Oncogenomic Screening Strategies to Identify Driver Genes in Acute Myeloid Leukaemia

A thesis submitted in part requirement for the degree of

Doctor of Philosophy from the Faculty of Medical Sciences

Jake Clayton

Northern Institute for Cancer Research

Faculty of Medical Sciences

Newcastle University

Newcastle-Upon-Tyne

April 2015

This thesis is dedicated to my father, Maurice.

Abstract

Deletions of the short arm of chromosome 12 (12p) are found in around 6% of acute myeloid leukaemia (AML). Particularly in paediatric AML they often occur as the sole cytogenetic change and are associated with a poor prognosis. Amplifications of the long arm of chromosome 11 (11q) are rarer, found in around 1% of AML, but are particularly prevalent in the poor prognosis complex karyotype group.

Despite multiple deletion mapping studies, single genes have not been identified from these regions that are responsible for driving leukaemic progression, thus it is clear that a functional approach is required. This study aimed to functionally implicate significant genes through the use of competitive selection assays.

Publicly available array data from eight studies were used to determine copy number from AML patient samples to define minimal regions of copy number alteration. Data were obtained from 866 samples which had been analysed on Affymetrix SNP array platforms. In total 58 samples (6.7%) were found to have deletions of 12p and were used to determine a minimally deleted region (MDR), whilst five samples (0.6%) were identified with amplifications of 11q and used to define a minimally amplified region (MAR).

AML cell lines with deletions of 12p (NKM-1 and GDM-1) and amplification of 11q (UoC-M1) were selected to investigate the functional relevance of the genes within the regions of interest. Cell lines were used *in vitro* and in immunocompromised mouse models, where leukaemia was established by intrafemoral injection and monitored by luminescent imaging of luciferase expression constructs. Taking a pooled approach, 11 genes within the 12p MDR were overexpressed and 30 genes from the 11q MAR were knocked down using integrated lentiviral vectors. To evaluate effects of expression on leukaemia growth or survival, changes in construct copy number after cell line expansion *in vitro* and *in vivo* were determined by targeted high throughput sequencing on the Illumina MiSeq platform.

Demonstrating a strong anti leukaemic effect for expression of this gene, a construct for cyclin dependent kinase inhibitor 1B (*CDKN1B*) was rapidly selected against in the 12p assay, whilst knockdown of *MPZL3* and *UBE4A* from 11q were selected against in the 11q assay. Functional follow up work mainly focussed on *CDKN1B*, where its expression was measured in a range of a cell lines and patient samples, its effects on the cell cycle assessed and correlation of *CDKN1B* expression and treatment with a cyclin dependent kinase inhibitor (Flavopiridol) was established.

Acknowledgements

Firstly, I wish to thank the European Research Council for providing the funding which made my PhD studentship possible. I would like to thank my supervisors Dr. Paul Sinclair and Prof. Christine Harrison for their valuable support and guidance throughout my time as a PhD student. I am also very grateful to Dr. Jim Allan and Dr. Dan Williamson for their detailed feedback and advice as members of my PhD review panel.

I would also like to thank Dr. Michelle Le Beau, Prof. Olaf Heidenreich and Dr. Sarah Fordham for donating cell lines for use in this project. Prof. Olaf Heidenreich also kindly donated a number of lentiviral plasmids that were extremely important to the completion of my work. I also wish to thank Prof. Claudia Haferlach for providing AML patient samples that were invaluable in this project.

With regards to the mouse work, my thanks go to Dr. Helen Blair, Dr. Lars Buechler and Christopher Huggins, who passed on their wealth of experience and provided assistance in both mouse handling and performing delicate surgeries. I would also like to thank all of the staff at the Comparative Biology Centre, who maintain and care for the mice on a daily basis, without whom the mouse work would not have been possible.

I wish to thank all the members of the LRCG, past and present, who helped out regularly and answered many of my questions. Particular thanks go to Joanna Cheng, who taught me many of the lab techniques required for this project, Dr. Alem Gabriel for teaching me how to analyse SNP microarray data and Dr. Sarra Ryan who greatly assisted in the analysis and interpretation of high throughput sequencing results.

Many thanks go to my friends and family, as without their encouragement, support and help, this thesis would not have been possible. Finally, special thanks go to my parents, for encouraging me to aim high and for believing in my continued education. I promise I'm done being a student!

Statement of Work Undertaken

The SIN-SIEW-Gateway construct used in this project was produced by Dr. Paul Sinclair from the SIN-SIEW construct kindly gifted by Prof. Olaf Heidenreich. The primers designed to amplify the shRNA sequence present in GIPZ lentiviral constructs were also designed by Dr. Paul Sinclair.

Mouse stocks were maintained by the staff of the Comparative Biology Centre. Dr. Helen Blair injected the mice used to produce Figure 31, in addition to monitoring mouse health on a regular basis. Dr. Helen Blair, Dr. Paul Sinclair, Dino Mašić and Dr. Stefano Tonin, in addition to myself, have all shared the responsibilities of regular mouse health and weight checks throughout the course of this project.

Sequencing reactions to verify construct inserts were performed by DBS Genomics (Durham University). Edinburgh Genomics (University of Edinburgh) performed high throughput sequencing runs on the Illumina MiSeq platform. Dr. Sarra Ryan prepared the NGS analysis software to count the number of reads associated to each construct used in this project and demonstrated how to process raw read files.

All other lab work and data analysis was performed by myself.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
aCGH	Array Comparative Genomic Hybridisation
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Myeloid Leukaemia
BAC	Bacterial Artificial Chromosome
BCA	Bicinchoninic Acid
BM	Bone Marrow
BN	Both Notch
BSA	Bovine Serum Albumin
cDNA	Complimentary Deoxyribonucleic Acid
CGH	Comparative Genomic Hybridisation
CLL	Chronic Lymphocytic Leukaemia
CML	Chronic Myeloid Leukaemia
CMP	Committed Myeloid Progenitor
CMV	Cytomegalovirus
COSMIC	Catalogue of Somatic Mutations in Cancer
DAPI	4',6-diamidino-2-phenylindole
ddNTPs	Dideoxynucleotide Triphosphates
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide
dUTP	Deoxyuridine Triphosphate
ECL	Enhanced Chemiluminscence
EDTA	Ethylenediaminetetraacetic Acid
eGFP	Enhanced Green Fluorescent Protein
ENU	N-ethyl-N-nitrosourea
ETP	Earliest Thymic Progenitor
FAB	French-American-British
FACS	Fluorescence Activated Cell Sorting
FCS	Foetal Calf Serum
FISH	Fluorescence *In Situ* Hybridisation
FITC	Fluorescein Isothiocyanate 1
FIV	Feline Immunodeficiency Virus

GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GI50	Growth Inhibition of 50%
GMP	Granulocytic-Monocytic Progenitor
HIV	Human Immunodeficiency Virus
HRP	Horseradish Peroxidase
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem Progenitor Cell
IKMC	International Knockout Mouse Consortium
IRES	Internal Ribosome Entry Site
ITD	Internal Tandem Duplication
IVC	Individually Ventilated Cage
IVIS	*In Vivo* Imaging System
LB	Lysogeny Broth
LBA	Lysogeny Broth Agar
Lin	Lineage Markers
LN	Left Notch
LOH	Loss of Heterozygosity
LRCG	Leukaemia Research Cytogenetics Group
LSC	Leukaemic Stem Cell
LTR	Long Terminal Repeat
MAR	Minimally Amplified Region
MB	Megabases
MDR	Minimally Deleted Region
MDS	Myelodysplastic Syndrome
MEP	Megakaryocytic-Erythrocytic Progenitor
miRNA	Micro Ribonucleic Acid
MLP	Immature Lymphoid Progenitor
MLPA	Multiplex Ligation-dependent Probe Amplification
MPP	Multipotent Progenitor
MRD	Minimal Residual Disease
mRNA	Messenger Ribonucleic Acid
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
NADPH	Nicotinamide Adenine Dinucleotide Phosphate

NGS	Next Generation Sequencing
NK	Natural Killer
NN	No Notch
NOD	Non Obese Diabetic
NOG	NOD/Shi-scid/IL-2Rγ^{null}
NSC	Non Silencing Control
NSG	NOD *Scid* Gamma
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PB	Peripheral Blood
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PE	Phycoerythrin
PI	Propidium Iodide
PVDF	Polyvinylidene Difluoride
qPCR	Quantitative Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
RN	Right Notch
RNA	Ribonucleic Acid
RNAi	Ribonucleic Acid Interference
ROS	Reactive Oxygen Species
RPMI	Roswell Park Memorial Institute
RRE	*rev* Response Element
RSV	Rous Sarcoma Virus
SB	Sleeping Beauty
Scid	Severe Combined Immune Deficient
SDS	Sodium Dodecyl Sulphate
SFFV	Spleen Focus Forming Virus
shRNA	Small Hairpin Ribonucleic Acid
SIN	Self Inactivating
siRNA	Small Interfering Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth with Catabolite Repression
t-AML	Therapy Related Acute Myeloid Leukaemia

TBS-T	Tris Buffered Saline-Tween-20
TE	Tris(hydroxymethyl)aminomethane-Ethylenediaminetetraacetic Acid
tGFP	Turbo Green Fluorescent Protein
TU	Transducing Units
UK	United Kingdom
USA	United States of America
UTR	Untranslated Region
VSV	Vesicular Stomatitis Virus
WHO	World Health Organization
WPRE	Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element

Chapter 1 Introduction

1.1 Haematopoiesis

Haematopoiesis is the process that drives the formation of the cellular components of blood. It allows the production of a wide range of cell types from a single type of progenitor cell – the multipotent haematopoietic stem cell (HSC). These stem cells reside within the bone marrow (BM) and give rise to approximately one trillion (10^{12}) new cells per day in an adult human (Doulatov *et al.*, 2012). HSCs demonstrate two main capabilities common to all stem cells. The first is that they are able to self-renew and produce additional HSCs with the same ability, and the second is that they are able to differentiate into all blood cell types (Seita and Weissman, 2010).

HSCs are defined operationally by their ability to reconstitute the entire blood system of a recipient who is haematopoietically deficient. This was first demonstrated in mice and guinea pigs who had been subjected to lethal doses of radiation, but were subsequently rescued following the injection of spleen or BM cells from unaffected donors (Lorenz *et al.*, 1951).

The development of fluorescence-activated cell sorting (FACS) techniques has been instrumental in tracking the range of cell populations present in the blood. This is achieved through the tracking of cell surface antigens, which are cell specific markers presented on cell membranes that can be detected using fluorescently labelled antibodies. Using FACS, HSCs can be purified by separating them from their more mature counterparts. Studies on these highly pure populations have demonstrated that a single HSC is able to provide haematopoietic reconstitution for a recipient for up to 4 months (Purton and Scadden, 2007).

The products of HSC differentiation can be most loosely grouped into two fundamental branches – lymphoid and myeloid. Lymphoid lineage cells consist of T, B and Natural Killer (NK) cells, and their roles lie mainly in immune response. Myeloid lineage cells consist of a number of distinct and short-lived cell types, including granulocytes, monocytes, erythrocytes and megakaryocytes (Seita and Weissman, 2010). Myeloid cells have a range of roles, from mounting immune responses to producing red blood cells and platelets. The differentiation process can be summarised in the form of a flow chart, with HSCs residing at the top of this hierarchy (Figure 1). As cells become more differentiated, they lose their ability to self-renew and the cell surface markers that they express change.

