra MA (2013) Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* 122(7): 1256–65.

Morise M, Taniguchi H, Saka H, Shindoh J, Suzuki R, Kojima E, Hase T, Ando M, Kondo M, Saito H and Hasegawa Y (2014) Phase II study of erlotinib for previously treated patients with EGFR wild-type non-small-cell lung cancer, following EGFR mutation status reevaluation with the Scorpion Amplified Refractory Mutation System. *Molecular and clinical oncology* 2(6): 991–996.

Mosca L, Fabris S, Lionetti M, Todoerti K, Agnelli L, Morabito F, Cutrona G, Andronache A, Matis S, Ferrari F, Gentile M, Spriano M, Callea V, Festini G, Molica S, Deliliers GL, Bicciato S, Ferrarini M and Neri A (2010) Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic lymphocytic leukemia patients with 13q14 deletion. *Clinical cancer research : an official journal of the American Association for Cancer Research* 16(23): 5641–53.

Mrózek K, Heinonen K, Lawrence D, Carroll AJ, Koduru PR, Rao KW, Strout MP, Hutchison RE, Moore JO, Mayer RJ, Schiffer CA and Bloomfield CD (1997) Adult patients with de novo acute myeloid leukemia and t(9; 11)(p22; q23) have a superior outcome to patients with other translocations involving band 11q23: a cancer and leukemia group B study. *Blood* 90(11): 4532–8.

Mrózek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP, Metzeler KH, Schwind S, Wu Y-Z, Kohlschmidt J, Pettenati MJ, Heerema NA, Block AW, Patil SR, Baer MR, Kolitz JE, Moore JO, Carroll AJ, Stone RM, Larson RA and Bloomfield CD (2012) Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30(36): 4515–23.

Mullaney JM, Mills RE, Pittard WS and Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. *Human molecular genetics* 19(R2): R131-6.

Murphy KM, Levis M, Hafez MJ, Geiger T, Cooper LC, Smith BD, Small D and Berg KD (2003) Detection of FLT3 internal tandem duplication and D835 mutations by a multiplex polymerase chain reaction and capillary electrophoresis assay (2003/04/23.). *The Journal of molecular diagnostics : JMD* 5(2): 96–102.

Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, Jares P, Navarro A, Martin-Garcia D, Bea S, Salaverria I, Oldreive C, Aymerich M, Suarez-Cisneros H, Rozman M, Villamor N, Colomer D, Lopez-Guillermo A, Gonzalez M, Alcoceba M, Terol MJ, Colado E, Puente XS, Lopez-Otin C, Enjuanes A and Campo E (2016) Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1 and ATM mutations in chronic lymphocytic leukemia. *Blood*. American Society of Hematology blood-2015-07-659144.

Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, Jares P, Navarro A, Martin-Garcia D, Bea S, Salaverría I, Oldreive C, Aymerich M, Suarez-Cisneros H, Rozman M, Villamor N, Colomer D, Lopez-Guillermo A, Gonzalez M, Alcoceba M, Terol MJ, Colado E, Puente XS, Lopez-Otin C, Enjuanes A, Campo E, Martín-García D, Beà S, Suárez-Cisneros H, López-Guillermo A, González M and López-Otín C (2016) Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood*. American Society of Hematology 127(17): 2122–30.

Neilson JR, Auer R, White D, Bienz N, Waters JJ, Whittaker JA, Milligan DW and Fegan CD (1997) Deletions at 11q identify a subset of patients with typical CLL who show consistent disease progression and reduced survival. *Leukemia* 11(11): 1929–32.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC and Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic acids research* 17(7): 2503–16.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt A, Martens JWM, Aparicio SAJR, Borg Å, Salomon AV, Thomas G, Børresen-Dale A-L, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ and Stratton MR (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5): 979–93.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn S-M, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GKJ, Jang SJ, Jones DR, Kim H-Y, King TA, Krishnamurthy S, Lee HJ, Lee J-Y, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L,

Caldas C, Veer L van't, Tutt A, Knappskog S, Tan BKT, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A-L, Richardson AL, Kong G, Thomas G and Stratton MR (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. Nature Publishing Group.

Nückel H, Switala M, Collins CH, Sellmann L, Grosse-Wilde H, Dührsen U and Rebmann V (2009) High CD49d protein and mRNA expression predicts poor outcome in chronic lymphocytic leukemia. *Clinical immunology (Orlando, Fla.)* 131(3): 472–80.

O'Brien S, Furman RR, Coutre SE, Sharman JP, Burger JA, Blum KA, Grant B, Richards DA, Coleman M, Wierda WG, Jones JA, Zhao W, Heerema NA, Johnson AJ, Izumi R, Hamdy A, Chang BY, Graef T, Clow F, Buggy JJ, James DF and Byrd JC (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): 48–58.

Oancea C, Rüster B, Henschler R, Puccetti E and Ruthardt M (2010) The t(6;9) associated DEK/CAN fusion protein targets a population of long-term repopulating hematopoietic stem cells for leukemogenic transformation. *Leukemia*. Macmillan Publishers Limited 24(11): 1910–9.

Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, Mino-Kenudson M, Lauwers GY, Loda M and Fuchs CS (2005) Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *The Journal of molecular diagnostics : JMD* 7(3): 413–21.

Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, Romero-Cordoba S, Treviño V, Vazquez-Santillan K, Guadarrama AS, Wright AA, Rosenberg MW, Duke F, Kaplan B, Wang R, Nickerson E, Walline HM, Lawrence MS, Stewart C, Carter SL, McKenna A, Rodriguez-Sanchez IP, Espinosa-Castilla M, Woie K, Bjorge L, Wik E, Halle MK, Hoivik EA, Krakstad C, Gabiño NB, Gómez-Macías GS, Valdez-Chapa LD, Garza-Rodríguez ML, Maytorena G, Vazquez J, Rodea C, Cravioto A, Cortes ML, Greulich H, Crum CP, Neuberg DS, Hidalgo-Miranda A, Escareno CR, Akslen LA, Carey TE, Vintermyr OK, Gabriel SB, Barrera-Saldaña HA, Melendez-Zajgla J, Getz G, Salvesen HB and Meyerson M (2014) Landscape of genomic alterations in cervical carcinomas. *Nature* 506(7488): 371–5.

Oscier D, Wade R, Davis Z, Morilla A, Best G, Richards S, Else M, Matutes E and Catovsky D (2010) Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica* 95(10): 1705–12.

Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, Behre G, Hiddemann W and Tenen DG (2001) Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha

(C/EBPalpha), in acute myeloid leukemia. *Nature genetics* 27(3): 263–70.

Packham D, Ward RL, Ap Lin V, Hawkins NJ and Hitchins MP (2009) Implementation of novel pyrosequencing assays to screen for common mutations of BRAF and KRAS in a cohort of sporadic colorectal cancers. *Diagnostic molecular pathology : the American journal of surgical pathology, part B* 18(2): 62–71.

Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, Godfrey AL, Rapado I, Cvejic A, Rance R, McGee C, Ellis P, Mudie LJ, Stephens PJ, McLaren S, Massie CE, Tarpey PS, Varela I, Nik-Zainal S, Davies HR, Shlien A, Jones D, Raine K, Hinton J, Butler AP, Teague JW, Baxter EJ, Score J, Galli A, Della Porta MG, Travaglino E, Groves M, Tauro S, Munshi NC, Anderson KC, El-Naggar A, Fischer A, Mustonen V, Warren AJ, Cross NC, Green AR, Futreal PA, Stratton MR and Campbell PJ (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts (2011/10/15.). *The New England journal of medicine* 365(15): 1384–1395.

Paschka P, Marcucci G, Ruppert AS, Whitman SP, Mrózek K, Maharry K, Langer C, Baldus CD, Zhao W, Powell BL, Baer MR, Carroll AJ, Caligiuri MA, Kolitz JE, Larson RA and Bloomfield CD (2008) Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26(28): 4595–602.

Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse HC, Nussenzweig MC and Dalla-Favera R (2008) AID is required for germinal center-derived lymphomagenesis. *Nature genetics* 40(1): 108–12.

Pasqualucci L, Neri A, Baldini L, Dalla-Favera R and Migliazza A (2000) BCL-6 mutations are associated with immunoglobulin variable heavy chain mutations in B-cell chronic lymphocytic leukemia. *Cancer research* 60(20): 5644–8.

Patel JP, Gönen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, Van Vlierberghe P, Dolgalev I, Thomas S, Aminova O, Huberman K, Cheng J, Viale A, Socci ND, Heguy A, Cherry A, Vance G, Higgins RR, Ketterling RP, Gallagher RE, Litzow M, van den Brink MRM, Lazarus HM, Rowe JM, Luger S, Ferrando A, Paietta E, Tallman MS, Melnick A, Abdel-Wahab O and Levine RL (2012) Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *The New England journal of medicine* 366(12): 1079–89.

Pekarsky Y, Zanesi N and Croce CM (2010) Molecular basis of CLL. *Seminars in cancer biology* 20(6): 370–6.

Pellagatti A and Boultwood J (2015) The molecular pathogenesis of the myelodysplastic syndromes. *European Journal of Haematology* 95(1): 3–15.

Pettitt AR, Jackson R, Carruthers S, Dodd J, Dodd S, Oates M, Johnson GG, Schuh A, Matutes E, Dearden CE, Catovsky D, Radford JA, Bloor A, Follows GA, Devereux S, Kruger A, Blundell J, Agrawal S, Allsup D, Proctor S, Heartin E, Oscier D, Hamblin TJ, Rawstron A and Hillmen P (2012) Alemtuzumab in combination with methylprednisolone is a highly effective induction regimen for patients with chronic lymphocytic leukemia and deletion of TP53: final results of the national cancer research institute CLL206 trial (2012/04/12.). *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30(14): 1647–1655.

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, Bauer K, Malchau G, Rabe KG, Stilgenbauer S, Döhner H, Jäger U, Eckart MJ, Hopfinger G, Busch R, Fink A-M, Wendtner C-M, Fischer K, Kay NE and Hallek M (2014) Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood*. American Society of Hematology 124(1): 49–62.

Pratt G, Thomas P, Marden N, Alexander D, Davis Z, Hussey D, Parry H, Harding S, Catovsky D, Begley J and Oscier D (2016) Evaluation of serum markers in the LRF CLL4 trial: β2-microglobulin but not serum free light chains, is an independent marker of overall survival. *Leukemia & lymphoma* 1–9.

Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrio A, Castellano G, Clot G, Colado E, Colomer D, Costa D, Delgado J, Enjuanes A, Estivill X, Ferrando AA, Gelpí JL, González B, González S, González M, Gut M, Hernández-Rivas JM, López-Guerra M, Martín-García D, Navarro A, Nicolás P, Orozco M, Payer ÁR, Pinyol M, Pisano DG, Puente DA, Queirós AC, Quesada V, Romeo-Casabona CM, Royo C, Royo R, Rozman M, Russiñol N, Salaverría I, Stamatopoulos K, Stunnenberg HG, Tamborero D, Terol MJ, Valencia A, López-Bigas N, Torrents D, Gut I, López-Guillermo A, López-Otín C and Campo E (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 526(7574): 519–24.

Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JMC, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente DA, Freije JMP, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, Hernández L, Yagüe J, Nicolás P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjosé S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpí JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigó R, Bayés M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR, Campbell PJ, Gut I, López-Guillermo A, Estivill X, Montserrat E, López-Otín C, Campo E, Ordonez GR, Bea S, Gonzalez-Diaz M, Lopez-Guerra M, Tubio JM, Lopez C, Hernandez JM, Gutierrez-Fernandez A, Carrio A, Hernandez L, Yague J, Nicolas P, de Sanjose S, Gelpi

JL, Guigo R, Bayes M, Lopez-Guillermo A and Lopez-Otin C (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia (2011/06/07.). *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. 475(7354): 101–105.

Puiggros A, Delgado J, Rodriguez-Vicente A, Collado R, Aventín A, Luño E, Grau J, Hernandez JÁ, Marugán I, Ardanaz M, González T, Valiente A, Osma M, Calasanz MJ, Sanzo C, Carrió A, Ortega M, Santacruz R, Abrisqueta P, Abella E, Bosch F, Carbonell F, Solé F, Hernández JM and Espinet B (2013) Biallelic losses of 13q do not confer a poorer outcome in chronic lymphocytic leukaemia: analysis of 627 patients with isolated 13q deletion. *British journal of haematology* 163(1): 47–54.

Qu Y, Lennartsson A, Gaidzik VI, Deneberg S, Karimi M, Bengtzén S, Höglund M, Bullinger L, Döhner K and Lehmann S (2014) Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics* 9(8): 1108–19.

Que D, Xiao H, Zhao B, Zhang X, Wang Q, Xiao H and Wang G (2016) EGFR mutation status in plasma and tumor tissues in non-small cell lung cancer serves as a predictor of response to EGFR-TKI treatment. *Cancer biology & therapy* 17(3): 320–7.

Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E and Lopez-Otin C (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia (2011/12/14.). *Nature genetics* 44(1): 47–52.

Quesada V, Ramsay AJ, Rodríguez D, Puente XS, Campo E and López-Otín C (2013) The genomic landscape of chronic lymphocytic leukemia: clinical implications. *BMC medicine* 11: 124.

Rachakonda PS, Hosen I, de Verdier PJ, Fallah M, Heidenreich B, Ryk C, Wiklund NP, Steineck G, Schadendorf D, Hemminki K and Kumar R (2013) TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proceedings of the National Academy of Sciences of the United States of America* 110(43): 17426–31.

Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, Chuang H-Y, Källberg M, Kumar SA, Liao A, Little KM, Strömberg MP and Tanner SW (2013) Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics (Oxford, England)* 29(16):

2041–3.

Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, Besenbacher S, Lundin P, Stacey SN, Gudmundsson J, Magnusson OT, le Roux L, Orlygsdottir G, Helgadottir HT, Johannsdottir H, Gylfason A, Tryggvadottir L, Jonasson JG, de Juan A, Ortega E, Ramon-Cajal JM, García-Prats MD, Mayordomo C, Panadero A, Rivera F, Aben KKH, van Altena AM, Massuger LFAG, Aavikko M, Kujala PM, Staff S, Aaltonen LA, Olafsdottir K, Bjornsson J, Kong A, Salvarsdottir A, Saemundsson H, Olafsson K, Benediktsdottir KR, Gulcher J, Masson G, Kiemeney LA, Mayordomo JI, Thorsteinsdottir U and Stefansson K (2011) Mutations in BRIP1 confer high risk of ovarian cancer. *Nature genetics* 43(11): 1104–7.

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN and Pasternack BS (1975) Clinical staging of chronic lymphocytic leukemia. *Blood*. American Society of Hematology 46(2): 219–34.

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, Li Y, Ahn J, Abdel-Wahab O, Shih A, Lu C, Ward PS, Tsai JJ, Hricik T, Tosello V, Tallman JE, Zhao X, Daniels D, Dai Q, Ciminio L, Aifantis I, He C, Fuks F, Tallman MS, Ferrando A, Nimer S, Paietta E, Thompson CB, Licht JD, Mason CE, Godley LA, Melnick A, Figueroa ME and Levine RL (2014) DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell reports* 9(5): 1841–55.

Rassenti LZ, Huynh L, Toy TL, Chen L, Keating MJ, Gribben JG, Neuberg DS, Flinn IW, Rai KR, Byrd JC, Kay NE, Greaves A, Weiss A and Kipps TJ (2004) ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *The New England journal of medicine*. Massachusetts Medical Society 351(9): 893–901.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V and Korbel JO (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England)* 28(18): i333–i339.

Ravandi F, Cortes JE, Jones D, Faderl S, Garcia-Manero G, Konopleva MY, O'Brien S, Estrov Z, Borthakur G, Thomas D, Pierce SR, Brandt M, Byrd A, Bekele BN, Pratz K, Luthra R, Levis M, Andreeff M and Kantarjian HM (2010) Phase I/II study of combination therapy with sorafenib, idarubicin, and cytarabine in younger patients with acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28(11): 1856–62.

Reiter E, Greinix H, Rabitsch W, Keil F, Schwarzinger I, Jaeger U, Lechner K, Worel N, Streubel B, Fonatsch C, Mitterbauer G and Kalhs P (2000) Low curative potential of bone marrow transplantation for highly aggressive acute myelogenous leukemia with inversioin inv (3)(q21q26) or homologous translocation t(3;3) (q21;q26). *Annals of hematology* 79(7): 374–7.

Renneville A, Boissel N, Gachard N, Naguib D, Bastard C, de Botton S, Nibourel O, Pautas C, Reman O, Thomas X, Gardin C, Terré C, Castaigne S, Preudhomme C and Dombret H (2009) The favorable impact of CEBPA mutations in patients with acute myeloid leukemia is only observed in the absence of associated cytogenetic abnormalities and FLT3 internal duplication. *Blood* 113(21): 5090–3.

Renneville A, Boissel N, Nibourel O, Berthon C, Helevaut N, Gardin C, Cayuela J-M, Hayette S, Reman O, Contentin N, Bordessoule D, Pautas C, Botton S de, Revel T de, Terre C, Fenaux P, Thomas X, Castaigne S, Dombret H and Preudhomme C (2012) Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia* 26(6): 1247–54.

Renneville A, Boissel N, Zurawski V, Llopis L, Biggio V, Nibourel O, Philippe N, Thomas X, Dombret H and Preudhomme C (2009) Wilms tumor 1 gene mutations are associated with a higher risk of recurrence in young adults with acute myeloid leukemia: a study from the Acute Leukemia French Association. *Cancer* 115(16): 3719–27.

Rice GI, Bond J, Asipu A, Brunette RL, Manfield IW, Carr IM, Fuller JC, Jackson RM, Lamb T, Briggs TA, Ali M, Gornall H, Couthard LR, Aeby A, Attard-Montalto SP, Bertini E, Bodemer C, Brockmann K, Brueton LA, Corry PC, Desguerre I, Fazzi E, Cazorla AG, Gener B, Hamel BCJ, Heiberg A, Hunter M, van der Knaap MS, Kumar R, Lagae L, Landrieu PG, Lourenco CM, Marom D, McDermott MF, van der Merwe W, Orcesi S, Prendiville JS, Rasmussen M, Shalev SA, Soler DM, Shinawi M, Spiegel R, Tan TY, Vanderver A, Wakeling EL, Wassmer E, Whittaker E, Lebon P, Stetson DB, Bonthron DT and Crow YJ (2009) Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature genetics* 41(7): 829–32.

Rickert RC (2013) New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nature Reviews Immunology*. Nature Research 13(8): 578–591.

Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, McVean G and Lunter G (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics* 46(8): 912–8.

Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, Bolland DJ, Chen HT, Corcoran AE, Nussenzweig A and Nussenzweig MC (2008) AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 135(6): 1028–38.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP (2011) Integrative genomics viewer. *Nature Biotechnology*. Nature Research 29(1): 24–26.

Ronaghi M (1998) DNA SEQUENCING:A Sequencing Method Based on Real-Time Pyrophosphate. *Science* 281(5375): 363–365.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M and Nyrén P (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry* 242(1): 84–9.

Rosati E, Sabatini R, Rampino G, Tabilio A, Di Ianni M, Fettucciari K, Bartoli A, Coaccioli S, Screpanti I and Marconi P (2009) Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* 113(4): 856–65.

Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, Guarini A, Del Giudice I, Cerri M, Cresta S, Deambrogi C, Gargiulo E, Gattei V, Forconi F, Bertoni F, Deaglio S, Rabadan R, Pasqualucci L, Foà R, Dalla-Favera R and Gaidano G (2011) Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: Association with progression and fludarabine-refractoriness (2011/11/01.). *Blood* 118(26): 6904–8.

Rossi D, Cerri M, Deambrogi C, Sozzi E, Cresta S, Rasi S, De Paoli L, Spina V, Gattei V, Capello D, Forconi F, Lauria F and Gaidano G (2009) The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness. *Clinical cancer research : an official journal of the American Association for Cancer Research* 15(3): 995–1004.

Rossi D, Fangazio M, Rasi S, Vaisitti T, Monti S, Cresta S, Chiaretti S, Del Giudice I, Fabbri G, Bruscaggin A, Spina V, Deambrogi C, Marinelli M, Famà R, Greco M, Daniele G, Forconi F, Gattei V, Bertoni F, Deaglio S, Pasqualucci L, Guarini A, Dalla-Favera R, Foà R and Gaidano G (2012) Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood*. American Society of Hematology 119(12): 2854–62.

Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R, Rasi S, Monti S, Deambrogi C, De Paoli L, Wang J, Gattei V, Guarini A, Foà R, Rabadan R and Gaidano G (2014) Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* 123(14): 2139–47.

Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, Marasca R, Laurenti L, Bruscaggin A, Cerri M, Monti S, Cresta S, Fama R, De Paoli L, Bulian P, Gattei V, Guarini A, Deaglio S, Capello D, Rabadan R, Pasqualucci L, Dalla-Favera R, Foa R and Gaidano G (2011) Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. (2011/11/15.). *Blood* 119(2): 521–529.

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, Deambrogi C, Khiabanian H, Serra R, Bertoni F, Forconi F, Laurenti L, Marasca R, Dal-Bo M, Rossi FM, Bulian P, Nomdedeu J, Del Poeta G, Gattei V, Pasqualucci L, Rabadan R, Foà R, Dalla-Favera R and Gaidano G (2013) Integrated

mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 121(8): 1403–12.

Rossi D, Rasi S, Spina V, Fangazio M, Monti S, Greco M, Ciardullo C, Fama R, Cresta S, Bruscaggin A, Laurenti L, Martini M, Musto P, Forconi F, Marasca R, Larocca LM, Foa R and Gaidano G (2012) Different impact of NOTCH1 and SF3B1 mutations on the risk of chronic lymphocytic leukemia transformation to Richter syndrome (2012/05/11.). *British journal of haematology*.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT and Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing (2011/07/22.). *Nature* 475(7356): 348–352.

el Rouby S, Thomas A, Costin D, Rosenberg CR, Potmesil M, Silber R and Newcomb EW (1993) p53 gene mutation in B-cell chronic lymphocytic leukemia is associated with drug resistance and is independent of MDR1/MDR3 gene expression. *Blood* 82(11): 3452–9.

Rubnitz JE, Raimondi SC, Tong X, Srivastava DK, Razzouk BI, Shurtleff SA, Downing JR, Pui C-H, Ribeiro RC and Behm FG (2002) Favorable impact of the t(9;11) in childhood acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 20(9): 2302–9.

Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK and Ley TJ (2014) The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer cell* 25(4): 442–54.

Russo G, Patrignani A, Poveda L, Hoehn F, Scholtka B, Schlapbach R and Garvin AM (2015) Highly sensitive, non-invasive detection of colorectal cancer mutations using single molecule, third generation sequencing. *Applied & translational genomics* 7: 32–9.

Samorodnitsky E, Datta J, Jewell BM, Hagopian R, Miya J, Wing MR, Damodaran S, Lippus JM, Reeser JW, Bhatt D, Timmers CD and Roychowdhury S (2015) Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing. *The Journal of Molecular Diagnostics* 17(1): 64–75.

Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, Damodaran S, Bhatt D, Reeser JW, Datta J and Roychowdhury S (2015) Evaluation of Hybridization Capture Versus Amplicon-Based Methods for

Whole-Exome Sequencing. *Human Mutation* 36(9): 903–914.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM and Smith M (1977) Nucleotide sequence of bacteriophage phi X174 DNA (1977/02/24.). *Nature* 265(5596): 687–695.

Sanger F and Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94(3): 441–448.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors (1977/12/01.). *Proceedings of the National Academy of Sciences of the United States of America* 74(12): 5463–7.

Sanz MA, Montesinos P, Kim HT, Ruiz-Argüelles GJ, Undurraga MS, Uriarte MR, Martínez L, Jacomo RH, Gutiérrez-Aguirre H, Melo RAM, Bittencourt R, Pasquini R, Pagnano K, Fagundes EM, Vellenga E, Holowiecka A, González-Huerta AJ, Fernández P, De la Serna J, Brunet S, De Lisa E, González-Campos J, Ribera JM, Krsnik I, Ganser A, Berliner N, Ribeiro RC, Lo-Coco F, Löwenberg B and Rego EM (2015) All-trans retinoic acid with daunorubicin or idarubicin for risk-adapted treatment of acute promyelocytic leukaemia: a matched-pair analysis of the PETHEMA LPA-2005 and IC-APL studies. *Annals of hematology* 94(8): 1347–56.

Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ and Cheetham RK (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28(14): 1811–1817.

Schaffner C, Stilgenbauer S, Rappold GA, Döhner H and Lichter P (1999) Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. *Blood*. American Society of Hematology 94(2): 748–53.

Scharenberg C, Giai V, Pellagatti A, Saft L, Dimitriou M, Jansson M, Jadersten M, Grandien A, Douagi I, Neuberg DS, LeBlanc K, Boultwood J, Karimi M, Jacobsen SEW, Woll PS and Hellstrom-Lindberg E (2016) Progression in patients with low- and intermediate1-risk del(5q) myelodysplastic syndromes is predicted by a limited subset of mutations. *Haematologica*.

Schlenk RF, Döhner K, Kneba M, Götze K, Hartmann F, Del Valle F, Kirchen H, Koller E, Fischer JT, Bullinger L, Habdank M, Späth D, Groner S, Krebs B, Kayser S, Corbacioglu A, Anhalt A, Benner A, Fröhling S and Döhner H (2009) Gene mutations and response to treatment with all-trans retinoic acid in elderly patients with acute myeloid leukemia. Results from the AMLSG Trial AML HD98B. *Haematologica* 94(1): 54–60.

Schlenk RF, Döhner K, Krauter J, Fröhling S, Corbacioglu A, Bullinger L, Habdank M, Späth D, Morgan M, Benner A, Schlegelberger B, Heil G, Ganser A and Döhner H (2008) Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *The New England journal*

of medicine 358(18): 1909–18.

Schnaiter A, Paschka P, Rossi M, Zenz T, Bühler A, Winkler D, Cazzola M, Döhner K, Edelmann J, Mertens D, Kless S, Mack S, Busch R, Hallek M, Döhner H and Stilgenbauer S (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): 1266–70.

Schneider F, Hoster E, Unterhalt M, Schneider S, Dufour A, Benthaus T, Mellert G, Zellmeier E, Bohlander SK, Feuring-Buske M, Buske C, Braess J, Fritsch S, Heinecke A, Sauerland MC, Berdel WE, Buechner T, Woermann BJ, Hiddemann W and Spiekermann K (2009) NPM1 but not FLT3-ITD mutations predict early blast cell clearance and CR rate in patients with normal karyotype AML (NK-AML) or high-risk myelodysplastic syndrome (MDS). *Blood* 113(21): 5250–3.

Schnittger S, Dicker F, Kern W, Wendland N, Sundermann J, Alpermann T, Haferlach C and Haferlach T (2011) RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood* 117(8).

Schnittger S, Eder C, Jeromin S, Alpermann T, Fasan A, Grossmann V, Kohlmann A, Illig T, Klopp N, Wichmann H-E, Kreuzer K-A, Schmid C, Staib P, Peceny R, Schmitz N, Kern W, Haferlach C and Haferlach T (2013) ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *Leukemia* 27(1): 82–91.

Schnittger S, Haferlach C, Ulke M, Alpermann T, Kern W and Haferlach T (2010) IDH1 mutations are detected in 6.6% of 1414 AML patients and are associated with intermediate risk karyotype and unfavorable prognosis in adults younger than 60 years and unmutated NPM1 status. *Blood* 116(25): 5486–96.

Schnittger S, Schoch C, Kern W, Mecucci C, Tschulik C, Martelli MF, Haferlach T, Hiddemann W and Falini B (2005) Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. *Blood* 106(12): 3733–9.

Scholl S, Mugge LO, Landt O, Loncarevic IF, Kunert C, Clement JH and Hoffken K (2007) Rapid screening and sensitive detection of NPM1 (nucleophosmin) exon 12 mutations in acute myeloid leukaemia (2007/02/20.). *Leukemia research* 31(9): 1205–1211.

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 MT, Taylor J and Bentley D (2012) Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns (2012/08/24.). *Blood* 120(20): 4191–4196.

Sciumè M, Vincenti D, Reda G, Orofino N, Cassin R, Giannarelli D, Gaidano G, Rossi D and Cortelezzi A (2015) Low-dose alemtuzumab in refractory/relapsed chronic lymphocytic leukemia: Genetic profile and long-term outcome from a single center experience. *American journal of hematology* 90(11): 970–4.

Sebaa A, Ades L, Baran-Marzack F, Mozziconacci M-J, Penther D, Dobbelstein S, Stamatoullas A, Récher C, Prebet T, Moulessehoul S, Fenaux P and Eclache V (2012) Incidence of 17p deletions and TP53 mutation in myelodysplastic syndrome and acute myeloid leukemia with 5q deletion. *Genes, chromosomes & cancer* 51(12): 1086–92.

Sehgal AR, Gimotty PA, Zhao J, Hsu J-M, Daber R, Morrissette JD, Luger S, Loren AW and Carroll M (2015) DNMT3A Mutational Status Affects the Results of Dose-Escalated Induction Therapy in Acute Myelogenous Leukemia. *Clinical cancer research : an official journal of the American Association for Cancer Research* 21(7): 1614–20.

Seiler T, Döhner H and Stilgenbauer S (2006) Risk stratification in chronic lymphocytic leukemia. *Seminars in oncology* 33(2): 186–94.

Shanafelt TD, Geyer SM, Bone ND, Tschumper RC, Witzig TE, Nowakowski GS, Zent CS, Call TG, Laplant B, Dewald GW, Jelinek DF and Kay NE (2008) CD49d expression is an independent predictor of overall survival in patients with chronic lymphocytic leukaemia: a prognostic parameter with therapeutic potential. *British journal of haematology* 140(5): 537–46.

Shayegi N, Kramer M, Bornhäuser M, Schaich M, Schetelig J, Platzbecker U, Röllig C, Heiderich C, Landt O, Ehninger G and Thiede C (2013) The level of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and survival in AML. *Blood* 122(1): 83–92.

Shivarov V, Gueorguieva R, Stoimenov A and Tiu R (2013) DNMT3A mutation is a poor prognosis biomarker in AML: results of a meta-analysis of 4500 AML patients. *Leukemia research* 37(11): 1445–50.

Shivarov V, Ivanova M and Naumova E (2014) Rapid Detection of DNMT3A R882 Mutations in Hematologic Malignancies Using a Novel Bead-Based Suspension Assay with BNA(NC) Probes. *PLoS ONE*. Public Library of Science 9(6): e99769.

Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, McLeod JL, Doedens M, Medeiros JJF, Marke R, Kim HJ, Lee K, McPherson JD, Hudson TJ, Pan-Leukemia Gene Panel Consortium TH, Brown AMK, Trinh QM, Stein LD, Minden MD, Wang JCY, Dick JE and Dick JE (2014) Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. Nature Research 506(7488): 328–333.

Siegel R, Naishadham D and Jemal A (2012) Cancer statistics, 2012. *CA: a cancer journal for clinicians* 62(1): 10–29.

Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S, Dyer M, Matutes E, Gonzalez D, Taylor AMR, Moss P, Thomas P, Oscier D and Stankovic T (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 : official journal of the American Society of Clinical Oncology* 30(36): 4524–32.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB and Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321(6071): 674–9.

Smith TFF and Waterman MSS (1981) Identification of common molecular subsequences (1981/03/25.). *Journal of molecular biology*. Academic Press 147(1): 195–197.

Snaddon J, Smith ML, Neat M, Cambal-Parrales M, Dixon-McIver A, Arch R, Amess JA, Rohatiner AZ, Lister TA and Fitzgibbon J (2003) Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. *Genes, chromosomes & cancer* 37(1): 72–8.

Soekarman D, von Lindern M, Daenen S, de Jong B, Fonatsch C, Heinze B, Bartram C, Hagemeijer A and Grosveld G (1992) The translocation (6;9) (p23;q34) shows consistent rearrangement of two genes and defines a myeloproliferative disorder with specific clinical features. *Blood* 79(11): 2990–7.

Soekarman D, von Lindern M, van der Plas DC, Selleri L, Bartram CR, Martiat P, Culligan D, Padua RA, Hasper-Voogt KP and Hagemeijer A (1992) Dek-can rearrangement in translocation (6;9)(p23;q34). *Leukemia* 6(6): 489–94.

Solh M, Yohe S, Weisdorf D and Ustun C (2014) Core-binding factor acute myeloid leukemia: Heterogeneity, monitoring, and therapy. *American journal of hematology* 89(12): 1121–31.

Spencer DH, Abel HJ, Lockwood CM, Payton JE, Szankasi P, Kelley TW, Kulkarni S, Pfeifer JD and Duncavage EJ (2013) Detection of FLT3 Internal Tandem Duplication in Targeted, Short-Read-Length, Next-Generation Sequencing Data. *The Journal of Molecular Diagnostics* 15(1): 81–93.

Sportoletti P, Baldoni S, Cavalli L, Del Papa B, Bonifacio E, Ciurnelli R, Bell AS, Di Tommaso A, Rosati E, Crescenzi B, Mecucci C, Screpanti I, Marconi P, Martelli MF, Di Ianni M and Falzetti F (2010) NOTCH1 PEST domain mutation is an adverse prognostic factor in B-CLL. *British journal of haematology* 151(4): 404–6.

Stamatopoulos B, Haibe-Kains B, Equeter C, Meuleman N, Sorée A, De Bruyn C, Hanosset D, Bron D, Martiat P and Lagneaux L (2009) Gene expression profiling reveals differences in microenvironment interaction between patients with chronic lymphocytic leukemia expressing high versus low ZAP70 mRNA. *Haematologica* 94(6): 790–9.

Stamatopoulos B, Timbs A, Bruce D, Smith T, Clifford R, Robbe P, Burns A, Vavoulis D V, Lopez L, Antoniou P, Mason J, Dreau H and Schuh A (2016) Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia* (October): 1–30.

Stankovic T, Weber P, Stewart G, Bedenham T, Murray J, Byrd PJ, Moss PA and Taylor AM (1999) Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. *Lancet (London, England)* 353(9146): 26–9.

Steensma DP, Higgs DR, Fisher CA and Gibbons RJ (2004) Acquired somatic ATRX mutations in myelodysplastic syndrome associated with ?? thalassemia (ATMDS) convey a more severe hematologic phenotype than germline ATRX mutations. *Blood* 103(6): 2019–2026.

Stieglitz E, Troup CB, Gelston LC, Haliburton J, Chow ED, Yu KB, Akutagawa J, Taylor-Weiner AN, Liu YL, Wang Y-D, Beckman K, Emanuel PD, Braun BS, Abate A, Gerbing RB, Alonzo TA and Loh ML (2015) Subclonal mutations in SETBP1 confer a poor prognosis in juvenile myelomonocytic leukemia. *Blood*. American Society of Hematology 125(3): 516–24.

Stilgenbauer S, Liebisch P, James MR, Schröder M, Schlegelberger B, Fischer K, Bentz M, Lichter P and Döhner H (1996) Molecular cytogenetic delineation of a novel critical genomic region in chromosome bands 11q22.3-923.1 in lymphoproliferative disorders. *Proceedings of the National Academy of Sciences of the United States of America* 93(21): 11837–41.

Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, Bühler A, Böttcher S, Ritgen M, Kneba M, Winkler D, Tausch E, Hoth P, Edelmann J, Mertens D, Bullinger L, Bergmann M, Kless S, Mack S, Jäger U, Patten N, Wu L, Wenger MK, Fingerle-Rowson G, Lichter P, Cazzola M, Wendtner CM, Fink AM, Fischer K, Busch R, Hallek M and Döhner H (2014) Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 123(21): 3247–54.

Stirewalt DL and Radich JP (2003) The role of FLT3 in haematopoietic malignancies. *Nature Reviews Cancer* 3(9): 650–665.

Stone RM, Fischer T, Paquette R, Schiller G, Schiffer CA, Ehninger G, Cortes J, Kantarjian HM, DeAngelo DJ, Huntsman-Labed A, Dutreix C, del Corral A and Giles F (2012) Phase IB study of the FLT3 kinase inhibitor midostaurin with chemotherapy in younger newly diagnosed adult patients with acute myeloid leukemia. *Leukemia* 26(9): 2061–8.

Summers K, Stevens J, Kakkas I, Smith M, Smith LL, Macdougall F, Cavenagh J, Bonnet D, Young BD, Lister TA and Fitzgibbon J (2007) Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia* 21(3): 550–1; author reply 552.

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J and Vardiman JW (2008) WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition. .

Tallman MS, Andersen JW, Schiffer CA, Appelbaum FR, Feusner JH, Ogden A, Shepherd L, Willman C, Bloomfield CD, Rowe JM and Wiernik PH (1997) All-trans-retinoic acid in acute promyelocytic leukemia. *The New England journal of medicine* 337(15): 1021–8.

Tam CS, O'Brien S, Plunkett W, Wierda W, Ferrajoli A, Wang X, Do K-AA, Cortes J, Khouri I, Kantarjian H, Lerner S, Keating MJ, O'Brien S, Plunkett W, Wierda W, Ferrajoli A, Wang X, Do K-AA, Cortes J, Khouri I, Kantarjian H, Lerner S, Keating MJ, O'Brien S, Plunkett W, Wierda W, Ferrajoli A, Wang X, Do K-AA, Cortes J, Khouri I, Kantarjian H, Lerner S and Keating MJ (2014) Long-term results of first salvage treatment in CLL patients treated initially with FCR (fludarabine, cyclophosphamide, rituximab). *Blood*. American Society of Hematology 124(20): 3059–3064.

Tang J-LL, Hou H-AA, Chen C-YY, Liu C-YY, Chou W-CC, Tseng M-HH, Huang C-FF, Lee F-YY, Liu M-CC, Yao M, Huang S-YY, Ko B-SS, Hsu S-CC, Wu S-JJ, Tsay W, Chen Y-CC, Lin L-II and Tien H-FF (2009) AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations (2009/10/08.). *Blood* 114(26): 5352–5361.

Thiede C, Koch S, Creutzig E, Steudel C, Illmer T, Schaich M and Ehninger G (2006) Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood* 107(10): 4011–20.

Thiede C, Prange-Krex G, Freiberg-Richter J, Bornhäuser M and Ehninger G (2000) Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. *Bone Marrow Transplantation*. Nature Publishing Group 25(5): 575–577.

Thiede C, Steudel C, Mohr B, Schaich M, Schäkel U, Platzbecker U, Wermke M, Bornhäuser M, Ritter M, Neubauer A, Ehninger G and Illmer T (2002) Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood* 99(12): 4326–35.

Thol F, Damm F, Lüdeking A, Winschel C, Wagner K, Morgan M, Yun H, Göhring G, Schlegelberger B, Hoelzer D, Lübbert M, Kanz L, Fiedler W, Kirchner H, Heil G, Krauter J, Ganser A and Heuser M (2011) Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29(21): 2889–96.

Thol F, Damm F, Wagner K, Göhring G, Schlegelberger B, Hoelzer D, Lübbert M, Heit W, Kanz L, Schlimok G, Raghavachar A, Fiedler W, Kirchner H, Heil G, Heuser M, Krauter J and Ganser A (2010) Prognostic impact of IDH2

mutations in cytogenetically normal acute myeloid leukemia. *Blood* 116(4): 614–6.

Thol F, Kölking B, Damm F, Reinhardt K, Klusmann J-H, Reinhardt D, von Neuhoff N, Brugman MH, Schlegelberger B, Suerbaum S, Krauter J, Ganser A and Heuser M (2012) Next-generation sequencing for minimal residual disease monitoring in acute myeloid leukemia patients with FLT3-ITD or NPM1 mutations. *Genes, Chromosomes and Cancer*. Wiley Subscription Services, Inc., A Wiley Company 51(7): 689–695.

Thorsélius M, Kröber A, Murray F, Thunberg U, Tobin G, Bühler A, Kienle D, Albesiano E, Maffei R, Dao-Ung L-P, Wiley J, Vilpo J, Laurell A, Merup M, Roos G, Karlsson K, Chiorazzi N, Marasca R, Döhner H, Stilgenbauer S and Rosenquist R (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): 2889–94.

Tobin G, Thunberg U, Johnson A, Thörn I, Söderberg O, Hultdin M, Botling J, Enblad G, Sällström J, Sundström C, Roos G and Rosenquist R (2002) Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood* 99(6): 2262–4.

Trapnell C, Pachter L and Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* 25(9): 1105–11.

Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR and Murphy KM (2010) Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *The Journal of molecular diagnostics : JMD*. American Society for Investigative Pathology 12(4): 425–32.

Vaisitti T, Aydin S, Rossi D, Cottino F, Bergui L, D'Arena G, Bonello L, Horenstein AL, Brennan P, Pepper C, Gaidano G, Malavasi F and Deaglio S (2010) CD38 increases CXCL12-mediated signals and homing of chronic lymphocytic leukemia cells. *Leukemia* 24(5): 958–69.

Vannucchi AM, Lasho TL, Guglielmelli P, Biamonte F, Pardanani A, Pereira A, Finke C, Score J, Gangat N, Mannarelli C, Ketterling RP, Rotunno G, Knudson RA, Susini MC, Laborde RR, Spolverini A, Pancrazzi A, Pieri L, Manfredini R, Tagliafico E, Zini R, Jones A, Zoi K, Reiter A, Duncombe A, Pietra D, Rumi E, Cervantes F, Barosi G, Cazzola M, Cross NCP and Tefferi A (2013) Mutations and prognosis in primary myelofibrosis. *Leukemia* 27(9): 1861–9.

Vardiman JW, Harris NL and Brunning RD (2002) The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. American Society of Hematology 100(7): 2292–302.

Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellström-Lindberg E, Tefferi A and Bloomfield CD

(2009) The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. American Society of Hematology 114(5): 937–51.

Venter JC et al (2001) The sequence of the human genome (2001/02/22.). *Science* 291(5507): 1304–1351.

Verhaak RGW, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, Uitterlinden AG, Erpelinck CAJ, Delwel R, Löwenberg B and Valk PJM (2005) Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* 106(12): 3747–54.

Villamor N, Conde L, Martínez-Trillos A, Cazorla M, Navarro A, Beà S, López C, Colomer D, Pinyol M, Aymerich M, Rozman M, Abrisqueta P, Baumann T, Delgado J, Giné E, González-Díaz M, Hernández JM, Colado E, Payer AR, Rayon C, Navarro B, José Terol M, Bosch F, Quesada V, Puente XS, López-Otín C, Jares P, Pereira A, Campo E and López-Guillermo A (2013) NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia* 27(5): 1100–6.

Vinagre J, Almeida A, Pópulo H, Batista R, Lyra J, Pinto V, Coelho R, Celestino R, Prazeres H, Lima L, Melo M, da Rocha AG, Preto A, Castro P, Castro L, Pardal F, Lopes JM, Santos LL, Reis RM, Cameselle-Teijeiro J, Sobrinho-Simões M, Lima J, Máximo V and Soares P (2013) Frequency of TERT promoter mutations in human cancers. *Nature communications* 4: 2185.

Virappane P, Gale R, Hills R, Kakkas I, Summers K, Stevens J, Allen C, Green C, Quentmeier H, Drexler H, Burnett A, Linch D, Bonnet D, Lister TA and Fitzgibbon J (2008) Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26(33): 5429–35.

Visser O, Trama A, Maynadié M, Stiller C, Marcos-Gragera R, De Angelis R, Mallone S, Tereanu C, Allemani C, Ricardi U and Schouten HCC (2012) Incidence, survival and prevalence of myeloid malignancies in Europe. *European journal of cancer (Oxford, England : 1990)* 48(17): 3257–66.

Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, Proszek PZ, Melchor L, Pawlyn C, Kaiser MF, Johnson DC, Qiang Y-W, Jones JR, Cairns DA, Gregory WM, Owen RG, Cook G, Drayson MT, Jackson GH, Davies FE and Morgan GJ (2015) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nature*

communications. Nature Publishing Group 6: 6997.

Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Kandoth C, Baty J, Westervelt P, DiPersio JF, Mardis ER, Wilson RK, Ley TJ and Graubert TA (2011) Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* 25(7): 1153–8.

Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, DiPersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK and Graubert TA (2012) Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine* 366(12): 1090–8.

Wang L, Fidler C, Nadig N, Giagounidis A, Della Porta MG, Malcovati L, Killick S, Gattermann N, Aul C, Boultwood J and Wainscoat JS (2008) Genome-wide analysis of copy number changes and loss of heterozygosity in myelodysplastic syndrome with del(5q) using high-density single nucleotide polymorphism arrays. *Haematologica* 93(7): 994–1000.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G and Wu CJ (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. (2011/12/14.). *The New England journal of medicine* 365(26): 2497–2506.

Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, Michor F, Meric-Bernstam F and Navin NE (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512(7513): 155–60.

Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, Wang P, Yang H, Ma S, Lin H, Jiao B, Ren R, Ye D, Guan K-L and Xiong Y (2015) WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Molecular cell* 57(4): 662–73.

Warren M, Luthra R, Yin CC, Ravandi F, Cortes JE, Kantarjian HM, Medeiros LJ and Zuo Z (2012) Clinical impact of change of FLT3 mutation status in acute myeloid leukemia patients. *Modern Pathology*. Nature Publishing Group 25(10): 1405–1412.

Watson L, Wyld P and Catovsky D (2008) Disease burden of chronic lymphocytic leukaemia within the European Union. *European journal of haematology* 81(4): 253–8.

Weinhold N, Jacobsen A, Schultz N, Sander C and Lee W (2014) Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature genetics* 46(11): 1160–5.

van de Werken HJG, Landan G, Holwerda SJB, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Öz Y, Bouwman BAM, Verstegen MJAM, de Wit E, Tanay A, de Laat W, Oz Y, Bouwman BAM, Verstegen MJAM, de Wit E, Tanay A and de Laat W (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. Nature Research 9(10): 969–972.

Westbrook CJ, Karl JA, Wiseman RW, Mate S, Koroleva G, Garcia K, Sanchez-Lockhart M, O'Connor DH and Palacios G (2015) No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human Immunology* 76(12): 891–896.

Whitman SP, Archer KJ, Feng L, Baldus C, Becknell B, Carlson BD, Carroll AJ, Mrózek K, Vardiman JW, George SL, Kolitz JE, Larson RA, Bloomfield CD and Caligiuri MA (2001) Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: a cancer and leukemia group B study. *Cancer research* 61(19): 7233–9.

Wiestner A, Rosenwald A, Barry TS, Wright G, Davis RE, Henrickson SE, Zhao H, Ibbotson RE, Orchard JA, Davis Z, Stetler-Stevenson M, Raffeld M, Arthur DC, Marti GE, Wilson WH, Hamblin TJ, Oscier DG and Staudt LM (2003) ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. *Blood*. American Society of Hematology 101(12): 4944–51.

Wouters BJ, Löwenberg B, Erpelinck-Verschueren CAJ, van Putten WLJ, Valk PJM and Delwel R (2009) Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* 113(13): 3088–91.

Wu S, Huang P, Li C, Huang Y, Li X, Wang Y, Chen C, Lv Z, Tang A, Sun X, Lu J, Li W, Zhou J, Gui Y, Zhou F, Wang D and Cai Z (2014) Telomerase reverse transcriptase gene promoter mutations help discern the origin of urogenital tumors: a genomic and molecular study. *European urology* 65(2): 274–7.

Yamamoto Y, Kiyoi H, Nakano Y, Suzuki R, Kodera Y, Miyawaki S, Asou N, Kuriyama K, Yagasaki F, Shimazaki C, Akiyama H, Saito K, Nishimura M, Motoji T, Shinagawa K, Takeshita A, Saito H, Ueda R, Ohno R and Naoe T (2001) Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* 97(8): 2434–9.

Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, Ueno T, Soda M, Hamada T, Haruta H, Takada S, Miyazaki Y, Kiyoi H, Ito E, Naoe T, Tomonaga M, Toyota M, Tajima S, Iwama A and Mano H (2010) Array-based genomic resequencing of human leukemia. *Oncogene* 29(25): 3723–31.

Yan X-J, Xu J, Gu Z-H, Pan C-M, Lu G, Shen Y, Shi J-Y, Zhu Y-M, Tang L, Zhang X-W, Liang W-X, Mi J-Q, Song H-D, Li K-Q, Chen Z and Chen S-J (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature genetics* 43(4): 309–15.

Ye K, Schulz MH, Long Q, Apweiler R and Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxford, England)* 25(21): 2865–71.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann W-KK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih L-YY, Haferlach T, Chiba S, Nakauchi H, Miyano S and Ogawa S (2011) Frequent pathway mutations of splicing machinery in myelodysplasia (2011/09/13.). *Nature* 478(7367): 64–69.

Zenz T, Eichhorst B, Busch R, Denzel T, Häbe S, Winkler D, Bühler A, Edelmann J, Bergmann M, Hopfinger G, Hensel M, Hallek M, Döhner H and Stilgenbauer S (2010) TP53 mutation and survival in chronic lymphocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28(29): 4473–9.

Zenz T, Kröber A, Scherer K, Häbe S, Bühler A, Benner A, Denzel T, Winkler D, Edelmann J, Schwänen C, Döhner H, Stilgenbauer S, Kro A, Scherer K, Ha S, Bu A, Benner A, Denzel T, Winkler D, Edelmann J, Schwa C, Kröber A, Scherer K, Häbe S, Bühler A, Benner A, Denzel T, Winkler D, Edelmann J, Schwänen C, Döhner H and Stilgenbauer S (2008) Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. *Blood* 112(8): 3322–9.

Zenz T, Mertens D, Küppers R, Döhner H and Stilgenbauer S (2010) From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature reviews. Cancer* 10(1): 37–50.

Zenz T, Mohr J, Edelmann J, Sarno A, Hoth P, Heuberger M, Helfrich H, Mertens D, Döhner H and Stilgenbauer S (2009) Treatment resistance in chronic lymphocytic leukemia: the role of the p53 pathway. *Leukemia & lymphoma* 50(3): 510–3.

Zenz T, Mohr J, Eldering E, Kater AP, Bühler A, Kienle D, Winkler D, Dürig J, van Oers MHJ, Mertens D, Döhner H and Stilgenbauer S (2009) miR-34a as part of the resistance network in chronic lymphocytic leukemia. *Blood* 113(16): 3801–8.

Zhang S, Fukuda S, Lee Y, Hangoc G, Cooper S, Spolski R, Leonard WJ and Broxmeyer HE (2000) Essential role of signal transducer and activator of transcription (Stat)5a but not Stat5b for Flt3-dependent signaling. *The*

Journal of experimental medicine 192(5): 719–28.

Zucchetto A, Benedetti D, Tripodo C, Bomben R, Dal Bo M, Marconi D, Bossi F, Lorenzon D, Degan M, Rossi FM, Rossi D, Bulian P, Franco V, Del Poeta G, Deaglio S, Gaidano G, Tedesco F, Malavasi F and Gattei V (2009) CD38/CD31, the CCL3 and CCL4 chemokines, and CD49d/vascular cell adhesion molecule-1 are interchained by sequential events sustaining chronic lymphocytic leukemia cell survival. *Cancer research* 69(9): 4001–9.

Zucchetto A, Bomben R, Dal Bo M, Bulian P, Benedetti D, Nanni P, Del Poeta G, Degan M and Gattei V (2006) CD49d in B-cell chronic lymphocytic leukemia: correlated expression with CD38 and prognostic relevance. *Leukemia* 20(3): 523-5-9.

APPENDICES

Appendix A: List of non-synonymous variants with COSMIC ID or not reported in dbSNP

Sample ID	Diagnosis	Gene	Genome coordinates	DNA change	Protein change	Variant call ratio [% (variant/total)]	COSMIC ID	dbSNP ID	Polyphen2
MDS16	RA (5q- syndrome)	*RUNX1*	chr21:36259324	A>AG	L29S	31.9 (23/72)	COSM24756	rs111527738	Probably damaging
MDS15	RA (5q- syndrome)	*SF3B1*	chr2:198266834	T>TC	K700E	11.4 (170/1496)	COSM84677	N/A	Probably damaging
MDS07	RA (5q- syndrome)	*DNMT3A*	chr2:25457242	C>CA	R882L	7.8 (92/1176)	N/A	N/A	Probably damaging
MDS08	RA (5q- syndrome)	*ASXL1*	chr20:31022449	insG	G646WfsX12	44.7 (174/389)	COSM34210	N/A	Truncated protein
MDS08	RA (5q- syndrome)	*WT1*	chr11:32413565	C>CT	R462Q	49.0 (174/355)	COSM21408	N/A	Probably damaging
MDS14	RA (5q- syndrome)	*TET2*	chr4:106193748	C>CT	R1404X	45.1 (309/685)	COSM42037	N/A	Truncated protein
MDS12	RA (5q- syndrome)	*ASXL1*	chr20:31022449	insG	G646WfsX12	10.5 (37/351)	COSM34210	N/A	Truncated protein
MDS12	RA (5q- syndrome)	*SF3B1*	chr2:198266834	T>TC	K700E	40.0 (620/1549)	COSM84677	N/A	Probably damaging
MDS12	RA (5q- syndrome)	*TET2*	chr4:106164896	insA	fs (Y1255X)	5.3 (41/771)	COSM110747	N/A	Truncated protein
MDS06	RA (5q- syndrome)	*TET2*	chr4:106197552	C>CT	P1962L	50.3 (303/602)	COSM41894	N/A	Probably damaging
MDS11	RA (5q- syndrome)	*ASXL1*	chr20:31022902	G>GA	W796X	35.8 (144/402)	COSM53207	N/A	Truncated protein
MDS10	RA (5q- syndrome)	*TP53*	chr17:7578413	C>CG	V173L	41.1 (109/265)	COSM43559	N/A	Probably damaging
MDS29	RA (5q- syndrome)	*JAK2*	Chr9:5073770	G>GT	V617F	7	COSM12600	rs77375493	Probably damaging

MDS34	RA (5q- syndrome)	*JAK2*	Chr9:5073770	G>GT	V617F	28	COSM12600	rs77375493	Probably damaging
MDS29	Del(5q) RA with additional cytogenetic abnormalities	*DNMT3A*	chr2:25457176	G>GA	P904L	44.0 (198/450)	COSM52989	rs149095705	Probably damaging
MDS28	Del(5q) RA with additional cytogenetic abnormalities	*U2AF1*	chr21:44514777	T>TC	Q157R	38.3 (242/632)	COSM144989	N/A	Probably damaging
MDS30	Del(5q) RA with additional cytogenetic abnormalities	*CBL*	chr11:119149332	C>CT	A447V	43.6 (99/227)	N/A	N/A	Possibly damaging
MDS26	Del(5q) RA with additional cytogenetic abnormalities	*TP53*	chr17:7577553	A>AG	M243T	28.0 (327/1166)	COSM43726	N/A	Probably damaging
MDS37	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577120	C>CT	R273H	82.2 (620/754)	COSM10660	rs28934576	Possibly damaging
MDS42	Advanced del(5q) MDS (CMML)	*ASXL1*	chr20:31023821	G>GT	E1102D	45.4 (366/806)	COSM36205	rs139115934	Possibly damaging
MDS42	Advanced del(5q) MDS (CMML)	*CBL*	chr11:119149004	G>GT	W408C	96.0 (267/278)	COSM34072	N/A	Probably damaging
MDS36	Advanced del(5q) MDS (RAEB)	*ASXL1*	chr20:31024704	G>GA	G1397S	49.9 (875/1755)	COSM133033	rs146464648	Possibly damaging
MDS33	Advanced del(5q) MDS (RAEB)	*TET2*	chr4:106196850	insCATG	E1728Dfs*13	17.0 (121/713)	COSM211745	N/A	Truncated protein
MDS43	Advanced del(5q) MDS (RAEB)	*TET2*	chr4:106164880	G>GT	E1250X	27.3 (313/1145)	N/A	N/A	Truncated protein
MDS43	Advanced del(5q) MDS (RAEB)	*ASXL1*	chr20:31022449	insG	G646WfsX12	42.8 (470/1097)	COSM34210	N/A	Truncated protein
MDS39	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7578190	T>TC	Y220C	38.7 (48/124)	COSM99719	rs121912666	Probably damaging
MDS39	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7578275	G>GA	Q192X	49.3 (99/201)	COSM117949	N/A	Truncated protein
MDS38	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577538	C>CA	R248L	44.1 (1168/2648)	COSM6549	rs11540652	Probably damaging
MDS38	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577568	C>CT	C238Y	37.5 (998/2664)	COSM11059	N/A	Probably damaging

Appendix B: List of clinically relevant mutations in 270 diagnostic samples

Sample	Finding	Gene	Nucleotide Change	Amino Acid Change
AML1	Good Risk	*NPM1*	c.859_860insTCTG	p.Trp288CysfsTer12
AML2	Targetable mutation	*IDH1*	c.394C>T	p.Arg132Cys
AML3	Poor risk	*KIT*	c.1248_1256GACTTACGA>TCCACC	p.Thr417_Asp419delins
AML7	Targetable mutation	*IDH1*	c.394C>A	p.Arg132Ser
AML8	Poor risk	*KIT*	c.2466T>G	p.Asn822Lys
AML9	MRD monitoring	*NPM1*	c.863_864insTCGG,	p.Trp288fs*12
AML11	Targetable mutation	*IDH1*	c.394C>T	p.Arg132Cys
AML13	Targetable mutation	*IDH2*	c.419G>A	p.Arg140Gln
AML19	Poor risk	*CEBPA*	c.110_111delCG	p.Arg37Glyfs*70
AML20	Poor risk	*KIT*	c.2466T>G	p.Asn822Lys
AML21	Targetable mutation	*IDH2*	c.515G>A	p.Arg172Lys
AML23	Targetable mutation	*IDH1*	c.394C>T	p.Arg132Cys
AML25	Poor risk	*TET2*	c.3781C>T	p.Arg1261Cys
AML27	Poor risk	*TET2*	c.412C>T	p.Gln138*
AML30	Targetable mutation	*IDH2*	c.419G>A	p.Arg140Gln
AML31	Poor risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
AML37	Targetable mutation	*IDH1*	c.395G>A	p.Arg132His
AML38	Poor risk	*KIT*	c.2447A>T	p.Asp816Val
AML40	Targetable mutation	*IDH2*	c.515G>A	p.Arg172Lys
AML41	Targetable mutation	*IDH1*	c.394C>T	p.Arg132Cys
AML43	Poor risk	*KIT*	c.2466T>G	p.Asn822Lys
AML44	Poor risk		*FLT3*-ITD	
AML45	Good Risk	*NPM1*	c.863_864insTCGG,	p.Trp288fs*12
AML46	Poor risk	*TET2*	c.5152G>T	p.Val1718Leu
AML47	Good Risk	*CEBPA*	c.912_913insTTG	p.Lys304_Q305insLeu
AML48	MRD monitoring	*NPM1*	c.859_860insTCTG	p.Trp288Cysfs*12
AML50	Poor risk	*TET2*	c.2856_2872del	p.Arg953Glufs*13
AML52	Targetable mutation	*IDH2*	c.515G>A	p.Arg172Lys
AML58	Targetable mutation	*IDH2*	c.515G>A	p.Arg172Lys
AML60	Good Risk	*NPM1*	c.863_864insTCGG,	p.Trp288fs*12
AML62	Poor risk	*CEBPA*	c.1042_1043dupAG	p.Ser348Argfs*75
AML64	Good Risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
AML65	Good Risk	*IDH1*	c.394C>A	p.Arg132Ser
MDS4	Targetable mutation	*IDH1*	c.395G>T	p.Ala132Leu
MDS5	Confirmed diagnosis	*WT1*	c.1108dupC	p.Arg370Profs*15
MDS6	Confirmed diagnosis	Absence of mutations added weight to patient not having a myeloid disorder		
MDS11	Targetable mutation	*SF3B1*	c.2098A>G	p.Lys700Glu
MDS15	Poor risk	*TP53*	c.808T>A	p.Phe270Ile
MDS16	Poor risk	*RUNX1*	c.335T>A	p.Leu112Gln
MDS17	Confirmed diagnosis	*TET2*	c.4354C>T	p.Arg1452*