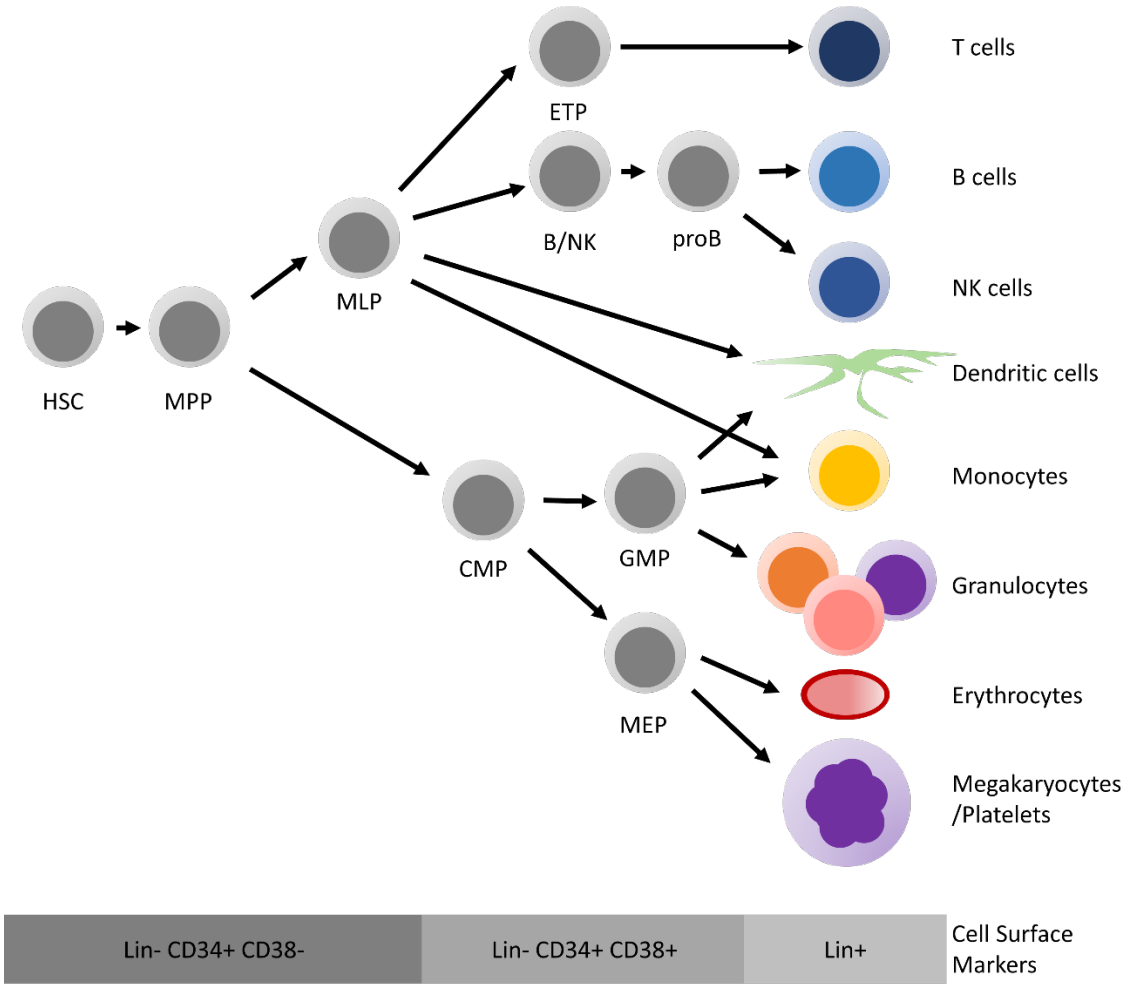


Figure 1. The adult human haematopoietic hierarchy. Haematopoietic stem cells (HSCs) give rise to the entire complement of human blood cell types, through a number of intermediary cell types. HSCs are able to self-renew and differentiate into multipotent progenitors (MPPs), which have a reduced self-renewal capacity. These go on to differentiate into the precursors of the two main lineages – immature lymphoid progenitors (MLPs) and commited myeloid progenitors (CMPs). MLPs give rise to earliest thymic progenitors (ETPs) which in turn produce T cells. CMPs differentiate into granulocytic-monocytic progenitors (GMPs) and megakaryocytic-erythrocytic progenitors (MEPs). Lineage markers (Lin) refer to a cocktail of markers that can be used to identify terminally differentiated haematopoietic cells.

1.2 Haematological Malignancies

Normal cells involved in haematopoiesis can acquire mutations that lead to a block in differentiation or provide a growth advantage. If these cells go on to replicate, this can lead to the development of a haematological malignancy. These diseases are primarily classified by the cell lineage involved, and this is typically determined through the staining and light microscopy of either peripheral blood or a lymph node biopsy. The numbers of patients diagnosed each year in the UK with a haematological malignancy is listed in Table 1 (Yorkshire Humberside Haematology Research Network, 2014).

Disease	**Children aged 0-14 years**	**Young adults aged 15-24 years**	**Adults 25+ years**	**All ages**	**% of total**
Acute lymphoblastic leukaemia (ALL)	370	90	290	750	2.65
Acute myeloid leukaemia (AML)	70	90	2090	2250	7.94
Chronic myeloid leukaemia (CML)	-	20	530	550	1.94
Chronic lymphocytic leukaemia (CLL)	-	-	3300	3300	11.64
Other leukaemias	20	10	670	700	2.47
Leukaemia (total)	**460**	**210**	**6880**	**7600**	**26.81**
Hodgkin lymphoma	70	250	1330	1650	5.82
Non-Hodgkin lymphoma	100	80	8820	9000	31.75
Other lymphoproliferative disorders	-	-	1050	1050	3.70
Lymphoma (total)	**170**	**330**	**11200**	**11700**	**41.28**
Myeloma (total)			**3750**	**3750**	**13.23**
Other haematological malignancies	10	35	-	45	0.16
Myelodysplastic syndromes	-	-	2000	2000	7.06
Myeloproliferative neoplasms	-	-	3300	3300	11.64
Other haematological malignancies (total)	**10**	**35**	**5300**	**5345**	**18.86**
All haematological malignancies (total)	**640**	**575**	**27130**	**28345**	**100.00**

Table 1. Number of people diagnosed each year with a haematological malignancy in the UK.

1.3 Acute Myeloid Leukaemia

1.3.1 Introduction

Acute myeloid leukaemia (AML) is the most common myeloid malignancy, and accounts for around 0.9% of all newly diagnosed cancers in the UK (CRUK 2014). AML carries the worst prognosis of the 4 main subtypes of leukaemia, with a median survival of around 9.5 months (Bhayat *et al.*, 2009). It is likely that this extremely poor overall survival rate is at least in part due to the majority of AML patients being older, with almost three quarters of cases being diagnosed in patients over the age of 60. These older patients are less likely to be able to tolerate the high dose chemotherapy required to achieve complete remission of AML. Despite a better overall outlook in paediatric cases, survival rates are still low when compared to other childhood leukaemias, with a 10 year event-free survival of 49% (Harrison *et al.* 2010).

1.3.2 Epidemiology

The annual incidence rate for AML in the UK can be estimated to be around 5 new AML cases for every 100,000 males, and 4 new cases for every 100,000 females. In 2011, this equated to 2921 new cases of AML being diagnosed: 1608 (55%) in males and 1313 (45%) in females (CRUK, 2014).

A number of factors that may increase an individual's risk of developing AML have been identified. Exposure to nuclear radiation, for example following the atomic bombing of Japan in 1945 (Nakanishi *et al.*, 1999) as well as working in the nuclear industry (Cardis *et al.*, 1995) have both been associated with an increased risk of AML, specifically with an increased incidence of chromosome 5 and 7 abnormalities. Radiation exposure in the cockpit of commercial jet airliners has also been associated with AML risk, with one study finding a 5 times increased incidence of AML in Danish male cockpit crew with over 5000 hours flying time (Gudestrup and Storm, 1999).

Exposure to the organic chemical compound benzene has been linked to AML development. The largest group exposed to benzene are cigarette smokers, who have been shown to have between a 1.2 and 2.3 times increased incidence of AML (Kane *et al.*, 1999). Occupational exposure to benzene has also been demonstrated to greatly increase AML risk, with a large study of almost 75,000 benzene exposed workers in China demonstrating an increased incidence of AML, and an increased incidence of chromosome 5 and 7 aberrations in these patients' leukaemias (Travis *et al.*, 1994).

Between 10-15% of patients who develop AML do so after treatment with chemotherapy for another cancer (therapy-related acute myeloid leukaemia; t-AML) (Smith *et al.*, 2003). These patients can be roughly divided into two classes. The first of these develop AML between 1 and

5 years post treatment with DNA topoisomerase II inhibitors such as doxorubicin and etoposide. These patients often have cytogenetic abnormalities involving the long arm of chromosome 11 (11q23 rearrangements involving *MLL*) and balanced translocations between chromosomes 15 and 17, t(15;17)(q24;q21) *PML/RARα*. The second group develop AML approximately 5-10 years after exposure to DNA damaging alkylating agents, and have abnormalities of chromosomes 5 and/or 7, usually either as the loss of part of the long arm (-5q/-7q) or as monosomy of the entire chromosome (-5/-7) (Smith *et al.*, 2003).

Patients with myelodysplastic syndrome (MDS) are at a high risk of developing AML. Approximately one third of patients with MDS will go on to develop AML (Barzi and Sekeres, 2010), and they have a particularly poor prognosis when compared to other AML patients. MDS patients most at risk of developing AML are typically younger, present with higher blast counts, a history of chemotherapy and have multiple karyotypic abnormalities (Shi *et al.*, 2004).

1.3.3 Origins of Disease

It is widely believed that AML leukaemogenesis requires a number of cooperating genetic changes. Mouse models of AML have found that a single genetic aberration, for example the expression of t(8;21)(q24;q22) *RUNX/RUNX1T1* or inv(16)(p13;q22) *CBFB-MYH11* fusion genes, impair myeloid differentiation but do not alone cause an overt leukaemic phenotype (Frohling *et al.*, 2005). Additionally, germline mutations in genes such as *RUNX1* and *CEBPA* provide a genetic predisposition to AML, but do not always result in complete disease development. If leukaemia does develop in these patients, it is likely to carry additional acquired somatic mutations. Further supporting this hypothesis is the observation that many molecular events in AML patients are observed in conjunction with another mutation, for example the frequent association between *FLT3*-Internal tandem duplications (*FLT3*-ITD) mutations and the expression of a fusion transcript such as *RUNX1/RUNX1T1* or *PML/RARA* (Dash and Gilliland, 2001). The high rate of transformation of MDS into AML supports this multi-step theory of leukaemogenesis, where initial mutations in normal haematopoietic cells are not able to induce leukaemia, but do confer a growth advantage or differentiation block. These cells often go on to acquire additional detectable abnormalities, which result in the transformation of MDS to AML (Ripperger *et al.*, 2009).

Approximately 55% of *de novo* AML cases have cytogenetically detectable abnormalities (Sanderson *et al.*, 2006). The most common of these are balanced translocations, where no genetic material is gained or lost, but the genetic location of a gene or genes is altered. Balanced translocations typically have one of two effects. Firstly, they can lead to aberrant

gene expression, for example when a gene is translocated to be in close proximity to a highly active promoter region (e.g. t(12;18)(p13;q12) *ETV6/SETBP1*, Cristóbal *et al.*, 2010). Alternatively, the translocation process can result in the formation of a novel fusion gene with oncogenic properties (e.g. inv(16)(p13;q22) *CBFB/MYH11*, Kundu and Liu, 2001). The remaining 45% of cases appear to be cytogenetically normal, however more sensitive diagnostic techniques are able to identify smaller changes in these patients, such as focal deletions, point mutations or aberrant CpG methylation leading to altered gene expression.

The majority of genetic mutations in AML have previously been broadly classed into one of two categories (Renneville *et al.*, 2008). Class I mutations activate signal transduction pathways and enable either an increased rate of proliferation or enhanced survival capabilities of affected cells. These mutations often enhance the activation of either the FLT3 or RAS signalling pathways. Class II mutations commonly affect pathways involved in the regulation of transcription, either transcription factors or members of the transcriptional co-activation complex. These mutations are able to block differentiation and aberrantly maintain cells in a pre-leukaemic state. They may also activate self-renewal pathways that are generally only active in HSCs (Kelly and Gilliland, 2002).

Recent studies have suggested a further two classes of mutation to be present in AML. Class III mutations affecting *TP53* and *WT1* have been suggested to stand apart from Class I and Class II mutations, as they act as classical tumour suppressor genes, involved in the regulation of cell cycle and/or apoptosis (Dombret *et al.*, 2011). A fourth class (Class IV) comprising of epigenetic regulator genes has also been identified. Despite being rarer than Class I and Class II mutations, some evidence suggests that Class IV mutations are associated with a worse patient outcome and are more frequently observed in older AML patients (Fathi and Abdel-Wahab, 2012). It is possible that this may explain the poorer effects of treatment seen in older patients, including those with an otherwise favourable prognosis.