MDS18	Targetable mutation	*IDH2*	c.419G>A	p.Arg140Gln
MDS20	Poor risk	*RUNX1*	c.601C>T	p.Arg201*
MDS21	Confirmed diagnosis	*U2AF1*	c.101C>T	p.Ser34Phe
MDS27	Targetable mutation	*SF3B1*	c.2098A>G	p.Lys700Glu
MDS29	Targetable mutation	*SF3B1*	c.2098A>G	p.Lys700Glu
MDS30	Targetable mutation	*SF3B1*	c.1874G>T	p.Arg625Leu
MDS32	Poor risk	*TP53*	c.722C>T	p.Ser241Phe
MDS33	Poor risk	*TP53*	c.526T>C	p.Cys176Arg
MDS36	Poor risk	*TP53*	c.821T>A	p.Val274Asp
MDS37	Poor risk	*TP53*	c.839G>T	p.ArgR280Ile
MDS38	Poor risk	*SF3B1*	c.1984C>G	p.His662Asp
MDS40	Poor risk	*SF3B1*	c.2098A>G	p.Lys700Glu
MDS41	Targetable mutation	*IDH2*	c.419G>A	p.Arg140Gln
MDS45	Targetable mutation	*IDH1*	c.394C>G	p.Arg132Gly
MDS46	Targetable mutation	*IDH1*	c.395G>A	p.Arg132His
MDS47	Poor risk	*EZH2*	c.1747C>T	p.Arg583*
MDS48	Poor risk	*TP53*	c.568C>A	p.Pro190Thr
MDS54	Poor risk	*RUNX1*	c.167T>C	p.Leu56Ser
MDS60	Poor risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
MDS63	Poor risk	*TP53*	c.838A>G	p.Arg280Gly
MDS64	Poor risk	*TP53*	c.839G>C	p.Arg280Thr
MPN14	Confirmed diagnosis	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
MPN17	Confirmed diagnosis	*RUNX1*	c.602G>A	p.Arg201Gln
MPN32	Confirmed diagnosis	*KRAS*	c.35G>A	p.Gly12Asp
MPN33	Poor risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
Other5	Poor risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12
Other15	Targetable mutation	*SF3B1*	c.2098A>G	p.Lys700Glu
Other24	Poor risk	*TP53*	c.341T>A	p.Lys114*
Other30	Poor risk	*ASXL1*	c.2077C>T	p.Arg693*
Other58	Good Risk	*NPM1*	c.863_864insTCGG,	p.Trp288fs*12
Other63	Targetable mutation	*IDH2*	c.515G>A	p.Arg172Lys
Other68	Poor risk	*ASXL1*	c.1926_1927insG	p.Gly646Trpfs*12

Appendix C: List of somatically acquired coding-region mutations in 42 CLL genomes

Sample ID	Gene	Variant	Chr	Coordinate	Mutation Type	Amino Acid Change	Read Depth	Variant Read Depth	VAF
CLL003	*PLEKHG5*	C>C/T	1	6,528,057	Missense	Gly1026Arg	42	16	0.38
CLL158	*CASZ1*	G>G/A	1	10,706,210	Missense	Ser1224Phe	18	11	0.61
CWL80	*CASZ1*	C>C/A	1	10,720,395	Missense	Arg235Leu	32	12	0.38
CLL372	*PRAMEF4*	C>C/A	1	12,942,917	Splice Site	-	30	14	0.47
CLL154	*C1orf234*	A>A/T	1	23,337,455	Missense	Met102Lys	21	11	0.52
CLL063	*RHCE*	C>C/T	1	25,735,223	Missense	Gly96Ser	39	19	0.49
CLL129	*ARID1A*	C>C/T	1	27,100,925	Nonsense	Gln1403Ter	40	12	0.3
CLL374	*BAI2*	C>C/T	1	32,198,683	Missense	Val1172Ile	28	16	0.57
CLL366	*TMEM234*	G>G/C	1	32,682,956	Splice Site	-	33	16	0.48
CLL346	*KIAA0319L*	G>G/T	1	35,900,575	Missense	Leu1024Ile	41	17	0.41
CLL301	*IPO13*	C>C/T	1	44,433,258	Splice Site	-	38	15	0.39
CLL321	*PTCH2*	G>G/A	1	45,293,783	Missense	Thr597Ile	32	15	0.47
CLL160	*ELAVL4*	G>G/A	1	50,663,132	Missense	Val273Ile	37	13	0.35
CLL063	*INADL*	C>C/T	1	62,614,029	Missense	Ser1782Leu	30	13	0.43
CLL348	*RABGGTB*	A>A/T	1	76,255,692	Missense	Tyr122Phe	42	20	0.48
CLL145	*LPAR3*	G>G/A	1	85,331,529	Missense	Pro92Leu	65	25	0.38
CLL006	*ARHGAP29*	G>G/A	1	94,650,953	Missense	Thr622Met	30	7	0.23
CLL374	*PALMD*	G>G/A	1	100,133,243	Missense	Gly58Arg	41	15	0.37
CLL348	*AGL*	T>T/G	1	100,336,052	Missense	Leu254Arg	38	19	0.5
CLL372	*TRMT13*	C>C/G	1	100,606,068	Missense	Gln167Glu	45	15	0.33
CLL307	*VCAM1*	T>T/A	1	101,194,737	Missense	Cys335Ser	38	20	0.53
CLL158	*OVGP1*	C>C/G	1	111,962,252	Missense	Ala334Pro	32	16	0.5
CLL369	*VANGL1*	T>T/C	1	116,206,451	Missense	Leu125Pro	50	32	0.64
CLL145	*GJA5*	G>G/A	1	147,231,087	Missense	Thr87Met	52	21	0.4
CLL182	*PI4KB*	G>G/C	1	151,288,917	Missense	Thr26Ser	27	16	0.59
CLL364	*FLG2*	T>T/G	1	152,329,415	Missense	Ser283Arg	34	14	0.41
CLL371	*MSTO1*	G>G/A	1	155,582,252	Missense	Gly286Asp	38	10	0.26
JB210308	*MAEL*	C>C/T	1	166,958,656	Missense	Arg23Trp	26	12	0.46
CWL80	*CCDC181*	G>G/A	1	169,364,308	Missense	Arg502Cys	42	24	0.57
CLL330	*TNFSF18*	A>A/G	1	173,010,861	Splice Site	-	38	14	0.37
CLL321	*TNFSF18*	T>T/A	1	173,019,974	Missense	Arg43Ser	39	17	0.44
CLL003	*SLC9C2*	T>T/C	1	173,569,336	Missense	Lys50Glu	39	21	0.54
CLL346	*SLC9C2*	T>T/C	1	173,570,830	Missense	Lys29Arg	31	15	0.48
CLL307	*LAMC2*	G>G/A	1	183,204,825	Missense	Gly806Ser	36	18	0.5
CLL347	*SWT1*	A>A/G	1	185,175,878	Missense	Tyr652Cys	36	20	0.56
CLL342	*ASPM*	G>G/A	1	197,053,473	Missense	Thr3472Met	41	19	0.46
CLL351	*ZBTB41*	C>C/T	1	197,150,209	Missense	Asp529Asn	39	17	0.44
JB210308	*ZBTB41*	T>T/C	1	197,160,918	Missense	His411Arg	30	12	0.4
CLL346	*LHX9*	C>C/T	1	197,890,624	Missense	Leu190Phe	31	10	0.32

CWL80	*KDM5B*	T>T/A	1	202,718,203	Missense	His629Leu	42	14	0.33
CLL331	*KLHL12*	A>A/G	1	202,880,288	Missense	Val204Ala	34	15	0.44
CLL332	*TMCC2*	A>A/G	1	205,238,228	Missense	Thr300Ala	34	14	0.41
CLL305	*CAMK1G*	C>C/T	1	209,779,776	Missense	Pro183Ser	19	6	0.32
CLL331	*LAMB3*	A>A/C	1	209,788,737	Missense	Leu1133Arg	34	11	0.32
CLL112	*ESRRG*	C>C/T	1	216,850,804	Missense	Arg34Gln	44	13	0.3
CLL342	*KIF26B*	G>G/A	1	245,851,141	Missense	Ser1619Asn	33	13	0.39
CLL003	*NLRP3*	G>G/T	1	247,592,886	Splice Site	-	49	26	0.53
CLL003	*TRIM58*	C>C/G	1	248,028,028	Missense	Gln180Glu	39	20	0.51
CLL331	*ZNF672*	A>A/G	1	249,142,215	Missense	Thr248Ala	37	18	0.49
CLL112	*SMC6*	T>T/A	2	17,897,517	Missense	Asp454Val	34	16	0.47
CLL321	*SMC6*	T>T/G	2	17,919,587	Missense	Lys82Thr	49	39	0.8
CLL369	*OTOF*	G>G/A	2	26,696,120	Missense	Arg1205Cys	36	15	0.42
CLL372	*OTOF*	G>G/A	2	26,700,111	Missense	Arg818Trp	26	13	0.5
CLL186	*PREB*	C>C/T	2	27,355,110	Missense	Cys305Tyr	25	14	0.56
CLL369	*NDUFAF7*	G>G/A	2	37,475,482	Missense	Ala439Thr	39	13	0.33
CLL364	*PSME4*	G>G/A	2	54,125,078	Nonsense	Arg1179Ter	39	14	0.36
CLL371	*XPO1*	T>T/C	2	61,719,471	Missense	Glu571Gly	42	11	0.26
CLL342	*WDPCP*	C>C/T	2	63,631,276	Missense	Val448Ile	30	14	0.47
CLL130	*VAX2*	C>C/T	2	71,148,389	Missense	Arg137Cys	24	11	0.46
CLL077	*EXOC6B*	G>G/C	2	72,802,673	Missense	Ser265Cys	52	19	0.37
CLL331	*CCT7*	G>G/A	2	73,466,870	Missense	Val36Ile	39	15	0.38
CLL364	*LRRTM4*	C>C/T	2	77,746,963	Missense	Gly11Asp	31	11	0.35
CLL369	*CTNNA2*	A>A/C	2	80,782,932	Missense	His552Pro	30	10	0.33
CLL347	*ZNF514*	T>T/A	2	95,815,443	Missense	Ser263Cys	40	18	0.45
CLL301	*DUSP2*	A>A/C	2	96,809,596	Missense	Leu304Arg	31	16	0.52
CLL149	*KIAA1211L*	A>A/G	2	99,411,057	Missense	Trp943Arg	37	17	0.46
CLL372	*MAP4K4*	TA>TA/T	2	102,504,401	Splice Site	-	33	13	0.39
CLL372	*IL1RL2*	T>T/C	2	102,851,455	Missense	Ser466Pro	46	11	0.24
CLL348	*CNTNAP5*	C>C/A	2	125,284,856	Splice Site	-	45	32	0.71
CLL344	*WDR33*	G>G/A	2	128,477,186	Missense	His805Tyr	34	17	0.5
CLL305	*LCT*	A>A/G	2	136,575,246	Missense	Ser458Pro	31	12	0.39
CWL80	*NEB*	G>G/C	2	152,382,708	Missense	Ile7338Met	39	15	0.38
CLL366	*PKP4*	T>T/G	2	159,526,318	Missense	Cys939Gly	50	11	0.22
CLL171	*FAP*	A>A/G	2	163,059,661	Splice Site	-	63	24	0.38
CLL348	*SCN2A*	T>T/C	2	166,243,509	Missense	Val1602Ala	36	7	0.19
CWL80	*SCN2A*	G>G/C	2	166,245,905	Missense	Lys1863Asn	29	13	0.45
CLL371	*ABCB11*	G>G/A	2	169,828,351	Splice Site	-	44	16	0.36
CLL144	*KIAA1715*	T>T/G	2	176,794,920	Missense	Leu354Phe	36	17	0.47
CLL330	*TTN*	A>A/T	2	179,426,199	Missense	Asp28220Glu	31	18	0.58
CLL347	*PMS1*	G>G/T	2	190,719,352	Nonsense	Glu452Ter	46	28	0.61
CLL154	*STAT4*	A>A/C	2	191,929,687	Splice Site	-	32	12	0.38
CLL252	*DNAH7*	G>G/A	2	196,740,516	Nonsense	Gln2057Ter	30	14	0.47
CLL160	*SF3B1*	C>C/T	2	198,266,611	Missense	Gly742Asp	45	15	0.33
CLL305	*SF3B1*	C>C/T	2	198,266,713	Missense	Gly740Glu	39	11	0.28
CLL158	*SF3B1*	A>A/T	2	198,266,821	Missense	Ile704Asn	31	13	0.42

CLL171	*SF3B1*	T>T/A	2	198,266,822	Missense	Ile704Phe	40	17	0.43
CLL003	*SF3B1*	T>T/C	2	198,266,834	Missense	Lys700Glu	45	21	0.47
CLL112	*SF3B1*	T>T/C	2	198,266,834	Missense	Lys700Glu	42	19	0.45
CLL145	*SF3B1*	T>T/C	2	198,266,834	Missense	Lys700Glu	49	22	0.45
CLL307	*SF3B1*	T>T/C	2	198,266,834	Missense	Lys700Glu	30	19	0.63
CLL130	*SF3B1*	T>T/C	2	198,267,361	Missense	Lys666Glu	53	22	0.42
CLL369	*SF3B1*	G>G/A	2	198,267,484	Missense	Arg625Cys	42	19	0.45
CLL156	*SF3B1*	T>T/C	2	198,267,489	Missense	Tyr623Cys	42	13	0.31
CLL129	*SF3B1*	C>C/G	2	198,267,491	Missense	Glu622Asp	42	19	0.45
CLL332	*SF3B1*	C>C/A	2	198,267,491	Missense	Glu622Asp	33	10	0.3
CLL347	*AOX1*	G>G/C	2	201,515,919	Missense	Gly1024Arg	26	7	0.27
CLL186	*ABCA12*	G>G/A	2	215,843,102	Missense	Ala1689Val	27	15	0.56
CLL330	*FN1*	G>G/A	2	216,286,817	Nonsense	Arg515Ter	45	17	0.38
CLL330	*TNS1*	G>G/A	2	218,712,872	Missense	Arg665Cys	19	8	0.42
CLL130	*TNS1*	A>A/G	2	218,750,770	Missense	Ile221Thr	38	15	0.39
CLL330	*PAX3*	G>G/A	2	223,161,923	Missense	Pro32Leu	34	15	0.44
CLL371	*IRS1*	G>G/A	2	227,662,157	Missense	Ser433Phe	32	9	0.28
CLL345	*AGAP1*	G>G/A	2	237,032,743	Missense	Gly851Arg	19	13	0.68
CLL186	*HES6*	C>C/T	2	239,147,581	Missense	Ala188Thr	28	15	0.54
CWL80	*ZNF860*	T>T/G	3	32,031,179	Missense	Ile203Ser	42	11	0.26
CLL346	*ARPP21*	GGATAAAAAC AAA>GGATAA AAACAAA/G	3	35,729,320	Indel	Asp118_Lys121deli nsdel	39	16	0.41
CLL171	*MYD88*	C>C/G	3	38,182,032	Missense	Ser219Cys	39	14	0.36
CLL344	*MYD88*	T>T/C	3	38,182,641	Missense	Leu273Pro	22	11	0.5
CLL186	*FAM198A*	G>G/C	3	43,074,354	Missense	Gly200Ala	38	22	0.58
CLL006	*MAP4*	T>T/C	3	47,898,936	Missense	Ile1045Val	42	14	0.33
CLL305	*DNAH1*	T>T/C	3	52,360,195	Missense	Met149Thr	35	15	0.43
CWL80	*PTPRG*	G>G/A	3	62,189,079	Missense	Arg537His	23	9	0.39
CLL301	*CADPS*	C>C/T	3	62,388,836	Missense	Val1268Met	44	18	0.41
CLL348	*CADPS*	C>C/T	3	62,636,542	Missense	Val395Met	48	11	0.23
CLL144	*MAGI1*	T>T/G	3	65,422,862	Missense	Asn444Thr	34	15	0.44
CLL301	*PDZRN3*	G>G/A	3	73,432,532	Missense	Ser1062Leu	38	20	0.53
CLL364	*PDZRN3*	C>C/T	3	73,673,717	Missense	Arg87His	39	14	0.36
CLL149	*ZNF717*	T>T/G	3	75,790,493	Missense	Ile71Leu	140	14	0.1
CLL350	*ROBO2*	G>G/A	3	77,572,051	Missense	Arg327Gln	44	8	0.18
CLL305	*ROBO1*	G>G/A	3	78,663,819	Missense	Arg1472Cys	41	11	0.27
CLL371	*LINC00971*	A>A/T	3	84,744,510	Splice Site	-	34	19	0.56
CLL145	*CADM2*	G>G/T	3	85,932,593	Missense	Gly124Cys	71	28	0.39
CLL331	*KIAA1407*	G>G/A	3	113,737,706	Missense	Arg328Trp	35	17	0.49
CLL331	*GTF2E1*	C>C/G	3	120,500,126	Missense	Arg377Gly	33	18	0.55
CLL186	*GOLGB1*	TG>TG/T	3	121,414,228	Indel	Thr1714LysfsTer16	32	10	0.31
CLL154	*IQCB1*	G>G/A	3	121,491,506	Nonsense	Arg489Ter	38	18	0.47
CLL372	*ALDH1L1*	C>C/T	3	125,855,672	Missense	Gly437Arg	47	24	0.51
CLL369	*KBTBD12*	G>G/A	3	127,642,243	Missense	Met113Ile	46	16	0.35
CLL331	*PLXND1*	G>G/C	3	129,285,429	Missense	Leu1378Val	31	13	0.42

CLL369	*COL6A6*	C>C/T	3	130,282,383	Missense	Thr179Met	51	21	0.41
CLL301	*RYK*	A>A/G	3	133,910,706	Missense	Cys337Arg	31	12	0.39
CLL366	*AMOTL2*	G>G/A	3	134,078,211	Missense	Ala673Val	31	17	0.55
CLL301	*SOX14*	G>G/A	3	137,484,309	Missense	Gly228Asp	26	11	0.42
CLL130	*PLS1*	T>T/C	3	142,402,993	Missense	Ile242Thr	39	16	0.41
CLL307	*PLSCR4*	T>T/C	3	145,917,655	Missense	Gln190Arg	37	18	0.49
CLL186	*TSC22D2*	G>G/A	3	150,128,527	Missense	Gly464Ser	34	12	0.35
CLL321	*SLC33A1*	T>T/C	3	155,547,682	Missense	Tyr426Cys	39	22	0.56
CLL154	*SLITRK3*	T>T/G	3	164,908,641	Splice Site	-	42	18	0.43
CLL154	*SLITRK3*	G>G/T	3	164,908,642	Splice Site	-	41	18	0.44
CLL144	*KLHL6*	C>C/T	3	183,211,965	Missense	Val418Met	28	12	0.43
CLL348	*KLHL6*	A>A/T	3	183,273,179	Missense	Val88Glu	31	15	0.48
CLL348	*KLHL6*	A>A/G	3	183,273,242	Missense	Met67Thr	33	12	0.36
CLL348	*KLHL6*	G>G/C	3	183,273,246	Missense	Arg66Gly	33	12	0.36
CLL301	*DGKG*	G>G/T	3	185,970,894	Missense	Arg530Ser	33	11	0.33
CLL006	*C3orf43*	C>C/A	3	196,242,031	Missense	Arg17Ile	28	12	0.43
CLL301	*KIAA0226*	A>A/G	3	197,401,948	Missense	Tyr954His	31	12	0.39
CLL145	*TMEM129*	C>C/T	4	1,722,396	Missense	Ala57Thr	47	24	0.51
CLL158	*TMEM129*	A>A/G	4	1,722,458	Missense	Leu36Pro	39	19	0.49
CLL144	*WFS1*	C>C/A	4	6,279,318	Missense	Pro46Thr	37	13	0.35
CLL364	*WFS1*	G>G/A	4	6,302,843	Missense	Val441Met	40	15	0.38
CLL348	*PROM1*	C>C/T	4	15,989,305	Missense	Arg704His	39	15	0.38
CLL129	*KDR*	G>G/A	4	55,991,411	Missense	Thr17Ile	34	10	0.29
CWL80	*CEP135*	C>C/T	4	56,818,360	Missense	Arg22Cys	39	8	0.21
CLL171	*AASDH*	G>G/A	4	57,204,917	Missense	Pro983Leu	72	33	0.46
CWL80	*UGT2B11*	G>G/C	4	70,080,232	Missense	Ser70Cys	31	8	0.26
CLL003	*AMTN*	C>C/T	4	71,396,909	Missense	Arg171Cys	34	10	0.29
CLL347	*ADAMTS3*	A>A/T	4	73,184,400	Missense	Asp458Glu	35	11	0.31
CLL346	*AFM*	G>G/GT	4	74,347,531	Indel	Leu16PhefsTer3	37	20	0.54
CLL307	*RASSF6*	A>A/T	4	74,442,412	Missense	Leu285Gln	29	13	0.45
CLL344	*RASSF6*	T>T/G	4	74,459,160	Splice Site	-	31	9	0.29
CLL144	*CDS1*	G>G/A	4	85,504,636	Missense	Gly28Ser	38	18	0.47
CLL374	*ABCG2*	C>C/T	4	89,015,811	Missense	Ala580Thr	37	17	0.46
CLL321	*MMRN1*	C>C/G	4	90,874,276	Missense	Pro1132Ala	29	13	0.45
CLL305	*MTTP*	A>A/C	4	100,512,476	Missense	Ile196Leu	35	11	0.31
CLL369	*ANK2*	T>T/C	4	114,278,689	Missense	Val2972Ala	53	29	0.55
CLL372	*QRFPR*	T>T/G	4	122,250,611	Missense	Glu385Ala	40	19	0.48
CLL003	*ADAD1*	T>T/A	4	123,304,983	Missense	Cys131Ser	37	19	0.51
CLL351	*FAT4*	C>C/A	4	126,336,381	Missense	Thr2088Asn	40	15	0.38
CLL154	*FAT4*	G>G/T	4	126,371,628	Missense	Asp3153Tyr	40	14	0.35
CLL348	*INTU*	T>T/G	4	128,564,730	Nonsense	Tyr67Ter	45	24	0.53
CLL372	*FREM3*	C>C/T	4	144,620,949	Missense	Val294Met	44	18	0.41
CLL171	*OTUD4*	T>T/C	4	146,059,289	Missense	Ile815Val	58	25	0.43
CLL330	*FBXW7*	T>T/A	4	153,247,310	Missense	Met498Leu	33	15	0.45
CLL305	*FGA*	C>C/T	4	155,509,966	Missense	Gly115Ser	41	11	0.27
CLL374	*SPOCK3*	C>C/A	4	167,833,859	Missense	Arg132Met	40	22	0.55