Class	**I**	**II**	**III**	**IV**
Pathway	**Signal Transduction**	**Differentiation**	**Tumour Suppression**	**Epigenetic Regulation**
Genes	*FLT3*	*RUNX1*	*TP53*	*TET2*
	NRAS, KRAS	*CBFβ*	*WT1*	*IDH1, IDH2*
	KIT	*CEBPA*		*DNMT3A*
	JAK2	*NPM1*		*ASXL1*
	PTPN11	*PU1*		*EZH2*
		MLL		
		RARα		
		EVI1		

Table 2. Summary of common genetic mutations observed in AML.

Figure 2 shows some of the key signalling pathways that are dysregulated in AML, and the downstream effects that these have, as well as the large amount of cross-talk that exists between pathways. Some of these pathways are discussed in more detail in later sections.

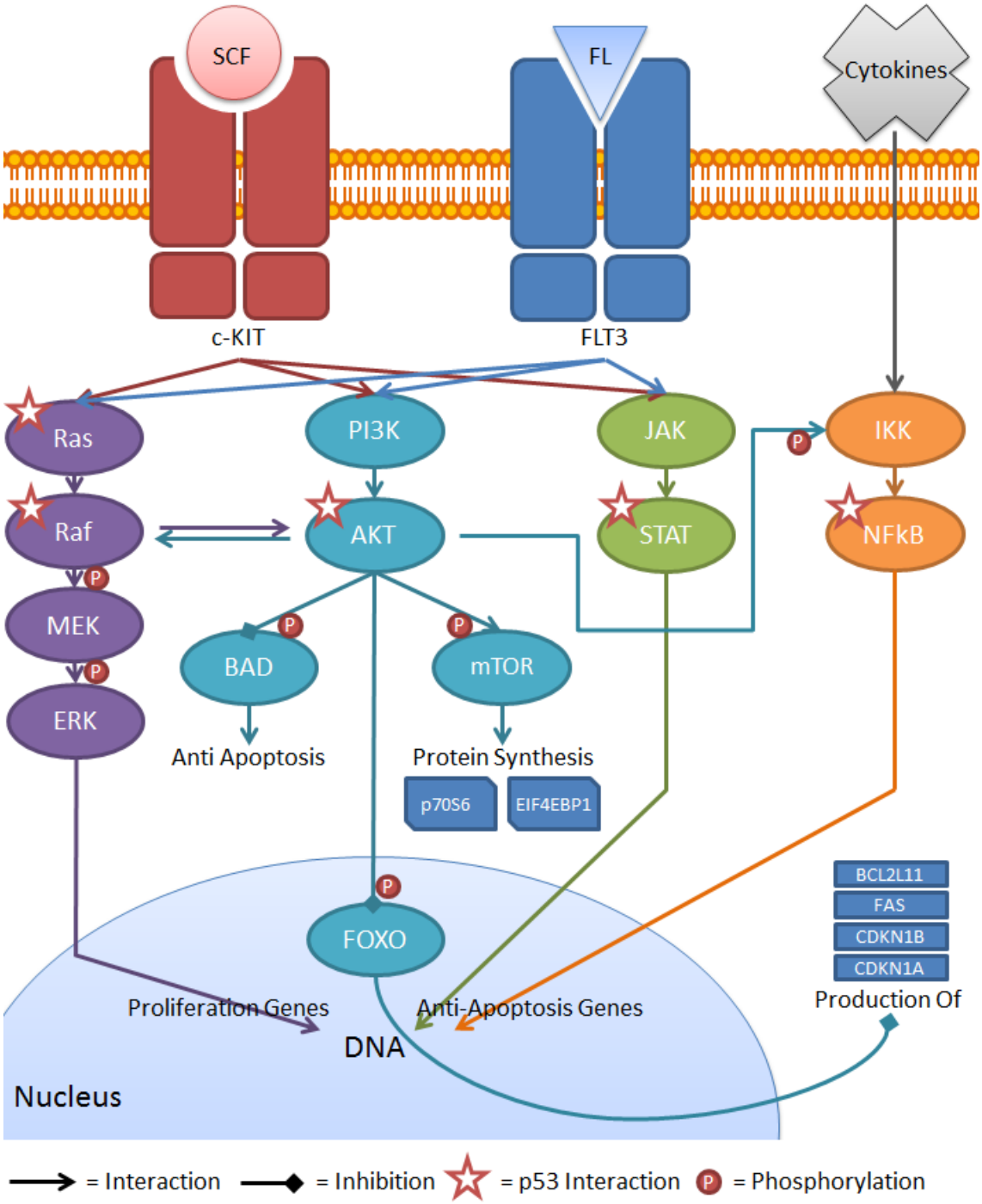


Figure 2. Key signalling pathways involved in AML development.

1.3.4 Disease Classification

The first widespread system used for the classification of AML was the French-American-British (FAB) classification, established in 1976. This system classified different AMLs based upon cellular morphology and cytochemical responses (Bennett *et al.*, 1976). Certain genetic changes are associated with particular morphological changes (Table 3), although with the exception of the M3 subtype, these are not absolute. The FAB subtype M3 is defined by t(15;17)(q24;q21) and the *PML/RARα* fusion (Rowley *et al.*, 1977).

Type	Cell Type	Common Cytogenetics
M0	Myeloblastic	
M1	Myeloblastic, with minimal maturation	
M2	Myeloblastic, with maturation	t(8;21)(q22;q22) *RUNX1/RUNX1T1*, t(6;9)(p22;q34) *DEK/NUP214*
M3	Promyelocytic	t(15;17)(q24;q21) *PML/RARα*
M4	Myelomonocytic	inv(16)(p13;q22) *CBFB/MYH11*, del(16q)
M5	Monocytic	del(11q), t(9;11)(q34;q23) *FNBP1/MLL*
M6	Erythroleukaemic	
M7	Megakaryocytic	t(1;22)(p13;q13) *RBM15/MKL1*
M8	Basophilic	

Table 3. The French-American-British classification of AML. Commonly associated cytogenetics are listed alongside associated disease classifications (Bennett et al., 1976).

Although there are some examples in the FAB system where cell morphology correlates with cytogenetics, there are many in which the cytogenetic change is seen over a range of FAB types. Thus, the FAB classification based on morphology alone does not fully reflect the genetic and clinical diversity of AML. The World Health Organization (WHO) system of classification was established to include strong associated genetic changes. The classification of AMLs under this system is outlined in Table 4 (Vardiman *et al.*, 2002; Vardiman *et al.*, 2009).

Acute myeloid leukaemia with recurrent genetic abnormalities	AML with t(8;21)(q22;q22) *RUNX1/RUNX1T1*
	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22) *CBFB/MYH11*
	APL with t(15;17)(q22;q12) *PML/RARα*
	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*
	Provisional entity: AML with mutated *NPM1*
	Provisional entity: AML with mutated *CEBPA*
Acute myeloid leukaemia with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
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
	Acute megakaryoblastic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis
	Myeloid leukaemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm	

Table 4. The World Health Organization classification of acute myeloid leukaemia (Vardiman et al., 2002; Vardiman et al., 2009).

A number of studies have utilised cytogenetic data to classify AML patients into different prognostic groups (Grimwade *et al.*, 1998; Slovak *et al.*, 2000; Grimwade *et al.*, 2001; Byrd *et al.*, 2002; Grimwade *et al.*, 2010). These studies mostly overlap with their findings, and a general consensus of prognostication is listed in Table 5. Whilst cytogenetic risk group is the

most important factor in risk stratification, more recent studies have devised stratification methods that take into account clinical features for those patients not placed into the good prognosis group by the presence of t(8;21)(q22;q22) *RUNX1/RUNX1T1,* inv(16)(p13.1q22) *CBFB/MYH11* or t(15;17)(q22;q12) *PML/RARα* (Grimwade *et al.*, 2010).

Prognosis	MRC	SWOG	CALGB
Good	t(8;21)(q22;q22) *RUNX1/RUNX1T1*	t(8;21)(q22;q22) *RUNX1/RUNX1T1* lacking del(9q) or complex karyotype	t(8;21)(q22;q22) *RUNX1/RUNX1T1*
	inv(16)(p13.1q22) /t(16;16)(p13;q22) *CBFB/MYH11* /del(16q)	inv(16)(p13.1q22) /t(16;16)(p13;q22) *CBFB/MYH11* /del(16q)	inv(16)(p13.1q22) /t(16;16)(p13;q22) *CBFB/MYH11* /del(16q)
	t(15;17)(q22;q12) *PML/RARα*	t(15;17)(q22;q12) *PML/RARα*	
	CEBPA double mutations		
	NPM mutations		
Intermediate	No abnormality	No abnormality	No abnormality
	Sole +8	+8	
	11q23 abnormality	+6	
	Other intermediate (not otherwise classified as favourable or adverse)	-Y	-Y
		del(12p)	12p abnormality
			del(7q)
			t(9;11)(p22;q23) *MLLT3/MLL*
			+11
			del(11q)
			+13
			del(20q)
			+21
Poor	-7	-7/del(7q)	-7
	del(5q)/-5	del(5q)/-5	
	3q abnormality	3q abnormality	inv(3)(q23q26) *RPN1/MECOM* or t(3;3)
	FLT3-ITD	9q	

		11q	
		20q	
		21q	
		17p	
		t(6;9)	t(6;9)
		t(9;22)	t(6;11)
			Sole +8/+8 with one other abnormality
			t(11;19)
	Complex karyotype (without favourable)	Complex karyotype	Complex Karyotype

Table 5. Prognostic implications of AML cytogenetic subgroups. MRC = Medical Research Council (Grimwade et al., 1998; Grimwade et al., 2001; Grimwade et al., 2010), SWOG = Southwest Oncology Group (Slovak et al., 2000), CALGB = Cancer and Leukemia Group B (Byrd et al., 2002).

As paediatric AML is rarer than adult AML, there are fewer studies looking at the effects on prognosis of cytogenetics specifically in children. Although the abnormalities seen are largely the same, the distribution of cytogenetic changes in paediatric AML is different from that of adult AML (Raimondi *et al.*, 1999; Betts *et al.*, 2007). Despite the lower number of cases, a large recent study of 729 paediatric AML patients in the UK established a similar scheme of disease classification for children, outlined in Table 6 (Harrison *et al.*, 2010) which was verified by a Dutch group (Balgobind *et al.*, 2011).

Prognosis	**Cytogenetics**
Good	t(8;21)
	inv(16)
Intermediate	Not included in Good or Poor risk
Poor	Abnormal 12
	t(6;9)
	5q Abnormalities
	-7
	t(9;22)

Table 6. Medical Research Council classification of prognostic risk in paediatric AML (Harrison et al., 2010).

1.3.5 Pathophysiology

The most common cause of death in AML is bone marrow failure. As AML blasts carry a genetic block to differentiation, these cells are unable to develop into mature red cells, neutrophils, monocytes and platelets. Cytokine expression by AML blasts prevents healthy BM cells from differentiating into their mature progeny, further compounding this effect (Youn *et al.*, 2000). This BM failure often leads to death due to infection (Anderlini *et al.*, 1996; Stalfelt *et al.*, 2001). Particularly high white blood cell counts (>50,000 cells per µL) often lead to potentially fatal organ infiltration, with an especially poor prognosis if the lung or brain is affected (Seymour *et al.*, 1994).

1.3.6 Treatment Strategies

The treatment of AML has remained essentially the same for almost 40 years. The first phase of therapy (induction) consists of 3 days of treatment with the anthracycline daunorubicin in combination with cytarabine for 7 days. This treatment produces complete remission rates of around 65-75% (Tallman *et al.*, 2005). Other treatments tested include the anthracycline idarubicin, the anthracenedione mitoxantrone, cytotoxic agents such as etoposide and fludarabine, as well as growth factors such as granulocyte colony-stimulating factor. However, these treatments have not been proven to be more successful than standard therapy except in a small number of subgroups (Estey and Dohner, 2006).

If remission is achieved with initial therapy, postremission therapy (consolidation) is administered. This treatment is directed according to the risk group that the patient has been classified into – either good, intermediate or poor risk, depending on the cytogenetic group, age and white blood cell count. Patients stratified into the good risk arm tend to be continue on therapy consisting of high dose cytarabine, which has been shown to be highly effective (Schaich *et al.*, 2013). Intermediate and particularly poor risk patients are often recommended for either autologous or allogeneic stem cell transplantation as an alternative to chemotherapy (Burnett *et al.*, 2009). Stem cell transplantation has improved outcome in terms of relapse-free survival, but carries a risk of treatment-related deaths of between 10 and 25%, as well as conferring an inferior long-term quality of life (Watson *et al.*, 2004). Patients classified as poor risk have less than 10% relapse-free survival on all treatments.

Patients who either fail to enter complete remission after induction therapy, or who relapse early are placed onto salvage treatment. The treatment given in these cases depends on the previous treatment received but will typically consist of either high-dose cytarabine or fludarabine plus cytarabine. In some cases allogeneic stem-cell transplantation will be recommended, which has proved to be successful in certain subgroups (Wong *et al.*, 2005).

The current definition of complete remission in AML is the presence of less than 5% myeloblasts in the BM. A number of studies have suggested that this number is too high, and there is increasing evidence that more sensitive techniques such as polymerase chain reaction (PCR) and FACS should be used to detect minimal residual disease (MRD) (Chen *et al.*, 2011; Kronke *et al.*, 2011; Roboz, 2011). It is believed that the high rate of relapse observed in AML is due to a small population of leukaemic stem cells (LSCs) that normally make up between 0.1% and 1% of the entire leukaemic population. These cells are mostly quiescent but are capable of self-renewal, and are highly resistant to both intensive chemotherapy and allogeneic stem cell transplants (Roboz and Guzman, 2009). For this reason, a number of novel therapies are currently in development with the aim of specifically targeting MRD. Such therapies include drugs that target the PI3K pathway or the transcription factor complex NF-κB, both of which have been shown to be crucial for LSC function (Frelin *et al.,* 2005, Thomas *et al*., 2013). Other possible treatments include antibody based therapies targeting cell surface antigens preferentially expressed by LSCs, such as CD33, CD44, CD47, CD123 and CLL-1 (Walter *et al.*, 2012). Finally, therapies that may disrupt regulatory pathways such as Wnt, Notch and HOX are being developed to disrupt the interaction of LSCs with the BM microenvironment (Roboz and Guzman, 2009; Schneider *et al.*, 2014).

There are currently a number of therapies in development to target commonly mutated genes in AML. For example, there are 5 drugs targeting *FLT3* currently in trials (Lestaurtinib, Quizartinib, Sorafenib, Midostaurin and PLX3397, Hatzimichael *et al.*, 2013). *FLT3* is an important target in AML since mutations in the gene are frequent and *FLT3*-ITD imparts a particularly poor prognosis (Grimwade *et al.*, 2010; Fathi and Chen, 2011). With therapies also in development for other genes involved in AML, including *NPM1*, *PLK1*, *DOT1L* and *JAK2,* there is hope that patient outcome can be improved in the future with new treatments.

1.4 Tumour Suppressor Genes

Tumour suppressor genes and oncogenes are crucial factors that contribute to cancer development in a wide range of malignancies, and are the main focus of this project. Historical examples of each will be discussed and select examples involved in AML development will be explored in these sections.

The classical "two-hit hypothesis" of cancer dates back to 1971, when it was first proposed by Alfred Knudson. In his hypothesis, he stated that loss of function mutations in tumour suppressor genes are recessive, so for a cell to become malignant, it would need to lose function of both alleles (Knudson, 1971). He reached this conclusion through the study of children who developed retinoblastoma, and found that 40% of cases were associated with a familial history of the disease. Additionally, these cases had a much higher rate of tumours that affected both eyes and occurred at a younger age, whilst nonhereditary cases were unilateral and typically occurred in older children. From this, Knudson concluded that retinoblastoma was caused by two separate mutation events, each affecting a single copy of a tumour suppressor gene (now known as the *RB1* gene). For patients with hereditary disease, he hypothesised that they already carried one of these mutations, leading to frequent bilateral tumours and early onset of disease. From his work, he estimated that each mutation event would occur at a rate of 2×10^{-7} per year, and whilst many people in the general population would have acquired a single *RB1* mutation in their lifetime, the cells carrying the mutations would have already differentiated and stopped dividing before they acquired a second mutation (Knudson, 2001).

1.4.1 TP53

The TP53 protein was originally identified as a component of the SV40 virus – a virus shown to transform cells and induce tumours in animals. It was known that the virus expressed 3 early antigens associated with cell transformation: a large antigen of 90kDa, a small antigen of 17kDa, and a middle T antigen of around 55kDa. The identification of the middle antigen became an important topic. It was eventually found to be a 53kDa protein that was not an antigen, and was not coded by the virus, but by the host cell itself (Kress *et al.*, 1979). This discovery sparked the field of *TP53* research into life, with subsequent publications showing that TP53 accumulated in the nuclei of tumour cells (Linzer and Levine, 1979), and was also associated with the T antigen of SV40 (Lane and Crawford, 1979).

In the early 1980s, the theory that cancer arose as the result of oncogene activation became a popular topic of research. At the same time, it was shown that transfected *TP53* could cooperate with other known oncogenes to lead to cellular transformation (Parada *et al.*, 1984),

leading to the classification of *TP53* as an oncogene. However, further studies indicated that the gene did not quite fit into the oncogene classification. It was reported that *TP53* alleles were inactivated in murine erythroleukaemias (Mowat *et al.*, 1985), and that it was frequently deleted or rearranged in human and murine osteosarcoma (Masuda *et al.*, 1987). It later became clear that early clones of mammalian *TP53* genes contained base pair substitutions in regions coding functional domains (Finlay *et al.*, 1989), which produced mutant TP53 protein with seemingly oncogenic properties, when it actually exhibited dominant negative features. The subsequent demonstration of the anti-proliferative characteristics of wild type *TP53*, in conjunction with its inactivation in human cancers, lead to its status finally being changed from oncogene to tumour suppressor.

The *TP53* gene is located on the short arm of chromosome 17. The encoded protein contains 393 amino acids, divided into 7 functional domains. These functional domains are summarised in Figure 3. The first activation domain (TAD1) is required for transcription factor activation, whilst the second activation domain (TAD2) is important for apoptotic activity. The central DNA binding domain (DBD) is crucial to many of the roles of *TP53* and is the hotspot for mutations in a wide range of cancers (Olivier *et al.*, 2010). The homo-oligomerisation domain (OD) allows TP53 to form tetramers with itself, essential for its functional activity.

Figure 3. A schematic representation of the p53 protein. TAD - Transactivation Domain; PRR - Proline Rich Domain; DBD - Central DNA-binding Core Domain; NLS - Nuclear Localisation Signalling Domain; OD - Homo-Oligomerisation Domain; CTD - C-Terminal Domain.

TP53 primarily functions as a transcription factor, by enhancing the transcription of a number of genes that carry out the TP53-dependent functions in a cell. It has a number of additional roles, including the activation of DNA repair proteins, the arrest of the cell cycle at G1/S phase, and the induction of apoptosis. Two homologues of *TP53* have been described: *TP63* and *TP73* (Kaghad *et al.*, 1997; Yang *et al.*, 1998). Both are believed to have similar functions to *TP53*, while the 1p36.32 region (where *TP73* is located) is frequently deleted in cancers. However deletions of *TP63* and *TP73* alone in murine models are insufficient to induce cancers, unlike deletions of *TP53*.

A large study investigated the prognostic value of *TP53* deletions in 2272 adult AMLs. Deletions were found in 105 patients (5%), and the majority of them occurred in conjunction with

multiple other aberrations (Seifert *et al.*, 2009). They showed that *TP53* deletions carried an inferior prognosis in both simple and complex karyotypes compared to patients without these deletions. Studies investigating the mutation rate in AML have found an occurrence of around 13% across all AML subtypes (Schoch *et al.*, 2006).

TP53 aberrations frequently occur in conjunction with other high-risk abnormalities, such as del(5q), -5 and -7, as well as complex karyotypes (Christiansen *et al.*, 2001; Rucker *et al.*, 2012). In many of these patients, deletions often occur in conjunction with mutation of the remaining allele, leading to a complete loss of wild type function TP53. A study of 234 complex karyotype AML patients showed that whilst 40% of patients had deletion of *TP53*, 60% of patients had a mutated *TP53*, and two thirds of those with a mutant *TP53* had deletion of the remaining allele, resulting in homozygous loss of wild type functional TP53 (Rucker *et al.*, 2012).

Some theories have implied that loss of *TP53* in AML facilitates the development of further abnormalities due to the genomic instability associated with *TP53* loss, leading not only to complex karyotypes but also to even rarer events such as gene amplification. It has also been related to the resistance to chemotherapy seen in patients with a complex karyotype, as the unstable genomic landscape is susceptible to changes that may be advantageous to their continued growth when faced with overcoming the effects of treatment (Schmitt *et al.*, 2002; Haferlach *et al.*, 2008).

1.4.2 WT1

The *WT1* gene was one of the first tumour suppressor genes to be cloned, in 1990 (Call *et al.*, 1990; Gessler *et al.*, 1990). When mutated, this gene led to predisposition to the development of a type of embryonal kidney malignancy, known as Wilms' tumour or nephroblastoma. The gene contains multiple zinc finger motifs, as well as a DNA-binding domain. *WT1* also contains multiple alternative splice sites, and 24 different isoforms of the gene have been identified (Hohenstein and Hastie, 2006). The exact mechanisms by which *WT1* functions have not been fully elucidated, because the gene itself has different functions depending on which tissue it is expressed in as well as at what stage of development it is expressed.

WT1 has been identified to have 3 main roles. The first is in transcriptional regulation. Through differential expression of *Wt1* in mutant mouse models, in combination with chromatin immunoprecipitation techniques, a number of target genes have been identified, which appear to be activated by *WT1,* including *Amphiregulin, Sprouty, TrkB, nephrin, nestin* and *Pou4f2*. *WT1* is also believed to have a role in RNA metabolism, as it has been shown to bind RNA *in vivo* (Ladomery *et al.*, 2003). However it is not clear whether this leads to alternative splicing or processing of the RNA that it binds to. The second role of *WT1* is believed to be in

differentiation, as *WT1* has been shown to be required for proper kidney development by stimulating differentiation of immature kidney cells (Davies *et al.*, 2004). In other organ types such as the heart, *WT1* has the opposite role; to maintain an undifferentiated state in progenitor cells (Moore *et al.*, 1998).

Although first identified as a tumour suppressor, recent evidence has accumulated suggesting a possible oncogenic role for *WT1* in tumours such as colorectal, breast, desmoid and brain. In these tumour types elevated *WT1* expression has been identified, on a wild type *WT1* background. Additionally, RNAi targeted to *WT1* in some cell lines has led to a reduction in proliferation rate, as well as the inhibition of apoptosis (Tuna *et al.*, 2005). In tumours where elevated *WT1* expression has been identified, the majority of the protein appeared to be located in the cytoplasm; an abnormal location for the majority of WT1 proteins which may reflect the upregulation of a specific *WT1* isoform. The cell of origin of these tumour types is also likely to be highly important in unravelling the role of *WT1*, especially as it is clear that *WT1* has a variety of roles in differentiation between tissue types, as well as in different stages of differentiation.

In AML, *WT1* mutations have been found to occur in 6-13% of all non-M3 AML patients, with a higher rate in younger patients with a normal karyotype (Gaidzik *et al.*, 2009; Hou *et al.*, 2010). Multivariate analysis of these data showed that *WT1* mutations were an independent poor prognostic factor for both overall survival and relapse-free survival in these patients. High levels of *WT1* expression have been associated with a worse outcome (Nomdedeu *et al.*, 2013; Luo *et al.*, 2014) and it has been suggested that levels of *WT1* could be used as a target for MRD detection in AML with a high degree of accuracy (Nomdedeu *et al.*, 2013).