CLL156	*VEGFC*	G>G/T	4	177,632,668	Missense	Pro230Gln	50	10	0.2
CLL321	*FAT1*	C>C/T	4	187,534,362	Missense	Ala3122Thr	37	17	0.46
CLL346	*FAT1*	C>C/G	4	187,628,094	Missense	Arg963Thr	38	16	0.42
CLL364	*FAT1*	C>C/T	4	187,628,602	Missense	Val794Ile	32	11	0.34
CLL342	*FAT1*	T>T/A	4	187,630,609	Nonsense	Lys125Ter	40	22	0.55
CLL144	*PLEKHG4B*	C>C/T	5	171,380	Missense	Arg935Trp	26	14	0.54
CLL331	*KIAA0947*	T>T/C	5	5,447,828	Splice Site	-	64	24	0.38
CLL252	*40603*	T>T/G	5	16,067,822	Missense	Asn323His	32	8	0.25
CLL171	*CDH12*	C>C/T	5	21,842,354	Missense	Gly244Arg	37	18	0.49
CLL252	*CDH12*	C>C/T	5	22,078,749	Missense	Val13Ile	47	14	0.3
CLL145	*PRDM9*	C>C/T	5	23,522,473	Missense	Ala190Val	67	30	0.45
CLL347	*CDH10*	T>T/A	5	24,537,560	Missense	Asp152Val	38	14	0.37
CLL171	*C5orf22*	G>G/T	5	31,552,890	Missense	Asp404Tyr	61	20	0.33
CLL145	*PDZD2*	C>C/T	5	32,052,757	Missense	Thr569Met	42	20	0.48
CLL374	*PDZD2*	G>G/C	5	32,090,702	Missense	Gly2383Ala	39	17	0.44
CLL144	*SPEF2*	A>A/C	5	35,793,438	Missense	Met1578Leu	42	13	0.31
CLL348	*IL7R*	A>A/G	5	35,876,487	Missense	Thr427Ala	32	8	0.25
CLL374	*UGT3A1*	G>G/A	5	35,954,333	Missense	Arg515Cys	39	15	0.38
CLL345	*DAB2*	G>G/A	5	39,390,614	Missense	Arg132Trp	40	20	0.5
CLL369	*PAPD4*	C>C/G	5	78,919,184	Missense	Arg113Gly	40	16	0.4
CLL366	*ACOT12*	C>C/A	5	80,643,687	Nonsense	Glu187Ter	35	8	0.23
CLL372	*ARRDC3*	GTC>GTC/G	5	90,670,824	Indel	Glu261AspfsTer20	48	19	0.4
CLL330	*ARRDC3*	T>T/C	5	90,678,723	Missense	Arg63Gly	38	20	0.53
CLL129	*KCNN2*	C>C/A	5	113,698,547	Missense	His25Gln	37	23	0.62
CLL321	*DMXL1*	C>C/T	5	118,533,442	Splice Site	-	26	9	0.35
CLL301	*SLC12A2*	G>G/T	5	127,516,625	Missense	Asp1051Tyr	38	16	0.42
CLL364	*FBN2*	G>G/A	5	127,670,870	Missense	Thr1322Ile	40	10	0.25
CLL372	*FBN2*	G>G/A	5	127,872,095	Missense	Pro113Ser	38	14	0.37
CLL307	*RAD50*	A>A/G	5	131,977,898	Missense	Asn1261Asp	43	27	0.63
CLL003	*SHROOM1*	A>A/G	5	132,160,714	Missense	Ser344Pro	48	23	0.48
CLL371	*WNT8A*	G>G/A	5	137,426,659	Nonsense	Trp318Ter	31	14	0.45
CLL160	*PCDHB3*	CT>CT/C	5	140,482,214	Indel	Leu661ProfsTer79	23	11	0.48
CLL369	*PCDHB8*	C>C/T	5	140,558,459	Missense	Leu282Phe	72	20	0.28
CLL149	*RNF14*	C>C/T	5	141,353,281	Missense	Pro43Leu	31	14	0.45
CLL347	*SH3RF2*	G>G/A	5	145,428,816	Splice Site	-	33	8	0.24
CLL144	*PPP2R2B*	A>A/C	5	145,969,658	Missense	Val398Gly	36	17	0.47
CLL374	*FBXO38*	CT>CT/C	5	147,812,978	Splice Site	-	43	16	0.37
CLL348	*ABLIM3*	C>C/A	5	148,630,028	Missense	Ser583Tyr	34	13	0.38
JB210308	*KIF4B*	T>T/C	5	154,394,237	Missense	Val273Ala	39	13	0.33
CLL348	*GABRA1*	A>A/G	5	161,324,181	Missense	Tyr375Cys	36	12	0.33
CLL348	*FGF18*	C>C/A	5	170,883,805	Missense	Ala207Asp	27	15	0.56
CLL344	*SH3PXD2B*	T>T/A	5	171,766,452	Missense	Thr553Ser	31	14	0.45
CLL186	*SH3PXD2B*	G>G/A	5	171,774,312	Missense	Pro346Leu	34	21	0.62
CLL342	*MGAT1*	G>G/A	5	180,219,755	Nonsense	Gln73Ter	31	8	0.26
CLL144	*PXDC1*	C>C/T	6	3,738,356	Missense	Asp95Asn	35	16	0.46

CLL372	*DSP*	C>C/T	6	7,581,486	Missense	Ala1688Val	34	20	0.59
CLL301	*CD83*	C>C/T	6	14,118,264	Missense	Pro41Ser	48	20	0.42
CLL149	*HIST1H2AG*	C>C/T	6	27,101,100	Missense	Leu84Phe	26	10	0.38
CLL301	*HIST1H2BM*	C>C/G	6	27,782,831	Missense	Pro4Ala	40	17	0.43
CLL366	*OR2B6*	G>G/T	6	27,925,466	Missense	Val150Phe	30	13	0.43
CLL144	*ZNF311*	T>T/G	6	28,963,397	Missense	His461Pro	50	24	0.48
CWL80	*PRRC2A*	C>C/T	6	31,603,374	Nonsense	Arg1797Ter	37	19	0.51
CLL182	*ZBTB22*	C>C/T	6	33,283,955	Missense	Val247Ile	33	13	0.39
CLL348	*KIF6*	A>A/G	6	39,607,414	Missense	Leu124Pro	36	19	0.53
CLL351	*DAAM2*	C>C/T	6	39,846,218	Missense	Arg467Trp	38	14	0.37
CLL347	*PTK7*	G>G/T	6	43,112,282	Missense	Arg790Leu	34	12	0.35
CLL305	*NFKBIE*	TGTAA>TGTAA/T	6	44,232,738	Indel	Tyr254SerfsTer13	29	10	0.34
CLL112	*TNFRSF21*	G>G/A	6	47,200,590	Missense	Arg627Trp	31	17	0.55
CLL305	*TFAP2D*	C>C/T	6	50,740,501	Missense	Ala428Val	36	10	0.28
CLL144	*PKHD1*	C>C/A	6	51,512,894	Missense	Gly3778Val	43	19	0.44
CLL144	*PKHD1*	A>A/T	6	51,524,477	Missense	Phe3483Ile	48	18	0.38
CLL112	*HMGCLL1*	G>G/T	6	55,378,954	Missense	Ala175Glu	53	27	0.51
CLL301	*COL21A1*	A>A/T	6	55,988,882	Nonsense	Leu579Ter	31	11	0.35
CLL350	*RIMS1*	C>C/G	6	72,889,572	Missense	Gln256Glu	34	11	0.32
CLL144	*KHDC1*	A>A/C	6	74,019,377	Missense	Tyr21Asp	39	9	0.23
CLL129	*KHDC3L*	C>C/T	6	74,073,459	Missense	Ala177Val	35	17	0.49
CLL364	*EPHA7*	A>A/C	6	93,967,229	Missense	Met708Arg	38	11	0.29
CLL347	*PNISR*	CCT>CCT/C	6	99,848,839	Indel	Arg665GlufsTer10	37	18	0.49
CLL331	*USP45*	T>T/A	6	99,914,598	Missense	Ile353Phe	39	21	0.54
CLL344	*AK9*	T>T/A	6	109,837,076	Missense	Tyr1350Phe	31	10	0.32
CLL346	*TRMT11*	A>A/T	6	126,333,974	Missense	Lys328Met	37	16	0.43
CLL331	*THEMIS*	G>G/C	6	128,134,980	Nonsense	Ser269Ter	45	19	0.42
CLL374	*SGK1*	G>G/C	6	134,495,183	Missense	Pro158Arg	33	15	0.45
CLL321	*MAP7*	G>G/A	6	136,667,042	Missense	Leu761Phe	38	19	0.5
CLL372	*PDE10A*	C>C/A	6	165,832,148	Missense	Val325Phe	16	9	0.56
CLL112	*KIF25*	G>G/T	6	168,443,240	Splice Site	-	33	11	0.33
CLL171	*IQCE*	C>C/T	7	2,627,479	Missense	Arg338Trp	48	16	0.33
CLL182	*CYTH3*	A>A/G	7	6,210,177	Missense	Leu271Pro	29	13	0.45
CLL171	*ISPD*	G>G/T	7	16,298,032	Missense	Pro368Thr	69	29	0.42
CWL80	*DNAH11*	T>T/C	7	21,856,121	Missense	Ile3464Thr	40	15	0.38
CLL364	*EVX1*	G>G/A	7	27,284,672	Missense	Gly145Ser	27	9	0.33
CLL372	*BMPER*	G>G/A	7	34,125,473	Missense	Arg505His	49	18	0.37
CLL348	*AOAH*	T>T/C	7	36,634,006	Missense	Thr293Ala	37	22	0.59
CLL369	*DBNL*	A>A/C	7	44,099,154	Missense	Glu324Ala	31	11	0.35
CLL158	*PKD1L1*	C>C/T	7	47,933,601	Missense	Arg776Gln	27	15	0.56
CLL347	*SUN3*	T>T/C	7	48,026,979	Missense	Tyr341Cys	39	14	0.36
JB210308	*DDC*	C>C/T	7	50,571,751	Missense	Ala241Thr	39	10	0.26
CLL144	*COBL*	T>T/G	7	51,096,337	Missense	Lys819Thr	33	15	0.45
CLL006	*EGFR*	TG>TG/T	7	55,227,941	Indel	Cys470SerfsTer6	42	20	0.48
CLL347	*CRCP*	A>A/G	7	65,599,285	Missense	Lys55Glu	28	11	0.39

CLL112	*MAGI2*	G>G/T	7	77,885,444	Missense	Asp621Glu	30	11	0.37
CLL330	*SEMA3C*	G>G/A	7	80,418,712	Missense	Arg422Cys	40	21	0.53
CLL006	*PCLO*	A>A/T	7	82,582,369	Missense	Ser2634Thr	46	13	0.28
CLL063	*SEMA3A*	A>A/G	7	83,606,456	Missense	Leu570Ser	49	20	0.41
CLL186	*ASB4*	G>G/A	7	95,157,380	Missense	Arg248Gln	37	8	0.22
CLL372	*ZSCAN25*	C>C/T	7	99,219,105	Missense	Pro166Leu	44	23	0.52
CLL006	*PILRB*	G>G/T	7	99,956,959	Splice Site	-	42	16	0.38
CLL330	*MUC3A*	C>C/G	7	100,552,712	Nonsense	Ser1098Ter	53	18	0.34
CLL348	*MUC17*	C>C/G	7	100,680,615	Missense	Thr1973Ser	21	7	0.33
CLL369	*PSMC2*	G>G/A	7	103,008,377	Missense	Gly393Asp	32	15	0.47
CWL80	*RELN*	G>G/A	7	103,155,701	Missense	Arg2684Cys	44	14	0.32
CLL158	*PNPLA8*	T>T/TA	7	108,155,909	Indel	Tyr10IlefsTer6	47	20	0.43
CLL145	*POT1*	C>C/A	7	124,503,668	Missense	Gln94His	53	13	0.25
CLL347	*BRAF*	T>T/C	7	140,453,193	Missense	Asn581Ser	38	9	0.24
CLL364	*CLEC5A*	T>T/C	7	141,646,016	Missense	Tyr26Cys	48	13	0.27
CLL321	*MGAM*	C>C/T	7	141,759,267	Splice Site	-	29	14	0.48
CLL252	*MGAM*	G>G/A	7	141,759,672	Missense	Gly1322Asp	44	14	0.32
CLL003	*KMT2C*	ATTGCCAACCTGCACG>ATTGCCAACCTGCACG/A	7	151,962,154	Indel	Arg380_Gln384delinsdel	88	6	0.07
CLL344	*LOC100996700*	G>G/C	7	153,932,892	Missense	Trp9Ser	38	15	0.39
CLL346	*FBXO25*	C>C/G	8	382,894	Missense	Arg83Gly	56	29	0.52
CLL345	*GATA4*	T>T/G	8	11,566,282	Missense	Phe154Cys	20	8	0.4
CLL129	*FAM86B1*	G>G/A	8	12,044,011	Nonsense	Arg164Ter	44	6	0.14
CLL063	*MICU3*	T>T/TA	8	16,942,809	Indel	Glu256ArgfsTer14	51	20	0.39
CLL006	*CNOT7*	G>G/A	8	17,088,342	Missense	His249Tyr	45	15	0.33
CLL003	*MTUS1*	C>C/A	8	17,507,447	Nonsense	Glu1137Ter	47	18	0.38
CLL331	*MTUS1*	T>T/A	8	17,611,963	Nonsense	Lys452Ter	32	16	0.5
CLL344	*CLU*	G>G/C	8	27,466,610	Splice Site	-	39	16	0.41
CLL186	*TEX15*	G>G/T	8	30,704,711	Missense	Thr608Asn	39	19	0.49
CLL145	*NRG1*	C>C/A	8	32,453,520	Missense	Pro92Gln	67	21	0.31
CLL145	*NRG1*	G>G/T	8	32,453,522	Missense	Gly93Trp	66	21	0.32
CLL346	*KCNU1*	T>T/C	8	36,768,622	Missense	Ser836Pro	25	7	0.28
CLL006	*LINC00293*	C>C/T	8	47,757,708	Splice Site	-	38	16	0.42
CLL063	*OPRK1*	G>G/A	8	54,147,555	Missense	Pro125Leu	41	16	0.39
CLL347	*RGS20*	C>C/T	8	54,852,163	Missense	Arg180Trp	30	9	0.3
CLL371	*CHD7*	A>A/T	8	61,765,885	Nonsense	Lys2201Ter	28	15	0.54
CLL182	*TRPA1*	C>C/T	8	72,964,955	Missense	Val564Ile	28	14	0.5
CLL330	*IL7*	C>C/T	8	79,710,366	Missense	Glu30Lys	42	18	0.43
CLL307	*MMP16*	C>C/A	8	89,198,807	Missense	Cys101Phe	21	10	0.48
JB210308	*VPS13B*	C>C/T	8	100,829,849	Nonsense	Gln2752Ter	41	19	0.46
CLL063	*CSMD3*	A>A/G	8	113,303,792	Missense	Val2974Ala	40	21	0.53
CLL144	*COL14A1*	G>G/A	8	121,209,122	Missense	Val177Ile	32	15	0.47
CLL344	*ATAD2*	A>A/G	8	124,358,860	Splice Site	-	39	16	0.41
CLL171	*ASAP1*	T>T/C	8	131,140,307	Missense	Glu416Gly	46	13	0.28
CLL321	*KHDRBS3*	C>C/A	8	136,594,147	Missense	Pro213Gln	26	6	0.23