1.4.3 CUX1

CUX1 was recently identified as a tumour suppressor gene in AML within the long-investigated long arm of chromosome 7 (7q). This region is frequently deleted in AML, but it was previously unknown how these deletions contributed to leukaemogenesis. *CUX1* was identified to be within a minimal region of deletion of 7q, and to have a reduced expression in AML cases where it was deleted (McNerney *et al.*, 2012). Additionally, one case with a *CUX1* fusion gene was identified in this study, providing an alternative process by which its expression was reduced.

CUX1 functions as a transcription factor involved in cell cycle progression and apoptosis, and has previously been implicated in the tumorigenesis of breast and pancreatic cancers (Hulea and Nepveu, 2012). However, these studies in solid tumours have found that its expression is increased in high grade tumours, and its high expression is associated with increased cell

proliferation as well as motility and invasiveness. Despite this seemingly conflicting evidence, McNerney *et al.* (2012) used RNAi to knock down the *CUX1* ortholog, *cut* in *D. melanogaster*. They found that this resulted in the formation of melanotic tumours in *Drosophila* larvae and pupae as the result of haemocyte overgrowth. These findings were consistent with a human model, where *CUX1* was knocked down in human haematopoietic progenitors and transplanted into immunocompromised mice. The proportions of cells expressing *CUX1* shRNA were found to be significantly higher compared to cells expressing nonspecific shRNAs, but this change did not affect the ability of the cells to differentiate into various haematopoietic lineages (McNerney *et al.*, 2012). These data suggest that *CUX1* may act as a tumour suppressor gene in other haematopoietic cell types.

Interestingly, unlike classical tumour suppressor genes such as *RB1*, *CUX1* was found to act as a haploinsufficient tumour suppressor gene, as the transcript levels were found to be half of those found in samples with both *CUX1* alleles intact. The large size of deletions within chromosome 7q in AML raise the possibility that there are other genes acting as tumour suppressors in this region, either through a heterozygotic or biallelic mutation. Indeed, this is likely to be the case as Chen *et al.* (2014) demonstrated that *MLL3* is likely to have tumour suppressor properties in AML, as *Mll3* suppression through shRNA knockdown and CRISPR gene editing in mice promoted the development of myeloid leukaemia (Chen *et al.*, 2014). *Mll3* suppression was found to impair haematopoietic stem cell differentiation both *in vitro* and *in vivo*, producing a MDS-like state and corresponding gene expression signature.

Further compounding the complex nature of deletions in this region, two other genes have been put forward as possible tumour suppressors. *EZH2* has been shown to act as a tumour suppressor in ALL (Simon *et al.*, 2012) yet also acts as an oncogene in other cancer types (Lund *et al.*, 2014). *MLL5,* another *MLL* family member has also been shown to be involved with differentiation and self-renewal in haematopoietic stem cells (Zhang *et al.*, 2009) suggesting that the 7q region may harbour multiple tumour suppressor genes.

1.5 Oncogenes

Whilst oncogenes play an equally large part in cancer development, their role is the opposite of a tumour suppressor gene. Rather than loss of function leading to tumourigenesis, a proto-oncogene is normally affected by activating mutations that lead to their overactivity and subsequent classification as an oncogene. This may be a point mutation that leads to a gene product being constitutively activated, or mutations of genetic material resulting in overexpression of the gene.

The first oncogene to be identified was *src* (Martin, 2001). It was identified after considerable research into the Rous sarcoma virus (RSV), which was first identified as a filterable agent that caused solid tumours in chickens. At the time, the research community was highly sceptical that cancers could be caused from an infectious origin, however further research demonstrated that the virus was released from infected tumour cells and led to the further propagation of both the tumour itself and viral particle release (Rubin, 1955). Subsequent work identified that different cellular morphology could be produced as a result of infection depending on which strain infected the cells, suggesting that the morphology of the transformed cell was controlled by a genetic property of the virus (Temin and Rubin, 1958; Temin, 1960). Although it was not known at the time, this was actually a property of the *v-src* gene.

Additional strains of RSV were identified in 1970 that were unable to induce cellular transformation and were key to the physical identification of *v-src* (Golde 1970; Toyoshima *et al.*, 1970). The strains that were unable to induce transformation were found to have a shorter viral genome, suggesting that these genes were lacking in a component that promoted transformation. This gene was first identified with oligonucleotide fingerprinting to allow it to be tracked through a series of genetic crosses, from which it was shown to be the gene responsible for transformation (Bernstein *et al.*, 1976; Wang *et al.*, 1976).

The question still remained why RSV carried a gene that induced cellular transformation. Further research using sequence probes for *v-src* found that the gene had homologous sequences in normal avian genomes, and that regions of these genes were also conserved in vertebrates. The fact that the gene was present in vertebrates suggested that it was of cellular and not viral origin. This led to *src* being designated as the first proto-oncogene, due to its nature as a cellular precursor of a retroviral transforming gene (Stehelin *et al.*, 1976).

1.5.1 MYC

Mutation, amplification or activation of *MYC* family genes is one of the most frequent events associated with cancer (Eilers and Eisenman, 2008). The discovery of the first *MYC* family genes

stemmed directly from the work researching RSV and the findings related to *src*. Another avian virus affecting chickens, MC29 was isolated and found to induce haematopoietic neoplasia, although it did not result in the development of overt leukaemia (Ivanov *et al.*, 1964). Instead, the virus transformed myeloid cells which went on to form either diffuse growths known as myelocytomatosis, or solid tumours known as myelocytomas. These growths were to give the *MYC* family its name.

In a similar story to s*rc,* finding the gene driving cellular transformation was difficult but the use of newly developed techniques (in this case, radiolabelled nucleic acid sequences) allowed researchers to identify and track the single functional gene it carried, the *gag-myc* fusion gene (Sheiness *et al.*, 1978). The *v-myc* gene within this fusion gene was identified and found to be distinct from *src*. Additional work on other avian leucosis viruses that did not carry *v-myc* found that they acted by directly stimulating a *v-myc* homologue through retrovirial insertion (Hayward *et al.*, 1981). This allowed the cellular homologue of *v-myc* to be identified, which was named *c-myc* and began the development of a huge amount of research into the *Myc* family of genes (Vennstrom *et al.*, 1982).

Mammalian cells express MYC proteins from 3 distinct genes, which are *c-Myc*, *N-myc* and *L-myc*. They have similar functions but different patterns of expression as well as differing potencies (Tansey, 2014). They have a similar general structure, with a DNA-binding domain at their C-termini and a transactivation domain at their N-termini. This means that upon binding specific genomic sites, *MYC* proteins are able to recruit RNA polymerase II associated proteins to stimulate gene transcription (Tansey, 2014).

The first human involvement of *MYC* driven cancer was found in Burkitt's lymphoma. Large chromosomal translocations had been noted in a range of human malignancies, but their biological effects were unknown. After *c-MYC* was mapped to chromosome 8, it was noted that in Burkitt's lymphoma, chromosomes 14, 2 or 22 were translocated with chromosome 8. The partners for *c-MYC* in these cases were found to be the loci for immunoglobulin heavy and light chains respectively, and the relocation of *c-MYC* to their highly active enhancer regions was found to result in overexpression of *c-MYC* (Adams *et al.*, 1985).

MYC family members typically drive cancer development through a number of mechanisms, however it is clear now that not all of these mechanisms are active in every cancer type where *MYC* is dysregulated. *MYC* has been shown to drive cellular growth through the activation of cyclin expression and through inhibiting cell cycle checkpoints, leading to cells being able to pass through G1 and G2 cell cycle phases faster (Oster *et al.*, 2002). *MYC* also drives growth through promoting protein synthesis and RNA transcription (Iritani and Eisenman, 1999). It

achieves this by activating the expression of protein coding genes responsible for ribosome biogenesis and protein translation. *MYC* is also able to stimulate the activity of RNA polymerases I and III leading to enhanced rRNA and tRNA synthesis (Gomez-Roman *et al.*, 2003; Arabi *et al.*, 2005). Finally, *MYC* is also able to converge with mTOR to increase the efficiency of mRNA-ribosome engagement (Pourdehnad *et al.*, 2013).

The increased levels of proliferation driven by *MYC* family members has been shown to lead to an increased level of cancer genome instability. This phenomenon is known as replication stress, and leads to the breakdown of DNA replication forks at fragile hotspots in the genome. As the cell cycle checkpoints in cancer cells are typically inhibited, these breaks are not resolved correctly and can lead to double-stranded DNA breaks, loss of heterozygosity, amplifications and rearrangements of select loci (Mai *et al.*, 1996; Felsher and Bishop, 1999). *MYC* overexpression has also been linked to an increased production of reactive oxygen species (ROS), which themselves contribute to further DNA damage (Vafa *et al.*, 2002).

Somewhat counterintuitively for an oncogene, *MYC* is also able to drive apoptosis. In the absence of cellular survival factors, forcing high levels of *MYC* expression has been shown to promote apoptosis (Askew *et al.*, 1991), and this is achieved mainly through stimulating expression of the *ARF* tumour suppressor gene, resulting in a rapid induction of *TP53* (Li and Hann, 2009). It is clear that this mechanism needs to be overcome in *MYC*-driven cancer development, and one mechanism through which this is achieved is by the commonly observed loss of *TP53* in cancer, as described previously.

In AML, *MYC* is upregulated by activating mutations of the *FLT3* receptor tyrosine kinase (Gilliland and Griffin, 2002), as well as by a number of fusion genes commonly found in AML, including *RUNX1-RUNX1T1*, *PML-RARα* and *PLZF-RARα* (Müller-Tidow *et al.*, 2004). Mutations in CEBPα have also been linked with elevated expression levels of *MYC* (Johansen *et al.*, 2001). Amplification of the genomic region containing *MYC* is one of the most common amplifications detected in AML, and these additional copies are often found on double minute chromosomes (Slovak *et al.*, 1994). Although microarray studies do not often report *MYC* overexpression, this may be due to stricter filtering protocols searching for fold changes greater than 2 times above control samples. However it is known that *MYC* can have relevant effects at a sub 2-fold expression level change, as demonstrated by the ability of *c-Myc* to cause transformation of rat fibroblasts and mouse embryonic stem cells with only modest overexpression (Murphy *et al.*, 2008).

1.5.2 RAS

The *Ras* family of genes contains 3 genes (*HRAS*, *KRAS*, and *NRAS*) that encode small GTPases critical to many signal transduction pathways within cells. Similar to *Src* and *Myc*¸ the first *Ras* genes were identified as homologues of the transforming components of Harvey and Kirsten murine sarcoma retroviruses, in 1981 (DeFeo *et al.*, 1981). The Harvey sarcoma virus-associated oncogene was named Ha-*ras* (*H-RAS* in mammals) and the Kirsten sarcoma virus transforming gene was named Ki-*ras* (*K-RAS* in mammals). After the identification of the human homologues of these genes, mutations in their sequences were discovered soon afterwards in a range of human cancer cell lines (Karnoub and Weinberg, 2008). The third member of the *Ras* family was identified in 1983 through cloning from neuroblastoma and leukaemia cell lines, and was named *N-RAS* (Shimizu *et al.*, 1983).

The GTPase activity of *Ras* family members was identified in 1980, when it was shown that *Ras* proteins bound to guanine nucleotides in a similar fashion to other previously identified GTPases (Shih *et al.*, 1980). In mutated *Ras*, this function is disrupted and the commonly observed mutations in cancer lead to impaired GTP hydrolysis. This leads to *Ras* remaining in an active conformation for an increased period of time, during which it stimulates its downstream pathways.

Downstream of Ras, the Ras/Raf/MEK/ERK pathway is central to the transduction of a variety of cell surface receptors to transcription factors in the nucleus, including mitogens and cytokines. The Raf protein family contains serine/threonine specific protein kinases related to retroviral oncogenes. There are 3 members, A-Raf, B-Raf and C-Raf (Raf-1), which are primarily activated by Ras family members. However they are also able to be regulated by AKT (Rommel *et al.*, 1999), cAMP-dependent protein kinase (PKA) and protein kinase C (PKC) (Dumaz *et al.*, 2002). The *Raf* family members contain a Ras-GTP binding domain and a catalytic protein kinase domain, linked by a serine rich hinge region.

The primary target of Raf proteins are the *MEK* family of gene products (MEK1-5), which are tyrosine/threonine kinases. The *MEK* gene products (primarily MEK1 and MEK2) then phosphorylate and activate mitogen activated protein kinases (MAPK). These have many targets of phosphorylation, including a wide range of proliferation related transcription factors such as c-Myc and CREB (through MNK phosphorylation) (Avruch *et al.*, 2001). Furthermore, translation of mRNA to protein is increased by phosphorylation of the 40S ribosomal protein S6 kinase, leading to phosphorylation of ribosomal protein S6 and enhanced translation.

In AML, mutations in *N-RAS* or *K-RAS* are found in 12-27% of patients, with around two thirds of these occurring in *N-RAS*, and one third in *K-RAS* (Bowen *et al.*, 2005; Neubauer *et al.*, 2008).

There is much conflicting evidence about the prognostic effects of *RAS* mutations in AML, with some studies finding inferior, the same or a more favourable prognosis associated with *RAS* mutations. Despite these observations, there is evidence that patients with *RAS* mutations may benefit from increased doses of cytarabine during consolidation treatment (Neubauer *et al.*, 2008; Ahmad *et al.*, 2011).

1.5.3 BCL2

BCL2 was one of the first oncogenes to be discovered that primarily drove oncogenesis through increased cellular survival rather than by promoting aberrant proliferation. It was first identified in cancer cells as being translocated in a t(14;18)(q32;q21) translocation to the immunoglobulin heavy chain locus in follicular lymphoma (Tsujimoto *et al.*, 1984), leading to *BCL2* overexpression. As *BCL2* normally has an anti-apoptotic function, this overexpression prohibits cell death through inhibition of pro-apoptotic *BCL*-family members. There have been a large number of *BCL2* family members identified, including inhibitors of apoptosis (for example *BCL2* and *BCLX*), promoters of apoptosis (such as *BAX and BAK*), as well as regulators that act as either activators or inhibitors (for example *BAD*, *BID* and *BIM*) (Youle *et al.*, 2008; Chipuk *et al.*, 2010).

BCL2 expression levels in AML have been associated with a number of AML disease characteristics. It is highly expressed in the immature FAB subtypes M0 and M1 (Venditti *et al.*, 2004), and its increased expression is associated with expression of the *RUNX1-RUNX1T1* fusion gene (Klampfer *et al.*, 1996). High levels of BCL2 protein have been proposed to be markers of poor prognosis (Karakas *et al.*, 1998) except in patients who have poor risk cytogenetics, for whom it is a marker of improved prognosis (Kornblau *et al.*, 1999).

Due to the widespread prevalence of altered *BCL2* expression in AML, a number of studies are looking to use it as a target for therapy with the aim of improving treatment. ABT-199, a *BCL2* inhibitor currently under investigation for the treatment of lymphoid cancers has been shown to be effective in inducing cell death in AML cell lines *in vitro* (Pan *et al.*, 2014). Similarly, another *BCL2* inhibitor known as ABT-737 has been shown to be effective in treating chemoresistant primary AML blasts (Kohl *et al.*, 2007), as well as in the treatment of AML in a transgenic mouse model (Beurlet *et al.*, 2013).

1.6 12p Deletions in AML

Both deletions and translocations of the short arm of chromosome 12 have been widely described in a variety of haematological malignancies, including AML, ALL, MDS and non-Hodgkin lymphoma (Andreasson *et al.*, 1997). Translocations in this region typically involve the Ets Variant 6 (*ETV6*) gene, and over 40 partner genes for *ETV6* have been identified. In AML cases with translocated *ETV6* however, it is rarely possible to detect the expression of a fusion protein and it is thought that the disruption of gene expression in this region leads to leukaemogenesis in these patients (Bohlander, 2005).

Deletions of 12p are found in between 1% and 5% of AML patients from conventional cytogenetic studies (Grimwade *et al.*, 2010; Harrison *et al.*, 2010) and in adult AML they are frequently associated with other abnormalities such as -7/-7q and -5/-5q (Wall *et al.*, 2012). It has been suggested that small deletions of 12p that are not detectable by cytogenetic analysis may occur in up to 10% of normal karyotype AML patients, as demonstrated by a sensitive FISH study performed by Andreasson *et al.* (1997). Deletions of 12p in paediatric AML have been found to have a poor prognosis, with a 35% event-free survival compared to 59% without deletions of 12p (Harrison *et al.*, 2010). Although there is limited data for adult AML with 12p deletions, analysis of survival data from annotated copy number microarray data demonstrates that a poor prognosis is also observed in these patients (Parkin *et al.*, 2010).

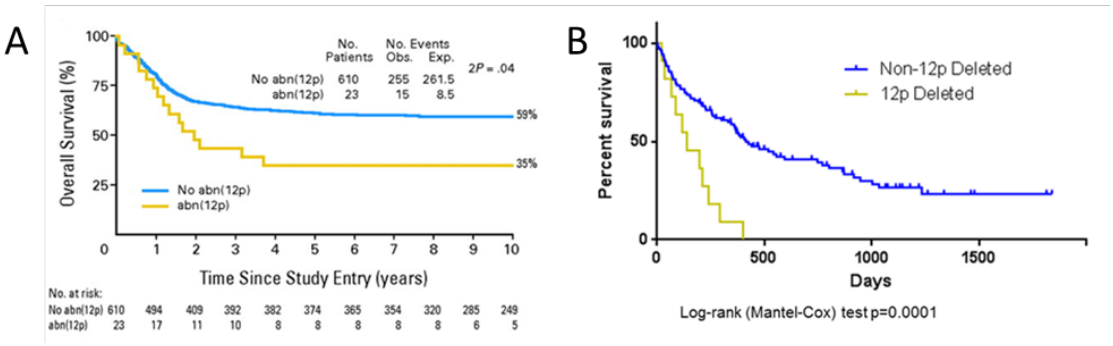


Figure 4. Prognostic effects of 12p deletion in AML. A. 12p deletions in paediatric AML, from Harrison et al. (2010). B. 12p deletions in adult AML, from Parkin et al. (2010) (GSE23452).

A number of studies have attempted to define the boundaries of 12p deletions in AML. Whilst some studies have suggested boundaries defined by the *ETV6* and *CDKN1B* genes, (Sato *et al.*, 1995; Wlodarska *et al.*, 1996; Andreasson *et al.*, 1997; Andreasson *et al.*, 1998; Baens *et al.*, 1999) two studies based on findings in single patients have disputed this. Silva *et al.* (2008) excluded *CDKN1B* from the MDR whilst retaining *ETV6*, whilst Haferlach *et al.* (2009) excluded *ETV6* whilst retaining *CDKN1B* in their MDR. Many of these studies used FISH probes to map the deletions relying on small numbers of patients, thus identifying the region of 12p deletions using a high resolution method and increased numbers of patients is warranted.

Few studies have measured the expression levels from within the deleted region of 12p and compared them to control samples; the studies that measured expression have generally focussed on single genes (Baens *et al.*, 1999; Silva *et al.*, 2008). In other haematological malignancies however, it has been demonstrated that loss of genes from this region results in reduced expression of affected genes, suggesting that the loss of material has functional consequences (Montpetit *et al.*, 2004).

1.7 11q Amplifications in AML

Gene amplification is an event rarely associated with AML and typically involves either the long arm of chromosome 8 (containing the *MYC* gene as described previously), or the long arm of chromosome 11 (11q). Amplifications of 11q are found in between 1 and 2% of AML patients overall. However, the gain of genetic material from 11q is the most commonly observed gain in complex karyotype AML, seen in between 15 and 40% of these patients (Schoch *et al.*, 2002; Rücker *et al.*, 2006). It is likely that if there is a gene or genes with an oncogenic dosage linked effect on 11q, it would reside within amplified regions that overlap with the commonly observed gains.

Amplified chromosomal material of 11q has been found in a variety of cytogenetic manifestations, including homogenously staining regions, double minutes, ring chromosomes and marker chromosomes. The number of additional copies found in amplifications has ranged from 4 to 50 copies per cell (Avet-Loiseau *et al.*, 1999). Typically, patients with amplifications of 11q have a complex karyotype, which are frequently associated with deletions of chromosome 5/5q, 7/7q and 17/17p (Zatkova *et al.*, 2009). AML patients with 11q amplification tend to be older (median age = 71, Michaux *et al.*, 2000) and respond poorly to treatment (Maitta *et al.*, 2009). Both complex karyotype AML and 11q amplification are associated with therapy-related AML, with 11q amplification found in 12% of t-AML patients (Andersen *et al.*, 2001; Mohamed, 2011).

Whilst there are a number of historical studies that have looked at 11q amplifications using Fluorescence *in situ* Hybridisation (FISH) (Michaux *et al.*, 2000; Streubel *et al.*, 2000; Andersen *et al.*, 2001; Dolan *et al.*, 2002), there are few that have made use of array based techniques to define the extent of gains and amplifications of 11q in AML. A study by Rücker *et al.* (2006) used microarray techniques to identify a wide range of chromosomal aberrations in 60 complex karyotype patients. Twenty-four of these patients (40%) had some form of chromosomal gain on 11q and three separate common regions of gain were identified. Fifteen cases had gains between 59.1 and 79.6 megabases (MB) on 11q12-q14. Twenty-one cases had gains between 116.3 and 119.6MB on 11q23.3 (containing the *MLL* locus), and of these, 5 cases showed high level amplification greater than 2 additional copies. The third common region of gain was identified in 20 cases as being between 127.1 and 129.3MB on 11q23.3-q24.1 (containing the *ETS1* and *FLI1* loci); 7 of these cases displayed high level DNA amplification. The authors also made use of expression microarray data to study a gene dosage effect in these patients, and found that patients with gains of 11q had a significantly increased average expression for gained genes compared to those patients without gains (1.17 vs. 1.00, p=0.005) (Rücker *et al.*, 2006).

One of the largest studies seeking to map the extent of 11q amplifications in AML was performed using array comparative genomic hybridisation (CGH) in 12 patients with 11q amplification (Zatkova *et al.*, 2009). This study found nine recurring amplified regions on 11q, 3 of which were amplified in all cases studied. These included a similar region containing *MLL* to the Rücker *et al.* paper (117.4 to 117.9MB, 11q23.3.) as well as two amplicons spanning 126.4 to 130.9MB containing *ETS1* and *FLI1* on 11q23.3-q24.1. Expression analysis of the same group of patients confirmed increased expression of probes from within the amplified regions, as well as a gene expression signature distinct from other complex karyotype patients and *MLL* translocated AML patients (Zatkova *et al.*, 2009).

A correlation between gene dosage and the presence of trisomy 11 has also been confirmed outside of complex karyotype samples, where 828 out of 1263 (65.6%) probe sets corresponding to chromosome 11 displayed an increased expression when compared to normal karyotype control AML samples (Schoch *et al.*, 2005).

1.7.1 *TP53* Status and 11q Amplification

Both AML and MDS patients with 11q amplifications have an unusually high frequency of *TP53* mutations and 17p deletions. *TP53* mutations have been observed at a frequency of up to 94% in 11q amplification patients (Zatkova *et al.*, 2009), compared to 78% in complex karyotype AML/MDS (Haferlach *et al.*, 2008) and 13.4% in AML as a whole (Schoch *et al.*, 2006). Similarly, 17p deletions including the *TP53* locus have been reported at 32% of 11q amplification patients (Zatkova *et al.*, 2009), compared to 53% of complex karyotype AML/MDS patients (Haferlach *et al.*, 2008). The loss of wild type *TP53* function may result in genetic instability leading to the acquisition of a complex karyotype including 11q amplifications in these patients. *TP53* mutations or losses may also provide a mechanism for the chemoresistance commonly seen in AML patients with a complex karyotype (Schmitt *et al.*, 2002).

1.8 Oncogenic Drivers and Passengers

Whilst it is clear that there are many recurrent genetic changes involved in the development of both AML and cancer as a whole, many of the affected genes are not likely to be directly involved in disease development. For this reason, the terms 'driver' and 'passenger' are used when referring to different classes of genetic changes (Stratton *et al.*, 2009).