CLL129	*FAM135B*	G>G/A	8	139,164,244	Missense	Ala825Val	39	23	0.59
CLL144	*DENND3*	ACAGGTAACA GCAT>ACAGGT AACAGCAT/A	8	142,202,432	Indel	Gly1049TrpfsTer3	29	13	0.45
CLL186	*LY6D*	T>T/G	8	143,866,816	Missense	Ser70Arg	22	9	0.41
CLL144	*GLIS3*	G>G/A	9	3,937,028	Splice Site	-	31	16	0.52
CLL330	*PTPRD*	T>T/C	9	8,486,190	Missense	Lys876Arg	37	21	0.57
CLL332	*TYRP1*	C>C/T	9	12,695,565	Missense	Arg146Trp	40	16	0.4
CLL129	*FREM1*	C>C/A	9	14,769,744	Missense	Gly1728Trp	48	14	0.29
CLL372	*ADAMTSL1*	C>C/T	9	18,721,614	Missense	Arg653Trp	37	19	0.51
CLL344	*RPS6*	C>C/T	9	19,376,365	Missense	Glu226Lys	24	8	0.33
CLL346	*MLLT3*	C>C/A	9	20,413,939	Missense	Ser302Ile	47	24	0.51
CLL003	*IL11RA*	C>C/T	9	34,655,300	Missense	Ala29Val	32	18	0.56
CLL301	*FAM214B*	G>G/T	9	35,107,995	Missense	Pro93Thr	40	15	0.38
CLL346	*LOC100996689*	C>C/T	9	66,454,921	Missense	Gly275Arg	84	6	0.07
CLL347	*TMEM2*	G>G/A	9	74,315,561	Missense	Thr1125Met	44	20	0.45
CLL063	*GCNT1*	C>C/T	9	79,118,139	Missense	Pro281Leu	40	16	0.4
CLL371	*PRUNE2*	C>C/T	9	79,244,178	Missense	Val3027Ile	35	15	0.43
CLL077	*TLE4*	C>C/CA	9	82,323,058	Indel	Ser323ValfsTer9	50	11	0.22
CLL154	*WNK2*	C>C/T	9	96,070,874	Splice Site	-	28	13	0.46
CLL342	*FSD1L*	G>G/A	9	108,296,776	Missense	Ser343Asn	30	12	0.4
CLL182	*ZNF462*	C>C/T	9	109,691,847	Missense	Pro1885Leu	32	9	0.28
CLL351	*SVEP1*	C>C/T	9	113,173,761	Missense	Arg2077His	42	19	0.45
CLL330	*LPAR1*	C>C/T	9	113,703,772	Missense	Arg241Gln	40	21	0.53
CLL321	*CDK5RAP2*	G>G/A	9	123,182,182	Missense	Ser1354Leu	39	19	0.49
CLL156	*TTLL11*	G>G/A	9	124,801,635	Missense	Arg249Trp	40	17	0.43
CLL351	*ZBTB6*	T>T/C	9	125,673,189	Missense	Asp388Gly	34	11	0.32
CLL345	*STRBP*	CATTTTCTAAC AGTAAAA>CAT TTTCTAACAGT AAAA/C	9	125,923,342	Splice Site	-	36	11	0.31
CLL374	*TTC16*	C>C/T	9	130,485,392	Splice Site	-	24	7	0.29
CLL003	*SPTAN1*	G>G/C	9	131,340,459	Missense	Ala386Pro	43	13	0.3
CLL112	*POMT1*	C>C/A	9	134,382,812	Missense	Ala113Asp	39	21	0.54
CLL301	*SETX*	T>T/G	9	135,140,240	Missense	Thr2474Pro	35	16	0.46
CLL369	*DDX31*	C>C/T	9	135,536,620	Missense	Val248Ile	39	25	0.64
CLL342	*REXO4*	G>G/A	9	136,279,924	Missense	Pro145Ser	42	17	0.4
CLL130	*NOTCH1*	CAG>CAG/C	9	139,390,648	Indel	Pro2514ArgfsTer4	40	16	0.4
CLL186	*NOTCH1*	CAG>CAG/C	9	139,390,648	Indel	Pro2514ArgfsTer4	36	12	0.33
CLL331	*NOTCH1*	CAG>CAG/C	9	139,390,648	Indel	Pro2514ArgfsTer4	36	13	0.36
JB210308	*NOTCH1*	CAG>CAG/C	9	139,390,648	Indel	Pro2514ArgfsTer4	22	13	0.59
CLL158	*NOTCH1*	G>G/T	9	139,390,678	Missense	Pro2505Thr	30	9	0.3
CLL369	*FBXW5*	C>C/T	9	139,835,384	Missense	Arg566His	32	11	0.34
CLL003	*EXD3*	C>C/T	9	140,245,984	Splice Site	-	54	19	0.35

CLL321	*WDR37*	C>C/T	10	1,132,303	Splice Site	-	26	13	0.5
CLL252	*MSRB2*	G>G/T	10	23,408,297	Missense	Gly121Cys	32	12	0.38
CLL171	*KIAA1217*	C>C/T	10	24,822,021	Missense	Pro1090Leu	37	18	0.49
CLL077	*GPR158*	C>C/T	10	25,888,057	Nonsense	Arg1168Ter	47	18	0.38
CLL346	*PTCHD3*	A>A/AT	10	27,687,469	Indel	Asn686LysfsTer6	41	13	0.32
CLL129	*ANKRD30A*	C>C/T	10	37,508,487	Missense	His1227Tyr	42	18	0.43
CLL371	*C10orf71*	AC>AC/A	10	50,532,583	Indel	His666ThrfsTer3	33	15	0.45
CLL372	*CHAT*	C>C/T	10	50,833,607	Missense	Arg281Trp	36	15	0.42
CLL351	*ANK3*	G>G/T	10	61,898,734	Missense	Ala909Glu	46	26	0.57
CLL186	*EGR2*	G>G/T	10	64,573,248	Missense	His384Asn	34	14	0.41
CLL330	*EGR2*	C>C/T	10	64,573,332	Missense	Glu356Lys	30	8	0.27
CLL130	*TACR2*	C>C/T	10	71,175,920	Missense	Val54Ile	31	17	0.55
CLL301	*UNC5B*	C>C/T	10	73,044,512	Missense	Arg114Trp	32	11	0.34
CLL348	*UNC5B*	G>G/T	10	73,053,326	Missense	Trp646Leu	28	11	0.39
CLL307	*CDH23*	C>C/T	10	73,461,867	Missense	Thr829Ile	32	9	0.28
CLL145	*USP54*	C>C/T	10	75,302,775	Splice Site	-	58	16	0.28
CLL332	*CAMK2G*	G>G/A	10	75,579,350	Missense	Arg403Trp	44	20	0.45
CLL130	*KAT6B*	T>T/C	10	76,790,464	Missense	Met1961Thr	35	17	0.49
CLL130	*KCNMA1*	A>A/T	10	78,649,280	Nonsense	Tyr1130Ter	32	14	0.44
CLL171	*SORBS1*	G>G/C	10	97,174,349	Missense	Pro238Ala	53	18	0.34
CLL351	*MRPL43*	A>A/G	10	102,747,068	Splice Site	-	29	13	0.45
CLL129	*HPS6*	C>C/T	10	103,826,400	Missense	Ala390Val	30	14	0.47
CLL145	*PPRC1*	A>A/T	10	103,892,826	Initiator	Met1	45	14	0.31
CLL112	*PPRC1*	CAT>CAT/C	10	103,901,047	Indel	His928ArgfsTer87	52	20	0.38
CLL145	*SLK*	A>A/G	10	105,727,659	Splice Site	-	34	11	0.32
CLL351	*BBIP1*	CATT>CATT/C	10	112,660,259	Indel	Asn98del	57	27	0.47
CLL351	*ENO4*	T>T/C	10	118,630,628	Splice Site	-	51	30	0.59
CLL003	*SLC18A2*	T>T/C	10	119,026,492	Splice Site	-	54	14	0.26
CLL369	*TACC2*	C>C/T	10	123,844,891	Missense	Ser959Leu	34	11	0.32
CLL350	*TACC2*	G>G/A	10	123,997,536	Missense	Glu2778Lys	32	12	0.38
CLL063	*DMBT1*	A>A/T	10	124,380,792	Missense	Asp1706Val	46	20	0.43
CLL369	*IKZF5*	C>C/T	10	124,754,167	Missense	Arg130His	44	26	0.59
CLL145	*TALDO1*	T>T/A	11	763,742	Splice Site	-	45	20	0.44
CLL332	*SLC25A22*	G>G/C	11	792,960	Missense	Leu108Val	31	11	0.35
CLL369	*MUC2*	C>C/G	11	1,092,941	Missense	Thr1586Ser	111	7	0.06
CLL321	*MUC5B*	A>A/G	11	1,265,854	Missense	Thr2582Ala	18	8	0.44
CLL321	*DNHD1*	A>A/C	11	6,532,648	Missense	Ile461Leu	40	19	0.48
CLL305	*PDE3B*	C>C/A	11	14,854,278	Missense	Thr702Asn	39	12	0.31
CLL369	*PIK3C2A*	T>T/C	11	17,190,732	Missense	Tyr186Cys	36	11	0.31
CLL347	*OTOG*	C>C/A	11	17,590,706	Missense	Gln562Lys	32	16	0.5
CLL129	*IGSF22*	C>C/A	11	18,733,731	Missense	Val766Leu	35	19	0.54
CLL348	*ANO5*	C>C/G	11	22,283,687	Missense	Thr548Arg	37	11	0.3
CLL366	*BBOX1*	A>A/G	11	27,114,885	Missense	Met169Val	32	11	0.34
CLL158	*CAPRIN1*	AAC>AAC/A	11	34,104,528	Indel	Gln304ValfsTer5	40	21	0.53
CLL158	*CAPRIN1*	TAC>TAC/T	11	34,118,133	Indel	Tyr605Ter	33	18	0.55

CLL171	*F2*	G>G/A	11	46,748,114	Missense	Arg314His	33	13	0.39
CLL371	*MYBPC3*	A>A/C	11	47,359,005	Missense	Tyr847Asp	28	13	0.46
CWL80	*OR4C15*	T>T/A	11	55,322,485	Missense	Tyr235Asn	29	13	0.45
CLL331	*OR5M8*	A>A/T	11	56,258,348	Missense	Phe167Ile	35	14	0.4
CLL305	*TNKS1BP1*	C>C/T	11	57,080,906	Missense	Arg419His	39	11	0.28
CLL331	*FAM111B*	C>C/T	11	58,892,585	Missense	Arg339Cys	45	22	0.49
CLL351	*SLC15A3*	A>A/C	11	60,706,987	Missense	Leu467Arg	43	14	0.33
CLL372	*FADS2*	G>G/A	11	61,630,541	Missense	Arg327Lys	35	12	0.34
CLL345	*NXF1*	A>A/C	11	62,568,594	Missense	Leu293Arg	36	10	0.28
CLL077	*PLA2G16*	C>C/T	11	63,342,473	Missense	Ala145Thr	50	18	0.36
CLL366	*PLCB3*	CAGGT>CAGGT/C	11	64,032,697	Splice Site	-	31	9	0.29
CLL345	*KCNK4*	C>C/T	11	64,064,648	Missense	Ala124Val	26	9	0.35
CLL186	*PACS1*	A>A/G	11	66,001,717	Splice Site	-	36	12	0.33
CLL301	*PPP6R3*	TCTATCAAATATTTTCCAC>TCTATCAAATATTTTCCAC/T	11	68,326,055	Indel	Leu252_His257delinsdel	46	22	0.48
CLL369	*TENM4*	G>G/A	11	78,574,151	Missense	His371Tyr	31	19	0.61
CLL332	*PCF11*	CAG>CAG/C	11	82,877,338	Indel	Gln467ArgfsTer10	38	10	0.26
CLL301	*FAT3*	G>G/C	11	92,086,092	Missense	Val272Leu	35	20	0.57
CLL171	*FAT3*	A>A/G	11	92,087,743	Missense	Glu822Gly	57	17	0.3
CLL171	*FAT3*	C>C/T	11	92,534,127	Missense	Leu2650Phe	59	19	0.32
CLL350	*FAT3*	T>T/A	11	92,616,197	Missense	Val4192Glu	30	7	0.23
CLL372	*MAML2*	T>T/C	11	95,825,240	Missense	Gln652Arg	16	14	0.88
CLL369	*BIRC3*	AG>AG/A	11	102,201,871	Indel	Asn409IlefsTer4	14	4	0.29
CLL330	*BIRC3*	T>T/A	11	102,207,782	Nonsense	Cys588Ter	17	14	0.82
CLL144	*ATM*	T>T/G	11	108,121,624	Missense	Ser478Ala	48	17	0.35
CLL144	*ATM*	C>C/T	11	108,121,625	Missense	Ser478Leu	48	17	0.35
CLL003	*ATM*	G>G/A	11	108,153,522	Nonsense	Trp1221Ter	32	26	0.81
CLL063	*ATM*	T>T/A	11	108,199,953	Missense	Ile2432Asn	31	8	0.26
CLL321	*ATM*	C>C/T	11	108,201,099	Missense	Ser2489Phe	15	14	0.93
CLL372	*ATM*	CTTATA>CTTATA/C	11	108,203,577	Indel	Ile2629SerfsTer25	29	20	0.69
CLL369	*ATM*	A>A/C	11	108,236,116	Missense	Lys3018Gln	23	22	0.96
CLL144	*HMBS*	C>C/T	11	118,959,849	Splice Site	-	48	21	0.44
CLL346	*CBL*	G>G/C	11	119,142,445	Missense	Arg148Ser	42	20	0.48
CLL171	*USP2*	G>G/A	11	119,227,604	Missense	Arg587Cys	28	12	0.43
CLL129	*USP2*	C>C/T	11	119,228,894	Missense	Asp436Asn	33	11	0.33
CLL344	*ARHGEF12*	C>C/T	11	120,322,406	Nonsense	Gln677Ter	30	18	0.6
CLL301	*TECTA*	C>C/A	11	121,039,469	Missense	Pro1945His	39	13	0.33
CLL351	*CLMP*	T>T/C	11	122,968,651	Missense	Tyr13Cys	39	21	0.54
CLL182	*OR10G4*	G>G/A	11	123,886,964	Missense	Arg228His	44	18	0.41
CLL158	*ADAMTS8*	C>C/T	11	130,275,559	Missense	Arg855Gln	35	16	0.46
CLL305	*DYRK4*	C>C/T	12	4,722,678	Missense	Ala441Val	53	17	0.32
CLL063	*KCNA5*	C>C/T	12	5,154,337	Missense	Arg342Cys	46	25	0.54
CLL154	*CD163L1*	C>C/T	12	7,528,532	Missense	Arg817His	45	19	0.42

CLL331	*SLC2A14*	A>A/T	12	7,984,356	Missense	Ile62Lys	49	16	0.33
CLL332	*CLEC4A*	A>A/G	12	8,288,207	Missense	Asn109Asp	35	17	0.49
CLL369	*RERGL*	C>C/T	12	18,234,195	Missense	Arg183His	47	21	0.45
CLL351	*CAPZA3*	C>C/T	12	18,891,446	Nonsense	Arg82Ter	35	13	0.37
CLL374	*PDE3A*	C>C/G	12	20,783,046	Missense	Pro582Arg	36	15	0.42
CLL182	*ABCC9*	T>T/G	12	21,998,547	Missense	Lys1029Thr	36	17	0.47
CLL112	*ABCD2*	T>T/TA	12	40,012,873	Indel	His182LeufsTer4	35	12	0.34
CLL374	*LRRK2*	T>T/G	12	40,619,384	Missense	Ile60Ser	34	14	0.41
CLL344	*MUC19*	G>G/A	12	40,875,962	Missense	Ala3054Thr	60	20	0.33
CLL342	*MUC19*	C>C/G	12	40,964,307	Missense	Leu8142Val	41	16	0.39
CWL80	*DIP2B*	A>A/G	12	51,092,990	Splice Site	-	41	24	0.59
CLL186	*TFCP2*	T>T/C	12	51,503,023	Missense	Lys200Glu	44	12	0.27
CLL154	*NR4A1*	T>T/A	12	52,451,019	Missense	Leu459His	30	15	0.5
CLL374	*OR10A7*	C>C/A	12	55,615,572	Missense	Ala255Asp	29	14	0.48
CLL307	*RDH16*	C>C/T	12	57,350,967	Missense	Ala94Thr	28	10	0.36
CLL149	*XPOT*	A>A/C	12	64,827,339	Missense	Gln803Pro	33	16	0.48
CWL80	*TBC1D30*	G>G/A	12	65,225,933	Missense	Gly148Asp	41	21	0.51
CLL330	*WIF1*	T>T/C	12	65,448,930	Missense	Gln329Arg	43	14	0.33
CLL186	*TMBIM4*	T>T/C	12	66,531,789	Missense	Asn223Ser	48	11	0.23
CLL156	*NAV3*	G>G/GT	12	78,443,882	Splice Site	-	52	20	0.38
CLL129	*SYT1*	C>C/T	12	79,679,738	Missense	Thr113Met	30	10	0.33
CLL149	*OTOGL*	T>T/C	12	80,764,475	Splice Site	-	23	13	0.57
CLL346	*DAO*	C>C/A	12	109,278,851	Nonsense	Tyr23Ter	31	18	0.58
CLL301	*BCL7A*	AAATGGTAAG CGGAGGCGCC CGC>AAATGGT AAGCGGAGGC GCCCGC/A	12	122,460,085	Splice Site	-	25	10	0.4
CLL369	*KNTC1*	G>G/C	12	123,055,647	Missense	Ala665Pro	25	10	0.4
CLL374	*SETD8*	A>A/T	12	123,892,045	Missense	Asp285Val	35	8	0.23
CLL369	*SFSWAP*	C>C/T	12	132,284,029	Missense	Ser1003Phe	30	4	0.13
CLL063	*CDK8*	C>C/A	13	26,975,653	Missense	His387Gln	39	39	1
CLL129	*WASF3*	G>G/A	13	27,250,750	Missense	Arg202His	47	22	0.47
CLL321	*MTUS2*	G>G/A	13	29,599,452	Missense	Arg216Gln	30	17	0.57
CLL342	*DCLK1*	C>C/T	13	36,367,609	Missense	Gly651Asp	36	15	0.42
CLL330	*FAM124A*	G>G/T	13	51,855,235	Missense	Arg531Leu	30	14	0.47
CLL182	*SLITRK6*	G>G/T	13	86,369,172	Missense	Thr491Lys	38	19	0.5
CLL371	*GPC5*	G>G/C	13	92,408,651	Missense	Trp419Cys	35	12	0.34
CLL372	*UGGT2*	T>T/C	13	96,642,325	Missense	Glu278Gly	38	17	0.45
CLL371	*UBAC2*	G>G/A	13	100,020,087	Missense	Arg285Gln	37	11	0.3
CLL171	*FGF14*	G>G/A	13	102,527,580	Missense	Pro92Leu	52	22	0.42
JB210308	*MYO16*	G>G/A	13	109,779,874	Missense	Ala1343Thr	28	14	0.5
CLL130	*ARHGEF7*	TG>TG/T	13	111,944,508	Indel	Asp748MetfsTer6	46	20	0.43
CLL003	*RPGRIP1*	G>G/A	14	21,793,478	Missense	Arg768Gln	42	18	0.43
CLL186	*OR4E2*	C>C/A	14	22,133,426	Missense	Leu44Ile	38	21	0.55
CLL366	*NFKBIA*	C>C/T	14	35,871,267	Splice Site	-	38	19	0.5

CLL145	*MDGA2*	G>G/A	14	47,426,751	Missense	Arg639Trp	47	24	0.51
CLL372	*NIN*	G>G/A	14	51,237,255	Nonsense	Arg429Ter	36	22	0.61
CLL171	*KIAA0586*	T>T/C	14	58,975,267	Splice Site	-	53	22	0.42
JB210308	*SNAPC1*	G>G/A	14	62,235,426	Missense	Val168Met	39	15	0.38
CLL374	*WDR89*	A>A/AT	14	64,065,524	Indel	Asp379GlufsTer4	38	21	0.55
CLL112	*DCAF4*	G>G/T	14	73,408,456	Missense	Ala119Ser	35	14	0.4
CLL372	*HEATR4*	C>C/T	14	73,973,175	Missense	Cys661Tyr	36	15	0.42
CLL331	*IRF2BPL*	A>A/G	14	77,492,878	Missense	Phe420Leu	31	15	0.48
CLL154	*CYP46A1*	G>G/A	14	100,166,408	Missense	Arg138Gln	33	14	0.42
CLL182	*TECPR2*	C>C/A	14	102,901,154	Missense	Thr667Asn	32	9	0.28
CLL171	*NC_000014*	G>G/C	14	106,691,909	Missense	Leu39Val	32	8	0.25
CLL171	*NC_000014*	T>T/C	14	107,259,467	Missense	Tyr71Cys	45	13	0.29
CLL171	*NC_000014*	G>G/A	14	107,259,533	Missense	Thr49Ile	43	11	0.26
CLL372	*GOLGA6L2*	T>T/A	15	23,689,147	Missense	Asp123Val	31	12	0.39
CLL186	*OCA2*	A>A/T	15	28,090,097	Splice Site	-	44	8	0.18
CLL077	*OCA2*	C>C/T	15	28,096,543	Missense	Gly775Ser	35	10	0.29
CLL003	*HERC2*	A>A/G	15	28,437,181	Missense	Ser2793Pro	47	13	0.28
CLL330	*THBS1*	T>T/A	15	39,880,810	Missense	Cys519Ser	44	20	0.45
CLL321	*PAK6*	G>G/T	15	40,564,522	Missense	Ser319Ile	40	23	0.58
CLL171	*TYRO3*	G>G/A	15	41,864,736	Missense	Ala617Thr	60	12	0.2
CLL330	*MGA*	TC>TC/T	15	41,962,023	Indel	Ser311Ter	38	11	0.29
CLL305	*MGA*	G>G/GT	15	42,041,328	Indel	Leu1842PhefsTer9	30	11	0.37
CLL077	*SLC12A1*	A>A/G	15	48,533,713	Missense	Asp406Gly	49	16	0.33
CLL144	*MYO5A*	A>A/G	15	52,725,448	Missense	Val21Ala	40	16	0.4
CLL171	*FAM214A*	G>G/A	15	52,901,667	Missense	Arg482Trp	40	18	0.45
CWL80	*FAM63B*	A>A/G	15	59,146,751	Missense	Glu603Gly	36	13	0.36
CLL348	*VPS13C*	C>C/T	15	62,276,096	Missense	Glu613Lys	40	19	0.48
CLL348	*CA12*	A>A/G	15	63,667,882	Splice Site	-	27	7	0.26
CLL371	*SLC51B*	A>A/T	15	65,343,854	Splice Site	-	33	18	0.55
CWL80	*LARP6*	G>G/A	15	71,125,137	Missense	Arg244Trp	36	20	0.56
CLL145	*THSD4*	G>G/A	15	71,507,745	Splice Site	-	44	20	0.45
CLL366	*SEMA7A*	T>T/C	15	74,709,001	Missense	Lys239Arg	28	9	0.32
CLL347	*PEAK1*	G>G/C	15	77,451,003	Missense	Pro1058Arg	34	12	0.35
CLL345	*ACSBG1*	G>G/A	15	78,472,067	Missense	Arg437Cys	34	18	0.53
CLL307	*TM6SF1*	TG>TG/T	15	83,793,468	Indel	Asp217IlefsTer2	31	19	0.61
CLL305	*BNC1*	G>G/A	15	83,926,459	Missense	Pro907Leu	33	12	0.36
JB210308	*TICRR*	G>G/C	15	90,127,629	Missense	Trp349Cys	31	11	0.35
CLL321	*ZNF774*	A>A/G	15	90,904,481	Missense	His473Arg	45	21	0.47
CLL144	*IQGAP1*	G>G/A	15	91,009,229	Splice Site	-	35	10	0.29
CLL371	*CHD2*	CA>CA/C	15	93,492,221	Indel	Gln473HisfsTer5	46	16	0.35
CLL346	*SYNM*	C>C/T	15	99,671,702	Missense	Arg1046Cys	27	12	0.44
CLL332	*RGS11*	C>C/T	16	321,425	Missense	Arg241Gln	30	12	0.4
CLL171	*RAB40C*	G>G/A	16	667,243	Missense	Gly60Ser	34	15	0.44
CLL160	*CACNA1H*	G>G/A	16	1,259,077	Missense	Ala1137Thr	27	13	0.48
CLL112	*TPSAB1*	G>G/A	16	1,291,159	Missense	Gly23Ser	35	12	0.34