Driver mutations are beneficial to cancer cells and are positively selected for by the microenvironment in which the cell is growing. These types of mutations confer a growth advantage through a range of mechanisms, for example, in addition to proliferative advantages, the ability to adapt their microenvironment (e.g. by promoting angiogenesis), by allowing the cancer to metastasise or through providing chemotherapeutic drug resistance to a cell.

Passenger mutations are not selected for and do not provide a growth advantage. They exist because multiple genetic mutations are acquired throughout cancer development which may become fixed in a population if advantageous driver mutations are also acquired. Hence the passenger mutations are carried in the clonally expanded population by the driver mutations (Stratton *et al.*, 2009).

A difficulty arises in distinguishing driver mutations from passenger mutations. The advent of whole genome sequencing has further demonstrated that there are huge numbers of mutations present in cancer genomes, and it is likely that the majority of these are passenger mutations. Approaches to address this issue will be discussed in Chapter 4.1.

1.9 Hypothesis and Study Objectives

As has been discussed above, deletions of chromosome 12p and amplifications of 11q have been identified as recurrent abnormalities in AML, with no clear single candidate gene within either chromosomal region known to contribute to leukaemogenesis. The presence of these abnormalities in patients often attributes towards a poor outcome, and patients may benefit from a therapy targeted to the leukaemogenic drivers behind their disease. Thus clear identification of the genes driving disease progression is required as well as an understanding of their functional effects.

The hypothesis behind this study was that:

Loss of material on chromosome 12p and gains of material on chromosome 11q result in the loss or gain of genes that contribute to the development of poor prognosis AML.

The objectives of this study were:

1. To clearly define the extent of deletions and amplifications in these regions and identify a set of candidate genes that may drive leukaemogenesis.
2. To identify which genes contribute to disease development through the use of a negative selection assay performed *in vitro* and *in vivo*.
3. To investigate the relevance of targets identified in the screening assay.

Chapter 2 General Materials and Methods

2.1 Materials

2.1.1 List of Manufacturers and Suppliers

Supplier	**Location**
Abbott Molecular	Chicago, IL, USA
Affymetrix	Santa Clara, CA, USA
Agilent Technlogies	Santa Clara, CA, USA
Beckman Coulter	Brea, CA, USA
Becton Dickinson	Franklin Lakes, NJ, USA
BioRad	Hercules, CA, USA
BMG Labtech	Ortenberg, Germany
Caliper Life Sciences	Hopkinton, MA, USA
Clontech	Mountain View, CA, USA
Corning Inc.	Corning, NY, USA
Cytocell	Cambridge, UK
Enzo Life Sciences	Farmingdale, NY, USA
Eppendorf	Hamburg, Germany
Fisher Scientific	Loughborough, UK
GE Healthcare	Little Chalfont, UK
Genecopoiea	Rockville, MD, USA
Grant Instruments	Cambridge, UK
Integrated DNA Technologies	Coralville, IA, USA
Labnet International	Edison, NJ, USA
Labtech International	Uckfield, UK
Leica Biosystems	Nussloch, Germany
Life Technologies (includes Applied Biosystems, Invitrogen and Gibco)	Carlsbad, CA, USA
Macherey-Nagel	Düren, Germany
Marabu	Tamm, Germany
Merck Millipore	Billerica, MA, USA
MRC Holland	Amsterdam, Netherlands
Olympus	Tokyo, Japan
Promega	Madison, WI, USA
Qiagen	Venlo, Netherlands

Santa Cruz Biotechnlogies	Dallas, TX, USA
SANYO	Osaka, Japan
Sigma-Aldrich	St. Louis, MO, USA
StarLab	Milton Keynes, UK
Syngene	Cambridge, UK
Thermo Scientific	Waltham, MA, USA
Vector Laboratories	Peterborough, UK
VELP Scientifica	Brianza, Italy
VWR International	Radnor, PA, USA

2.1.2 List of Equipment

Equipment	**Manufacturer**
2100 Bioanalyser	Agilent Technologies
2720 PCR Machine	Applied Biosystems
5415R Microcentrifuge	Eppendorf
5418 Microcentrifuge	Eppendorf
800W Compact Microwave	Sanyo
Agarose Gel Electrophoresis Tank	BioRad
Amersham ECL Gel Box	GE Life Sciences
Bacterial Incubator	Fisher Scientific
BX61 Fluorescent Microscope	Olympus
CEQ 8800	Beckman Coulter
CO_2 Incubator	Sanyo
Evos FL Cell Imaging System	Life Technologies
FACS Calibur	Becton Dickinson
FACS Canto II	Becton Dickinson
Fluostar Omega plate reader	BMG Labtech
G-box Imaging System	Syngene
GD100 Waterbath	Grant Instruments
Gel Electrophoresis Tanks	BioRad
GeneAmp PCR System 2700	Applied Biosystems
Heraeus Multifuge 3S+ Centrifuge	Thermo Scientific
Heraeus Multifuge 3SR Chilled Centrifuge	Thermo Scientific
HYBrite Probe Hybridisation Platform	Abbott Molecular

IVIS Spectrum Bioluminescent Imaging System	Caliper Life Sciences
Nanodrop 1000 Spectrophotometer	Labtech International
Neubauer Improved Haemocytometer	Sigma-Aldrich
Optima L-100 XP Ultracentrifuge	Beckman Coulter
PCR Hood	Labcaire
REC Digital Heating Ceramic Plate	VELP Scientifica
Shaking Incubator	Eppendorf
Sub Aqua Waterbath	Grant Instruments
Technico Mini Microcentrifuge	Fisher Scientific
Veriti PCR Machine	Applied Biosystems
ViiA 7 Real Time PCR System	Life Technologies
Vortex Mixer	Labnet

2.1.3 List of Software

Software	**Developer**
7-Zip	Igor Pavlov
CytoVision 7.1	Leica Biosystems
GeneMarker	SoftGenetics
Genotyping Browser	Affymetrix
Genotyping Console	Affymetrix
Graphpad Prism 6	Graphpad Software
ImageJ 1.48	National Institutes of Health
Living Image 4.0	Caliper Life Sciences
Primer 3	Whitehead Institute for Biomedical Research
SeqNext	JSI Medical Systems
STATA 11	StataCorp
ViiA 7 Software 1.1	Life Technologies

2.1.4 List of Online Resources

Resource	**Developer**
ArrayExpress	European Bioinformatics Institute
Catalogue of Somatic Mutations in Cancer	Wellcome Trust Sanger Institute
Gene Expression Omnibus	National Center for Biotechnology Information
MAPD: MLPA Probe Design Tool	Stony Brook University
R2: Genomics Analysis and Visualization Platform	Academic Medical Center Amsterdam

2.1.5 List of Chemicals, Reagents and Materials

Reagent	**Supplier**
0.45µm PVDF Filter	Merck Millipore
100bp Ladder	Promega
10cm Tissue Culture Plates	Corning Inc.
125cm^2 Tissue Culture Flask	Corning Inc.
13mm No. 1.5 Glass Coverslip	VWR International
1kb Ladder	Promega
24mm x 50mm No. 1 Glass Cover Slide	VWR International
25cm^2 Tissue Culture Flask	Corning Inc.
75cm^2 Tissue Culture Flask	Corning Inc.
Agar	Sigma-Aldrich
Ampicillin	Sigma-Aldrich
AMPure XP Beads	Agilent Technlogies
Beckman Sample Loading Solution	Beckman Coulter
Bovine Serum Albumin	Sigma-Aldrich
Caffeine	Sigma-Aldrich
CellTiter 96 AQueous One	Promega
CEQ Size Standard 600	Beckman Coulter
Chloramphenicol	Sigma-Aldrich
Chloroquine	Sigma-Aldrich
DAPI	Vector Laboratories
Dimethylsulphoxide	Sigma-Aldrich
DMSO	Sigma-Aldrich
EDTA	Sigma-Aldrich

EndoFectin	GeneCopoeia
Ethanol	Fisher Scientific
Ficoll-Paque PREMIUM 1.077	GE Healthcare
Fixogum	Marabu
Flavopiridol	Santa Cruz Biotechnologies
GelRed	VWR International
Glycerol	Sigma-Aldrich
Human Cot-1 DNA	Invitrogen
Human Reference cDNA, Random Primed	Clontech
Hybridisation Buffer	Cytocell
IGEPAL CA-630	Sigma-Aldrich
Illustra ProbeQuant G-50 Micro Columns	GE Healthcare
Isopropanol	Fisher Scientific
Kanamycin	Sigma-Aldrich
Lenti-X Concentrator	Clontech
Loading Dye (6x)	Promega
LR Clonase II	Invitrogen
Methanol	Fisher Scientific
Mineral Oil	Beckman Coulter
Monosodium Phosphate	Sigma-Aldrich
Nitrocellulose Membrane	Sigma-Aldrich
One Shot Stbl3 Chemically Competent *E. coli*	Life Technologies
Parafilm	StarLab
PCR Markers	Promega
Phosphate Buffered Saline Tablets	Sigma-Aldrich
Polybrene	Sigma-Aldrich
Proteinase K	Sigma-Aldrich
PVDF Membrane	Merck Millipore
Saline-Sodium Citrate Buffer (20x)	Fisher Scientific
Skimmed Milk Powder	Sigma-Aldrich
SOC Medium	Life Technologies
Sodium Acetate	Sigma-Aldrich
Sodium Butyrate	Sigma-Aldrich
Sodium Chloride	Sigma-Aldrich

Spectra Multicolour BR Protein Ladder	Thermo Scientific
Stbl3 Competent Cells	Life Technologies
Superfrost Glass Microscope Slides	Thermo Scientific
Synthetic MLPA Probes	Integrated DNA Technologies
Tris Base	Sigma-Aldrich
Trypan Blue	Bio-Rad
Trypsin	Life Technologies
Tryptone	Sigma-Aldrich
Urea	Sigma-Aldrich
Vybrant DyeCycle Ruby Stain	Life Technologies
XenoLight™ D-Luciferin	Caliper Life Sciences
Yeast Extract	Sigma-Aldrich

2.1.6 List of Kits

Kit	**Supplier**
Amersham ECL Prime Western Blotting Detection Kit	GE Healthcare
BCA Protein Assay Kit	Thermo Scientific
DNeasy Blood and Tissue Kit	Qiagen
EndoFree Plasmid Maxi Kit	Qiagen
FIV Lentiviral Packaging Kit	GeneCopoeia
FXCycle PI/RNase Staining Kit	Life Technologies
High Capacity RNA-to-cDNA Kit	Life Technologies
Nick Translation DNA Labelling System	Enzo Life Sciences
Nucleobond Xtra Midi	Macherey-Nagel
P200-A1 Human DNA Reference-1 Probemix	MRC Holland
QIAPrep Spin Mini Kit	Qiagen
QIAshredder	Qiagen
RNeasy Mini Kit	Qiagen
SALSA MLPA Reagent Kit	MRC Holland
TransLenti Viral GIPZ Packaging System	GE Dharmacon
Xfect Lentiviral Packaging Kit	Clontech

2.1.7 List of Solutions

Name	**Composition**
10x TBS-T (1L)	12.11g Tris 29.22g NaCl 10ml Tween-20
FIX	75% Methanol 25% Acetic Acid
LB Agar Plates	LB Medium 1.5% w/v agar
LB Medium	1% w/v tryptone 0.5% w/v yeast extract 1% w/v NaCl pH 7
TBE 5x (1L)	54g Tris 27.5g Boric Acid 20ml 0.5M EDTA
TE	10mM Tris-HCl 0.1mM EDTA pH 8.2
Urea Buffer	7.92M Urea 100mM NaH_2PO_4 80mM Tris-HCl pH 8
WASH 1 (1L)	20ml 20X SSC 3ml NP40
WASH 2 (1L)	100ml 20X SSC 1ml igepal-CA-630
Western Blotting 10x Running Buffer (1L)	30.3g Tris 144g Glycine 10g SDS
Western Blotting 10x Transfer Buffer (1L)	144g Glycine 30.3g Tris