CLL371	*MAPK8IP3*	G>G/A	16	1,816,361	Missense	Val923Ile	27	13	0.48
CLL186	*PKD1*	G>G/A	16	2,147,400	Missense	Ala3442Val	25	11	0.44
CLL345	*CREBBP*	G>G/A	16	3,786,819	Splice Site	-	34	22	0.65
CLL182	*CREBBP*	TGTGCTGGA>T GTGCTGGA/T	16	3,820,837	Indel	Leu869HisfsTer98	26	8	0.31
CLL347	*MAZ*	AAGGCC>AAG GCC/A	16	29,818,962	Indel	Lys286IlefsTer43	36	23	0.64
CLL149	*CCNYL3*	TGTCTGCCA>T GTCTGCCA/T	16	34,290,520	Indel	Ser390CysfsTer8	42	13	0.31
CLL144	*CHD9*	A>A/G	16	53,337,916	Missense	Asn2000Asp	32	9	0.28
CLL130	*IRX3*	G>G/T	16	54,319,908	Missense	Arg19Ser	36	16	0.44
CLL158	*CPNE2*	A>A/G	16	57,168,709	Missense	Asn380Ser	18	11	0.61
CLL158	*FAM192A*	T>T/C	16	57,188,212	Missense	Glu252Gly	35	19	0.54
CLL348	*PRSS54*	C>C/T	16	58,314,181	Missense	Glu379Lys	27	6	0.22
CLL154	*TMED6*	TTC>TTC/T	16	69,381,731	Indel	Lys150GlyfsTer7	47	12	0.26
CLL252	*PKD1L3*	T>T/C	16	71,963,589	Splice Site	-	53	23	0.43
CLL301	*CNTNAP4*	C>C/T	16	76,555,115	Missense	Ala742Val	29	12	0.41
CLL112	*ADAMTS18*	G>G/A	16	77,465,425	Nonsense	Arg88Ter	46	15	0.33
CLL160	*PKD1L2*	G>G/A	16	81,208,401	Missense	Ala901Val	45	12	0.27
CLL305	*WFDC1*	A>A/G	16	84,346,574	Missense	Glu51Gly	29	9	0.31
CLL006	*GSE1*	G>G/A	16	85,690,134	Missense	Arg392Gln	54	27	0.5
CLL156	*ZC3H18*	G>G/T	16	88,689,753	Splice Site	-	35	10	0.29
CLL252	*ANKRD11*	CG>CG/C	16	89,346,157	Indel	Ala2265ProfsTer72	37	11	0.3
CLL154	*ANKRD11*	TTC>TTC/T	16	89,349,133	Indel	Arg1272LysfsTer10	41	15	0.37
CLL158	*PRPF8*	G>G/A	17	1,577,866	Missense	Arg1057Trp	33	16	0.48
CLL158	*ANKFY1*	T>T/C	17	4,120,285	Missense	Arg193Gly	34	15	0.44
CLL371	*EIF5A*	G>G/A	17	7,214,714	Missense	Gly136Arg	27	10	0.37
CLL305	*GPS2*	CCCTAGAAAG GGAGAAGGGC TTCA>CCCTAG AAAGGGAGAA GGGCTTCA/C	17	7,217,306	Splice Site	-	26	9	0.35
CLL331	*GPS2*	AC>AC/A	17	7,217,608	Splice Site	-	19	17	0.89
CLL156	*TNK1*	G>G/A	17	7,291,719	Splice Site	-	47	14	0.3
CLL345	*LOC100996282*	G>G/A	17	7,335,668	Missense	Glu258Lys	32	13	0.41
CLL346	*POLR2A*	C>C/T	17	7,411,625	Missense	Ala1099Val	21	7	0.33
CLL145	*TP53*	C>C/T	17	7,577,120	Missense	Arg273His	49	24	0.49
CLL364	*TP53*	C>C/T	17	7,577,538	Missense	Arg248Gln	16	9	0.56
CLL332	*TP53*	T>T/C	17	7,578,190	Missense	Tyr220Cys	32	26	0.81
CLL252	*TP53*	A>A/C	17	7,578,204	Missense	Ser215Arg	16	10	0.63
CLL331	*TP53*	TGTAGATGGCC ATGGCGCGG> TGTAGATGGCC ATGGCGCGG/T	17	7,578,440	Indel	Arg158SerfsTer6	13	8	0.62
CLL345	*TP53*	T>T/G	17	7,578,536	Missense	Lys132Gln	29	13	0.45

CLL307	*TP53*	TGGTA>TGGTA/T	17	7,579,376	Indel	Tyr103ArgfsTer19	27	8	0.3
CLL307	*TP53*	G>G/A	17	7,579,382	Missense	Thr102Ile	23	7	0.3
CLL307	*TP53*	T>T/A	17	7,579,383	Missense	Thr102Ser	23	7	0.3
CWL80	*TP53*	G>G/A	17	7,579,717	Missense	Pro27Ser	20	7	0.35
CLL342	*ARHGEF15*	C>C/T	17	8,218,891	Missense	Arg474Trp	22	13	0.59
CWL80	*MYH10*	AT>AT/A	17	8,415,785	Indel	Met979CysfsTer12	45	12	0.27
CLL145	*DHRS7C*	G>G/A	17	9,676,195	Nonsense	Arg207Ter	45	28	0.62
CLL158	*DNAH9*	A>A/G	17	11,696,850	Missense	Lys2698Glu	30	13	0.43
CLL366	*ZNF286A*	G>G/A	17	15,619,550	Missense	Arg171Lys	43	8	0.19
CLL154	*NCOR1*	C>C/CT	17	15,974,831	Indel	Arg1349AlafsTer5	31	13	0.42
CLL366	*NCOR1*	C>C/G	17	16,012,099	Splice Site	-	34	11	0.32
CLL321	*SUZ12*	A>A/G	17	30,310,126	Splice Site	-	31	17	0.55
CLL344	*C17orf75*	C>C/G	17	30,669,083	Missense	Glu26Gln	31	13	0.42
CLL112	*ASIC2*	G>G/T	17	31,618,678	Nonsense	Tyr152Ter	24	13	0.54
CLL364	*SLFN12L*	A>A/C	17	33,801,971	Missense	Cys580Gly	31	14	0.45
CLL366	*IKZF3*	A>A/C	17	37,947,776	Missense	Leu162Arg	31	10	0.32
JB210308	*IKZF3*	A>A/C	17	37,947,776	Missense	Leu162Arg	35	15	0.43
CLL321	*CASC3*	ATGG>ATGG/A	17	38,319,088	Indel	Gly241del	50	15	0.3
CLL003	*KRTAP4-1*	C>C/T	17	39,340,769	Splice Site	-	28	10	0.36
CLL144	*KRT36*	C>C/T	17	39,644,900	Missense	Arg179Gln	43	19	0.44
CLL252	*KRT14*	C>C/T	17	39,742,686	Missense	Arg134His	39	12	0.31
CLL077	*GHDC*	A>A/G	17	40,345,464	Missense	Trp46Arg	38	17	0.45
CLL171	*PSME3*	G>G/C	17	40,986,400	Splice Site	-	43	14	0.33
CLL369	*CRHR1*	C>C/T	17	43,911,985	Splice Site	-	27	10	0.37
CLL158	*KAT7*	C>C/G	17	47,875,696	Missense	Ala119Gly	40	17	0.43
JB210308	*CA10*	T>T/C	17	49,726,575	Missense	Asn201Ser	50	21	0.42
CLL369	*STXBP4*	A>A/C	17	53,077,117	Missense	Ser138Arg	39	16	0.41
CLL347	*VMP1*	G>G/A	17	57,915,711	Missense	Ala344Thr	31	13	0.42
CLL321	*PPM1D*	A>A/T	17	58,740,824	Nonsense	Lys577Ter	33	15	0.45
CLL342	*CACNG4*	G>G/T	17	65,026,889	Missense	Glu251Asp	30	12	0.4
CLL301	*CASKIN2*	G>G/A	17	73,498,766	Missense	Arg797Trp	20	8	0.4
CLL321	*GALK1*	C>C/G	17	73,759,505	Missense	Gly124Ala	36	13	0.36
CLL301	*FBF1*	A>A/T	17	73,922,843	Missense	Asp183Glu	22	7	0.32
CLL342	*SRSF2*	G>G/A	17	74,732,365	Missense	Arg182Trp	24	10	0.42
CLL342	*ACTG1*	G>G/C	17	79,479,359	Missense	Leu8Val	38	17	0.45
CLL332	*L3MBTL4*	C>C/T	18	6,213,153	Missense	Val326Ile	44	14	0.32
CLL154	*SOGA2*	C>C/T	18	8,784,510	Missense	Thr467Met	28	15	0.54
CLL331	*ZNF521*	C>C/T	18	22,804,982	Missense	Arg967Gln	38	17	0.45
CLL307	*CCDC178*	C>C/A	18	30,926,174	Splice Site	-	53	11	0.21
CLL156	*TCEB3B*	C>C/T	18	44,561,056	Missense	Gly194Arg	52	19	0.37
CWL80	*ALPK2*	A>A/C	18	56,204,267	Missense	Phe1051Cys	37	15	0.41
CLL332	*ALPK2*	G>G/A	18	56,205,296	Missense	Ser708Leu	42	12	0.29
CLL301	*BCL2*	G>G/T	18	60,985,508	Missense	Ala131Asp	39	13	0.33
CLL063	*CDH19*	T>T/C	18	64,239,244	Splice Site	-	44	23	0.52

CLL063	*DSEL*	A>A/G	18	65,180,227	Missense	Ile550Thr	34	15	0.44
CLL351	*TMX3*	C>C/T	18	66,381,102	Missense	Val28Ile	27	14	0.52
CLL003	*NFATC1*	T>T/C	18	77,211,128	Splice Site	-	57	21	0.37
CWL80	*NFATC1*	C>C/T	18	77,227,543	Nonsense	Arg672Ter	27	11	0.41
CLL144	*PARD6G*	G>G/A	18	77,917,848	Missense	Pro313Ser	28	9	0.32
CLL252	*RPS15*	C>C/T	19	1,440,439	Missense	Ser139Phe	34	11	0.32
CLL330	*RPS15*	G>G/T	19	1,440,458	Missense	Lys145Asn	24	10	0.42
CLL332	*BTBD2*	G>G/C	19	1,990,039	Missense	Arg318Gly	40	15	0.38
CLL112	*AMH*	G>G/C	19	2,249,739	Missense	Glu136Asp	31	14	0.45
CLL006	*TMPRSS9*	C>C/T	19	2,424,182	Missense	Arg848Cys	48	20	0.42
CWL80	*DUS3L*	G>G/A	19	5,786,847	Missense	Arg467Cys	29	13	0.45
CLL369	*C3*	TG>TG/T	19	6,718,177	Splice Site	-	24	12	0.5
CLL144	*FBN3*	C>C/T	19	8,203,180	Missense	Arg349Gln	37	18	0.49
CLL145	*ZNF491*	T>T/G	19	11,909,544	Splice Site	-	54	22	0.41
CLL144	*ZNF763*	G>G/A	19	12,089,142	Missense	Glu138Lys	31	10	0.32
CLL332	*CCDC105*	C>C/T	19	15,132,337	Splice Site	-	37	13	0.35
CLL345	*NOTCH3*	C>C/T	19	15,272,228	Missense	Gly2071Arg	33	14	0.42
CLL156	*NOTCH3*	C>C/A	19	15,288,804	Missense	Arg1312Leu	28	15	0.54
CLL371	*CPAMD8*	G>G/A	19	17,010,309	Nonsense	Arg1656Ter	24	9	0.38
CLL301	*ZNF676*	G>G/A	19	22,363,009	Missense	Arg504Cys	41	20	0.49
CLL364	*ZNF91*	G>G/A	19	23,543,172	Missense	Thr870Met	43	10	0.23
JB210308	*ZNF536*	C>C/T	19	31,025,799	Missense	Ala739Val	38	19	0.5
CLL158	*CATSPERG*	C>C/T	19	38,861,163	Splice Site	-	28	18	0.64
CLL346	*NFKBIB*	CCT>CCT/C	19	39,398,225	Indel	Pro299ArgfsTer16	30	9	0.3
CLL003	*FCGBP*	G>G/T	19	40,362,885	Missense	Thr5062Asn	38	16	0.42
CWL80	*C19orf47*	G>G/A	19	40,832,340	Missense	Arg202Trp	36	20	0.56
CLL145	*PRX*	C>C/G	19	40,901,309	Missense	Gly984Arg	41	20	0.49
CLL307	*CYP2A7*	C>C/T	19	41,383,114	Missense	Arg381Gln	18	12	0.67
CLL344	*DEDD2*	T>T/G	19	42,721,780	Splice Site	-	30	14	0.47
CLL321	*ZNF180*	C>C/A	19	44,980,955	Missense	Gln581His	44	20	0.45
CLL171	*RSPH6A*	T>T/C	19	46,313,967	Missense	Asp261Gly	55	25	0.45
CLL301	*SPHK2*	C>C/T	19	49,129,346	Missense	Arg80Cys	34	12	0.35
CLL351	*MED25*	G>G/T	19	50,334,121	Missense	Ala360Ser	29	14	0.48
CLL321	*SHANK1*	G>G/A	19	51,189,591	Missense	Ala827Val	34	12	0.35
CLL342	*ACPT*	T>T/A	19	51,297,217	Missense	Met284Lys	35	11	0.31
JB210308	*CACNG8*	G>G/A	19	54,483,121	Missense	Arg123Gln	32	13	0.41
CLL347	*RPS9*	T>T/G	19	54,705,110	Missense	Phe20Val	32	10	0.31
CLL330	*ZNF580*	C>C/A	19	56,154,265	Missense	Arg131Ser	30	16	0.53
CLL301	*NLRP8*	C>C/T	19	56,466,176	Missense	Thr251Met	40	21	0.53
CLL369	*SOX12*	G>G/A	20	306,792	Missense	Gly75Asp	38	20	0.53
CLL348	*ANKEF1*	A>A/C	20	10,033,863	Missense	Lys658Asn	36	19	0.53
CLL348	*ANKEF1*	TA>TA/T	20	10,033,865	Indel	Gly661AlafsTer7	35	18	0.51
JB210308	*KIF16B*	T>T/C	20	16,348,288	Missense	Ile1228Val	40	15	0.38
CLL154	*BFSP1*	C>C/G	20	17,511,658	Missense	Arg106Pro	26	17	0.65
CLL156	*RRBP1*	T>T/C	20	17,601,357	Splice Site	-	21	10	0.48
CLL347	*RIN2*	C>C/T	20	19,937,365	Missense	Arg138Trp	27	11	0.41

CLL345	*APMAP*	G>G/A	20	24,950,241	Missense	Arg257Trp	41	18	0.44
CLL307	*BPIFB6*	A>A/G	20	31,630,645	Missense	Thr405Ala	35	10	0.29
CLL331	*NCOA6*	T>T/C	20	33,345,116	Missense	Met479Val	36	16	0.44
CLL186	*TRPC4AP*	G>G/T	20	33,632,322	Missense	Ala284Asp	32	14	0.44
CLL130	*NFS1*	C>C/T	20	34,263,103	Missense	Arg271His	35	19	0.54
JB210308	*RBM39*	T>T/A	20	34,313,042	Missense	Asp151Val	27	10	0.37
CLL077	*SAMHD1*	A>A/T	20	35,526,336	Missense	Phe545Leu	47	41	0.87
CWL80	*SAMHD1*	C>C/A	20	35,532,562	Missense	Asp501Tyr	39	13	0.33
CLL063	*IFT52*	A>A/G	20	42,265,888	Missense	Asn372Ser	31	13	0.42
JB210308	*NCOA3*	A>A/G	20	46,276,057	Missense	Asn1165Asp	35	16	0.46
CLL372	*ARFGEF2*	A>A/G	20	47,633,921	Missense	His1484Arg	47	22	0.47
CLL305	*KCNB1*	C>C/T	20	47,991,256	Missense	Glu281Lys	33	12	0.36
CLL144	*ADNP*	C>C/A	20	49,509,294	Missense	Val653Phe	40	18	0.45
CLL063	*CDH4*	G>G/A	20	60,427,955	Splice Site	-	42	19	0.45
CLL371	*POTED*	A>A/T	21	14,995,146	Missense	Asp359Val	16	7	0.44
CLL331	*BTG3*	C>C/T	21	18,966,481	Nonsense	Trp274Ter	45	21	0.47
JB210308	*TIAM1*	C>C/G	21	32,575,227	Missense	Glu830Asp	36	20	0.56
CLL145	*CHAF1B*	C>C/T	21	37,781,769	Splice Site	-	35	17	0.49
CLL112	*TTC3*	G>G/A	21	38,510,951	Missense	Met532Ile	43	21	0.49
CLL330	*MX2*	C>C/T	21	42,752,021	Missense	Arg174Trp	33	14	0.42
CLL006	*U2AF1*	C>C/T	21	44,514,662	Missense	Arg165Gln	42	18	0.43
CLL006	*SIK1*	C>C/T	21	44,839,738	Splice Site	-	39	14	0.36
CLL129	*HSF2BP*	C>C/T	21	45,050,224	Missense	Ala185Thr	40	19	0.48
CLL364	*PDXK*	C>C/T	21	45,168,940	Missense	Thr148Met	35	14	0.4
CLL301	*POTEH*	C>C/G	22	16,287,300	Missense	Val196Leu	24	7	0.29
CLL077	*BCL2L13*	C>C/T	22	18,210,041	Missense	Pro400Leu	40	20	0.5
CLL347	*LOC642643*	A>A/G	22	18,737,516	Missense	Asp338Gly	64	8	0.13
CLL369	*MED15*	C>C/T	22	20,939,218	Missense	Ala627Val	36	17	0.47
CLL171	*NC_000022*	A>A/C	22	23,029,598	Missense	Tyr67Ser	47	14	0.3
CLL171	*NC_000022*	G>G/A	22	23,040,689	Missense	Ser46Asn	48	16	0.33
CLL171	*NC_000022*	A>A/T	22	23,040,704	Missense	Tyr51Phe	39	11	0.28
CLL171	*NC_000022*	G>G/A	22	23,040,737	Missense	Gly62Asp	41	11	0.27
CLL156	*IGLL5*	G>G/C	22	23,230,278	Missense	Glu15Asp	41	28	0.68
CLL331	*IGLL5*	G>G/A	22	23,230,301	Missense	Arg23His	32	13	0.41
CLL342	*IGLL5*	G>G/C	22	23,230,363	Missense	Ala44Pro	21	19	0.9
CLL342	*IGLL5*	A>A/G	22	23,230,438	Missense	Arg69Gly	15	14	0.93
CLL374	*GGT1*	G>G/A	22	25,019,872	Missense	Asp337Asn	21	5	0.24
CLL350	*CRYBB3*	G>G/A	22	25,601,239	Missense	Arg127His	25	13	0.52
CLL171	*OSBP2*	G>G/T	22	31,266,577	Missense	Gly339Cys	50	26	0.52
CLL252	*SLC5A4*	T>T/C	22	32,651,232	Missense	Ile29Val	39	13	0.33
CLL003	*BPIFC*	C>C/T	22	32,831,696	Missense	Glu307Lys	59	24	0.41
CLL063	*FAM227A*	C>C/T	22	38,982,176	Missense	Arg556Lys	59	29	0.49
CLL144	*MCAT*	C>C/T	22	43,529,078	Missense	Asp382Asn	40	13	0.33
CLL332	*PNPLA3*	G>G/A	22	44,333,090	Missense	Arg306His	37	11	0.3
CLL366	*LDOC1L*	C>C/T	22	44,893,030	Missense	Arg136His	30	9	0.3
CLL156	*MAPK8IP2*	C>C/T	22	51,049,170	Missense	Arg813Cys	37	13	0.35

CLL149	*GPR143*	G>G/A	X	9,733,850	Missense	Ser3Phe	30	13	0.43
CLL301	*BEND2*	T>T/G	X	18,189,190	Missense	Ser706Arg	32	13	0.41
CLL063	*DMD*	G>G/T	X	31,986,508	Missense	Leu2188Met	38	17	0.45
JB210308	*FAM47C*	G>G/T	X	37,029,141	Missense	Lys886Asn	40	13	0.33
CLL347	*RPGR*	T>T/C	X	38,144,939	Missense	Lys1105Glu	26	12	0.46
CLL331	*TSPAN7*	G>G/A	X	38,533,526	Missense	Gly133Ser	22	22	1
CLL077	*BCOR*	C>C/T	X	39,913,134	Splice Site	-	21	20	0.95
CLL112	*DDX3X*	A>A/T	X	41,203,361	Missense	Ile282Phe	26	15	0.58
CLL006	*MED12*	G>G/C	X	70,339,253	Missense	Gly44Arg	20	18	0.9
CLL364	*MED12*	G>G/C	X	70,339,253	Missense	Gly44Arg	30	14	0.47
CLL371	*TAF1*	AGATGTAAGCT >AGATGTAAGC T/A	X	70,643,086	Splice Site	-	14	11	0.79
CLL063	*POF1B*	G>G/T	X	84,600,896	Nonsense	Cys231Ter	42	16	0.38
CLL006	*KLHL4*	C>C/G	X	86,873,001	Missense	Thr265Ser	16	15	0.94
CLL171	*KLHL4*	G>G/T	X	86,877,226	Missense	Val314Leu	52	24	0.46
CLL366	*RNF128*	G>G/A	X	106,034,395	Missense	Ala362Thr	27	14	0.52
CLL364	*COL4A5*	G>G/T	X	107,823,942	Missense	Gly289Cys	30	10	0.33
CLL145	*HTR2C*	G>G/A	X	114,141,205	Missense	Val202Met	23	21	0.91
CLL372	*SLC6A14*	C>C/T	X	115,590,000	Missense	Ala603Val	48	10	0.21
CLL369	*NKAP*	TCCCTTTGGTC TTTCTGTAAAC >TCCCTTTGGT CTTTCTGTAAA C/T	X	119,059,197	Indel	Val405_Gly411delin sdel	31	14	0.45
CLL345	*AIFM1*	T>T/A	X	129,272,652	Nonsense	Lys295Ter	42	22	0.52
CLL321	*LINC00893*	A>A/T	X	148,611,046	Splice Site	-	42	19	0.45
CLL112	*GABRA3*	C>C/A	X	151,424,470	Nonsense	Glu111Ter	21	15	0.71
CLL006	*NLGN4Y*	G>G/A	Y	16,942,161	Missense	Ala455Thr	16	13	0.81

Abbreviations: Chr, chromosome; VAF, variant allele frequency. Chromosomal coordinates refer to hg19 human genome build.

Appendix D: List of sequencing validation mutations

Sample ID	Gene	Chr	Coordinate (hg19)	Amino Acid Change	Whole Genome Sequencing			Targeted Sequencing		
					RD	VR	VAF	RD	VR	VAF
CLL371	*XPO1*	2	61,719,471	Glu571Gly	42	11	0.26	5000	1807	0.36
CLL160	*SF3B1*	2	198,266,611	Gly742Asp	45	15	0.33	4647	1570	0.34
CLL305	*SF3B1*	2	198,266,713	Gly740Glu	39	11	0.28	3801	1281	0.34
CLL158	*SF3B1*	2	198,266,821	Ile704Asn	31	13	0.42	2468	1220	0.49
CLL145	*SF3B1*	2	198,266,834	Lys700Glu	49	22	0.45	3341	1480	0.44
CLL307	*SF3B1*	2	198,266,834	Lys700Glu	30	19	0.63	256	164	0.64
CLL130	*SF3B1*	2	198,267,361	Lys666Glu	53	22	0.42	3504	1466	0.42
CLL316	*SF3B1*	2	198,267,371	His662Gln	29	6	0.21	1197	214	0.18
CLL369	*SF3B1*	2	198,267,484	Arg625Cys	42	19	0.45	1570	747	0.48
CLL156	*SF3B1*	2	198,267,489	Tyr623Cys	42	13	0.31	2430	1078	0.44
CLL332	*SF3B1*	2	198,267,491	Glu622Asp	33	10	0.30	3028	1140	0.38
CLL129	*SF3B1*	2	198,267,491	Glu622Asp	42	19	0.45	300	129	0.43
CLL344	*MYD88*	3	38,182,641	Leu273Pro	22	11	0.50	2445	1113	0.46
CLL144	*KLHL6*	3	183,211,965	Val418Met	28	12	0.43	3437	2169	0.63
CLL348	*KLHL6*	3	183,273,174	Leu90Phe	30	13	0.43	4682	3380	0.72
CLL348	*KLHL6*	3	183,273,179	Val88Glu	31	15	0.48	4682	1094	0.23
CLL348	*KLHL6*	3	183,273,242	Met67Thr	33	12	0.36	578	253	0.44
CLL348	*KLHL6*	3	183,273,246	Arg66Gly	33	12	0.36	1200	555	0.46
CLL330	*FBXW7*	4	153,247,310	Met498Leu	33	15	0.45	3231	1348	0.42
CLL006	*IRF4*	6	394,954	Asp117Val	40	4	0.10	1726	196	0.11
CLL145	*POT1*	7	124,503,668	Gln94His	53	13	0.25	1882	745	0.40
CLL372	*BRAF*	7	140,453,150	Phe595Leu	45	8	0.18	2942	277	0.09
CLL347	*BRAF*	7	140,453,193	Asn581Ser	38	9	0.24	116	24	0.21
CLL186	*NOTCH1*	9	139,390,648	Pro2514ArgfsX4	36	12	0.33	5000	1533	0.31
CLL331	*NOTCH1*	9	139,390,648	Pro2514ArgfsX4	36	13	0.36	100	45	0.45
CLL130	*NOTCH1*	9	139,390,648	Pro2514ArgfsX4	40	16	0.40	5000	1880	0.38
JB 210308	*NOTCH1*	9	139,390,648	Pro2514ArgfsX4	22	13	0.59	3748	1845	0.49
CLL317	*NOTCH1*	9	139,390,648	Pro2514ArgfsX4	37	21	0.57	217	121	0.56
CLL158	*NOTCH1*	9	139,390,678	Pro2505Thr	30	9	0.30	87	44	0.51
CLL186	*EGR2*	10	64,573,248	His384Asn	34	14	0.41	1992	906	0.45
CLL330	*EGR2*	10	64,573,332	Glu356Lys	30	8	0.27	5000	1584	0.32
CLL369	*BIRC3*	11	102,201,871	Asn409IlefsTer4	14	4	0.29	1116	569	0.51
CLL330	*BIRC3*	11	102,207,782	Cys588Ter	17	14	0.82	2189	1841	0.84
CLL144	*ATM*	11	108,121,624	Ser478Ala	48	17	0.35	4127	1889	0.46
CLL144	*ATM*	11	108,121,625	Ser478Leu	48	17	0.35	4116	1874	0.46
CLL112	*ATM*	11	108,199,953	Ile2432Asn	31	8	0.26	2285	883	0.39
CLL321	*ATM*	11	108,201,099	Ser2489Phe	15	14	0.93	1769	1743	0.99
CLL372	*ATM*	11	108,203,577	Ile2629SerfsTer25	29	20	0.69	2193	1872	0.85
CLL182	*ATM*	11	108,224,493	Gly2891Asp	18	4	0.22	2302	305	0.13
CLL369	*ATM*	11	108,236,116	Lys3018Gln	23	22	0.96	1069	987	0.92
CLL307	*TP53*	17	7,577,018	c.919+1G>A	25	6	0.24	1136	285	0.25
CLL145	*TP53*	17	7,577,120	Arg273His	49	24	0.49	1967	981	0.50

CLL364	*TP53*	17	7,577,538	Arg248Gln	16	9	0.56	1701	1091	0.64
CLL332	*TP53*	17	7,578,190	Tyr220Cys	32	26	0.81	4999	3950	0.79
CLL252	*TP53*	17	7,578,204	Ser215Arg	16	10	0.63	3189	1955	0.61
CLL317	*TP53*	17	7,578,280	Pro190Leu	24	11	0.46	4990	2345	0.47
CLL331	*TP53*	17	7,578,440	Arg158SerfsTer6	13	8	0.62	4344	4142	0.95
CLL145	*TP53*	17	7,578,455	Ala159Pro	43	4	0.09	3992	379	0.09
CLL345	*TP53*	17	7,578,536	Lys132Gln	29	13	0.45	2575	1186	0.46
CLL307	*TP53*	17	7,579,376	Tyr103ArgfsTer19	27	8	0.30	311	38	0.12
CLL307	*TP53*	17	7,579,382	Thr102Ile	23	7	0.30	298	38	0.13
CLL307	*TP53*	17	7,579,383	Thr102Ser	23	7	0.30	298	38	0.13
L80	*TP53*	17	7,579,717	Pro27Ser	20	7	0.35	111	25	0.23
L80	*SAMHD1*	20	35,532,562	Asp501Tyr	39	13	0.33	2311	985	0.43
CLL307	*DDX3X*	X	41,204,790	Leu435Pro	18	18	1.00	96	96	1.00
CLL364	*MED12*	X	70,339,253	Gly44Arg	30	14	0.47	4739	2217	0.47
CLL006	*MED12*	X	70,339,253	Gly44Arg	20	18	0.90	610	598	0.98

Abbreviations: Chr, chromosome; RD, total read depth; VR, number of reads containing the variant; VAF, variant allele frequency. Chromosomal coordinates refer to hg19 human genome build.

Appendix E: Full List of Kataegis Regions

Sample ID	Chr	Region Start (hg19)	Region End (hg19)	Number of Mutations	Size (bp)	Muts/bp	Genes in Region
CLL144	2	89,159,052	89,185,267	51	26,215	0.001945	*IGK*
CLL331	2	89,157,593	89,165,310	32	7,717	0.004147	*IGK*
CLL342	2	89,140,694	89,161,003	32	20,309	0.001576	*IGK*
CLL149	2	89,158,625	89,247,194	28	88,569	0.000316	*IGK*
CLL301	2	89,159,105	89,400,894	28	241,789	0.000116	*IGK*
CLL344	2	89,159,538	89,160,713	23	1,175	0.019574	*IGK*
CLL346	2	89,158,227	89,185,578	23	27,351	0.000841	*IGK*
CLL348	2	89,157,619	89,327,508	23	169,889	0.000135	*IGK*
CLL145	2	89,066,584	89,160,438	13	93,854	0.000139	*IGK*
CLL374	2	89,159,642	89,161,244	11	1,602	0.006866	*IGK*
CLL156	2	140,885,911	141,045,820	8	159,909	0.00005	*LRP1B*
CLL348	3	187,660,014	187,660,463	9	449	0.020045	-
CLL154	3	157,290,339	157,295,704	6	5,365	0.001118	*C3orf55*
CLL348	3	183,273,058	183,273,364	6	306	0.019608	*KLHL6*
CLL371	4	59,097,561	59,114,524	11	16,963	0.000648	-
CLL301	5	26,070,177	26,074,185	10	4,008	0.002495	-
CLL063	5	21,810,369	21,843,504	9	33,135	0.000272	*CDH12*
CLL301	7	122,433,699	122,517,296	12	83,597	0.000144	*CADPS2*
CLL371	7	49,672,829	49,682,118	9	9,289	0.000969	-
CLL301	7	122,622,586	122,638,447	6	15,861	0.000378	*TAS2R16*
CLL351	9	123,416,699	123,479,606	43	62,907	0.000684	*MEGF9*
CLL342	9	76,656,503	76,656,760	6	257	0.023346	-
CLL144	11	108,121,624	108,129,499	9	7,875	0.001143	*ATM*
CLL307	13	51,664,141	51,665,954	6	1,813	0.003309	*LINC00371*
CLL144	14	106,152,090	107,259,530	99	1,107,440	0.000089	*IGH*
CLL171	14	106,151,270	107,259,533	85	1,108,263	0.000077	*IGH*
CLL145	14	106,211,966	106,330,421	57	118,455	0.000481	*IGH*
CLL348	14	106,326,027	107,259,254	52	933,227	0.000056	*IGH*
CLL301	14	106,323,279	107,179,197	51	855,918	0.00006	*IGH*
CLL331	14	106,213,235	107,259,530	50	1,046,295	0.000048	*IGH*
CLL156	14	106,325,382	107,178,939	46	853,557	0.000054	*IGH*
CLL346	14	106,240,547	107,179,603	46	939,056	0.000049	*IGH*
CLL063	14	106,211,717	107,170,126	34	958,409	0.000035	*IGH*
CLL344	14	106,325,822	107,176,711	33	850,889	0.000039	*IGH*
CLL371	14	106,210,618	106,726,347	33	515,729	0.000064	*IGH*
CWL80	14	106,112,882	106,376,235	32	263,353	0.000122	*IGH*
CLL342	14	106,213,606	107,179,703	31	966,097	0.000032	*IGH*
CLL351	14	106,213,664	107,259,538	31	1,045,874	0.00003	*IGH*
CLL154	14	106,329,401	107,269,172	29	939,771	0.000031	*IGH*
CLL160	14	106,213,001	107,179,214	29	966,213	0.00003	*IGH*
CLL149	14	106,325,402	107,179,022	23	853,620	0.000027	*IGH*

CLL372	14	106,325,813	106,800,347	21	474,534	0.000044	*IGH*
CLL252	14	106,240,007	106,748,391	17	508,384	0.000033	*IGH*
CLL158	14	106,176,677	106,327,757	15	151,080	0.000099	*IGH*
CLL003	14	106,239,549	106,326,360	14	86,811	0.000161	*IGH*
CLL145	14	106,725,391	107,179,022	12	453,631	0.000026	*IGH*
CLL307	14	106,212,147	106,326,741	12	114,594	0.000105	*IGH*
CLL369	14	106,326,563	106,384,422	11	57,859	0.00019	*IGH*
CLL112	14	106,324,258	106,326,713	10	2,455	0.004073	*IGH*
CLL130	14	106,237,375	106,392,658	9	155,283	0.000058	*IGH*
CLL129	14	106,212,023	106,326,234	8	114,211	0.00007	*IGH*
CLL182	14	106,323,732	107,049,039	8	725,307	0.000011	*IGH*
CLL366	14	106,112,760	106,325,372	8	212,612	0.000038	*IGH*
CLL186	14	106,239,597	106,329,751	7	90,154	0.000078	*IGH*
CLL156	14	21,835,327	21,835,546	6	219	0.027397	*SUPT16H*
CLL321	14	106,211,966	106,673,293	6	461,327	0.000013	*IGH*
CLL344	14	104,922,215	106,112,736	6	1,190,521	0.000005	*IGH*
CLL301	18	60,873,525	60,988,029	12	114,504	0.000105	*BCL2*
CLL348	18	60,906,440	60,988,117	10	81,677	0.000122	*BCL2*
CLL252	18	9,284,149	9,374,417	6	90,268	0.000066	*ANKRD12, TWSG1*
CLL348	22	23,198,941	23,233,399	50	34,458	0.001451	*IGL*
CLL156	22	23,040,749	23,247,495	21	206,746	0.000102	*IGL*
CLL331	22	23,230,222	23,232,006	19	1,784	0.01065	*IGL*
CLL171	22	23,040,689	23,247,586	17	206,897	0.000082	*IGL*
CLL351	22	22,516,929	23,228,510	16	711,581	0.000022	*IGL*
CLL371	22	23,236,283	23,247,607	15	11,324	0.001325	*IGL*
CLL144	22	23,223,275	23,231,983	14	8,708	0.001608	*IGL*
CLL342	22	23,153,963	23,247,619	14	93,656	0.000149	*IGL*
CLL346	22	23,223,254	23,231,457	13	8,203	0.001585	*IGL*
CLL344	22	23,223,277	23,231,043	9	7,766	0.001159	*IGL*
CLL145	22	23,055,186	23,055,480	8	294	0.027211	*IGL*
CLL063	22	23,241,895	23,247,626	6	5,731	0.001047	*IGL*
CLL369	22	23,230,520	23,236,398	6	5,878	0.001021	*IGL*
CLL301	23	98,765,633	98,769,284	10	3,651	0.002739	*XRCC6P5*

Appendix F: Publications arising from this work

DELETED FROM ELECTRONIC VERSION

Publications

BURNS A, ALSOLAMI R, BECQ J, TIMBS A, BRUCE D, ROBBE P, VAVOULIS D, CABES M, DREAU H, TAYLOR J, KNIGHT SJ, MANSSON R, BENTLEY D, RIDOUT K, SCHUH A (2017) Whole genome sequencing reveals differences in the mutation spectrum of IgHV hypermutated and unmutated Chronic Lymphocytic Leukemia. Under Review.

FERNANDEZ-MERCADO M, **BURNS A**, PELLAGATTI A, GIAGOUNIDIS A, GERMING U, AGIRRE X, PROSPER F, AUL C, KILLICK S, WAINSCOAT JS, SCHUH A AND BOULTWOOD J (2013) Targeted re-sequencing analysis of 25 genes commonly mutated in myeloid disorders in del(5q) myelodysplastic syndromes. *Haematologica* 98(12): 1856–64

Abstracts

BURNS A, WELLER S, HUMPHRAY S, BECQ J, KINGSBURY Z, ALSOLAMI R, HAMBLIN A, CLIFFORD R, KNIGHT S, ROBBE P, CABES M, TIMBS A, MASON J, HATTON C, TAYLOR J, BENTLEY D, SCHUH A (2014) Whole Genome Sequencing, Targeted Deep Sequencing and Copy-Number Analysis Provides a Comprehensive Picture of the Mutational Landscape in 41 Clinically Annotated CLL Cases; 56th American Society of Hematology Annual Meeting, San Francisco

BURNS A, CLIFFORD R, DREAU H, HATTON C, HENDERSON S, TAYLOR J, SCHUH A (2012) Targeted gene profiling identifies differential genetic make-up depending on chronic lymphocytic leukaemia subtype; 54th American Society of Hematology Annual Meeting, Atlanta