2.1.8 List of Culture Media

Use	Cell Lines	Identifier	Manufacturer	Supplements
General	UoC-M1, NKM-1, GDM-1, KG-1, SKNO, U937, NB4, AML-3, Kasumi-1, MV4-11, AML-2, AML-3, HL-60, THP-1	RPMI-1640	Sigma-Aldrich	10% v/v foetal calf serum 2mM GlutaMAX
General	293T, 293FT, LX293T	Dulbecco's Modified Eagle's Medium	Sigma-Aldrich	10% v/v foetal calf serum 4mM GlutaMAX 1mM sodium pyruvate
Lentiviral Production	293T, 293FT, LX293T	Dulbecco's Modified Eagle's Medium	Sigma-Aldrich	5% v/v foetal calf serum 4mM GlutaMAX 1mM sodium pyruvate
Lentiviral Packaging	293T, 293FT, LX293T	Opti-MEM I	Gibco	None

2.1.9 List of Culture Supplements

Supplement	Manufacturer
Foetal Calf Serum	Sigma-Aldrich
Foetal Calf Serum	Gibco
GlutaMAX	Gibco
Sodium Pyruvate	Sigma-Aldrich

2.1.10 List of Cell Lines Used

Cell Line Identifier	Cell of Origin	Source	Growth Density (cells/ml)
293FT	Human Embryonic Kidney	Life Technologies, USA	N/A (Adherant)
293T	Human Embryonic Kidney	Kind gift from Prof. Olaf Heidenreich, Newcastle University	N/A (Adherant)

AML-2	Acute Myeloid Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
AML-3	Acute Myeloid Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
GDM-1	Human Myelomonocytic Leukaemia	DSMZ, Germany	0.5-1.5x10^6
HL-60	Acute Promyelocytic Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
Kasumi-1	Acute Myeloblastic Leukaemia	Kind gift from Prof. Olaf Heidenreich, Newcastle University	0.5-1.5x10^6
KG-1	Human Erythroblastic Leukaemia	DSMZ, Germany	0.5-1.5x10^6
LX-293T	Human Embryonic Kidney	Clontech, USA	N/A (Adherant)
MV4-11	Acute Monocytic Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
NB4	Human Promyelocytic Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
NKM-1	Human Myeloblastic Leukaemia	HSRRB, Japan	0.5-1.5x10^6
SKNO	Human Myeloblastic Leukaemia	Kind gift from Prof. Olaf Heidenreich, Newcastle University	0.5-1.5x10^6
THP-1	Acute Monocytic Leukaemia	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
U937	Myelomonocytic Development	Kind gift from Dr. Sarah Fordham, Newcastle University	0.5-1.5x10^6
UoC-M1	Human Megakaryoblastic Leukaemia	University of Chicago, USA	0.5-1.5x10^6

2.2 General Laboratory Techniques

2.2.1 DNA Extraction

2.2.1.1 Cell Line

The DNeasy Blood and Tissue kit (Qiagen) was used to extract DNA from cultured cell lines and patient material, either grown *in vitro* culture or from xenograft mouse samples. Key points of the process involve cell lysis, before adsorption of DNA to a silica membrane contained within a spin column. The membrane is washed before elution into a microcentrifuge tube.

Up to $5x10^6$ cells per spin column were taken from cell lines or primary patient material, either fresh or as cell pellets stored at -20°C. Cell pellets were allowed to reach room temperature, and resuspended in 200µl Phosphate Buffered Saline (PBS). To the cell suspension 20µl Proteinase K and 200µl Lysis Buffer AL were added, followed by pulse vortexing to mix and a subsequent incubation at 56°C for 10 minutes. 200µl ethanol was added and the suspension again mixed by vortexing. The cell lysate was applied to the top of a DNeasy Mini spin column placed inside a collection tube and centrifuged at 6000 x *g* for 1 minute. The flow through was discarded and the spin column was transferred to a new collection tube. 500µl Wash Buffer AW1 was applied to the spin column, and centrifuged again at 6000 x *g* for 1 minute. After transfer to a new collection tube, 500µl Wash Buffer AW2 was added and centrifuged at 20000 x *g* for 3 minutes. The collection tube was then transferred to a 1.5ml microcentrifuge tube and 200µl elution buffer AE applied directly to the silica membrane of the spin column. After a 1 minute wait at room temperature, the column was centrifuged at 6000 x *g* for 1 minute. The elution process was repeated a second time to give a separate second elution. DNA samples were then measured for concentration and purity using a Nanodrop 1000 Spectrophotometer, before storage at -20°C.

2.2.1.2 Bacterial Plasmid Mini Prep

To isolate DNA plasmids for use in cloning and sequencing, the QIAprep Spin Miniprep Kit (Qiagen) was used. This kit utilises a modified alkaline lysis protocol to release DNA from smaller quantities of bacterial cells (cultures in less than 5ml Lysogeny Broth). In this protocol, an alkaline mixture of sodium dodecyl sulphate and sodium hydroxide is added to bacterial cells. This mixture lyses bacterial cell walls and membranes, but also denatures genomic and plasmid DNA. Potassium acetate solution is then added to the solution to allow the plasmid to reanneal. As genomic DNA is too long to easily renature, it becomes tangled and remains attached to bacterial proteins which are removed when the lysate is centrifuged and the plasmid containing supernatant transferred to a spin column, where it is bound to a silica

membrane. Subsequent wash steps remove endonucleases and salts, and the DNA is then eluted into a microcentrifuge tube.

Using a sterile pipette tip, a single colony was isolated from transformed bacteria spread onto Lysogeny Broth Agar (LBA) plates containing a suitable selection antibiotic. This pipette tip was then transferred to a 15ml conical centrifuge tube containing 5ml Lysogeny Broth (LB) culture media also containing selection antibiotics. The culture was incubated at 37°C overnight in a shaking incubator. Bacterial cells were harvested from the culture media by centrifugation at 6800 x *g* for 3 minutes. The bacterial pellets were then resuspended in 250µl buffer P1. To achieve cell lysis, 250µl buffer P2 was added and the suspension mixed by inversion 6 times. After a 5 minute room temperature incubation, the cell lysis was terminated by the addition of 350µl buffer N3 and mixing by inversion 6 times, causing precipitation of bacterial proteins. The lysate was then centrifuged at 18000 x *g* for 10 minutes to pellet the precipitate. The supernatant from this process was then applied to the QIAprep spin column containing the DNA binding silica membrane, before centrifugation at 18000 x *g* for 1 minute. The flow through was discarded, and 500µl buffer PB was applied to wash the membrane, with another centrifugation at 18000 x *g* for 1 minute. A second wash was then performed using 750µl buffer PE and centrifugation at 18000 x *g* for 1 minute. After discarding the flow through, an additional centrifugation at 18000 x *g* for 1 minute was performed to ensure that the silica membrane was free from ethanol. The QIAprep spin column was then placed into a clean 1.5ml microcentrifuge tube and 50µl elution buffer EB applied directly to the membrane. The spin column was allowed to stand for 1 minute at room temperature, before centrifugation at 18000 x *g* for 1 minute. The eluted sample was measured for DNA concentration and purity using a Nanodrop 1000 Spectrophotometer, before storage at -20°C.

2.2.1.3 Bacterial Plasmid Maxi Prep

For the isolation of DNA plasmids to use in the production of lentiviral particles, a larger scale extraction was required to obtain sufficient material in endotoxin free buffers amenable to tissue culture. In the EndoFree Plasmid Maxi Kit (Qiagen), bacteria are lysed using an alkaline lysis protocol, and the cell lysates containing genomic DNA are cleared by cartridge filtration. The plasmid DNA is bound to a silica membrane contained within a flow column and subjected to a number of wash steps to remove contaminants including endotoxins. Following these washes, the DNA is eluted and precipitated in isopropanol, before resuspension in TE buffer.

Bacteria were seeded onto an antibiotic selection LBA plate by spreading from frozen glycerol stocks using a sterilised inoculation loop. After overnight incubation at 37°C, a single colony was picked using a sterile pipette tip and dropped into 200ml of LB media containing a selection antibiotic, which was subsequently incubated overnight at 37°C in a shaking

incubator. In the morning, the bacteria were harvested and pelleted by centrifugation at 6000 x *g* and 4°C for 15 minutes. Pelleted bacteria were then resuspended in 8ml Buffer P1.To this, 8ml Buffer P2 was added and mixed by inversion, followed by a 3 minute incubation at room temperature to allow bacterial lysis to occur. 8ml Buffer S3 was then added to the lysate and immediately inverted 6 times, then poured into a QIAfilter cartridge. The lysate was then incubated at room temperature for 10 minutes to allow a precipitate containing proteins, genomic DNA and detergents to float to the top of the QIAfilter cartridge. At this stage, the plunger was inserted into the QIAfilter cartridge and the cell lysate filtered into a new tube. The QIAfilter cartridge containing the remaining precipitate was discarded at this stage.

To the filtered lysate, 2.5ml buffer ER was added, the solution mixed and incubated on ice for 30 minutes to remove endotoxins. At this stage, the Qiagen-tip 500 was equilibrated by adding 10ml buffer QBT. The filtered lysate was then added to the Qiagen-tip 500 after incubation on ice and allowed to flow through by gravity. Two 30ml volumes of wash buffer QC were added and allowed to flow through, before plasmid DNA was eluted using 15ml of buffer QN. The plasmid DNA was then precipitated with 10.5ml isopropanol and centrifuged for 30 minutes at 15000 x *g* and 4°C. The supernatant was carefully decanted to ensure that the plasmid DNA pellet was not disturbed. The pellet was then washed with 5ml 70% ethanol in endotoxin free water and centrifuged for 10 minutes at 15000 x *g* and 4°C. The supernatant was again carefully decanted to ensure that the DNA pellet was not disturbed. The pellet was then air dried at room temperature for 10 minutes and resuspended in 300µl buffer TE. The DNA yield and purity was then determined using a Nanodrop 1000 Spectrophotometer and plasmid DNA stored at -20°C.

2.2.2 Freezing Bacterial Stocks

For the long term storage of bacterial stocks, 500µl of overnight bacterial culture was mixed with 500µl of 50% glycerol diluted with distilled H_2O. The resulting suspension was mixed by pulse vortexing and transferred to a 1.5ml microcentrifuge tube, before storage at -80°C.

2.2.3 RNA Extraction

2.2.3.1 QIAshredder

The QIAshredder kit (Qiagen) was used to homogenise cell lines or mouse tissues prior to RNA extraction, providing an increased yield of RNA. The QIAshredder columns homogenise lysate as they pass through the column under forces created by centrifugation. Up to 1 x 10^6 cells were resuspended in 700µl of buffer RLT and loaded into the column. The column was placed into a 2ml collection tube and centrifuged for 2 minutes at 20000 x *g*.

2.2.3.2 RNeasy Mini Kit

The RNeasy Mini kit (Qiagen) was used to extract RNA from homogenised lysates produced by the QIAshredder kit. The RNeasy spin columns allow for on column DNase digestion, before binding RNA. The bound RNA is then washed with a series of buffers and finally eluted into RNase-free water for downstream applications. The homogenised lysate from the QIAshredder kit was applied to an RNeasy spin column and centrifuged for 15 seconds at 8000 x *g*. The flow through was discarded and 350µl of buffer RW1 was added to the column. Following centrifugation at 8000 x *g* for 15 seconds, the flow through was discarded and 10µl DNase I solution was mixed with 70µl of buffer RDD before being added directly to the RNeasy spin column membrane. After incubation at room temperature for 15 minutes, 350µl buffer RW1 was added to the spin column and centrifuged for 15 seconds at 8000 x *g*, and the flow through discarded. Then 700µl buffer RW1 was added and the column was again centrifuged at 8000 x *g* for 15 seconds and the flow through discarded. This was followed by the addition of 500µl buffer RPE with the same centrifugation step and again the flow through was discarded, before repeating the addition of buffer RPE and subsequent steps again. The RNeasy spin column was then placed into a 1.5ml collection tube and 30µl RNase-free water was added directly to the column membrane. The column was centrifuged for 1 minute at 8000 x *g*, and this process repeated with another 30µl RNase-free water to give a total elution volume of 60µl. RNA concentration was then calculated through the measurement of spectral absorbance using a Nanodrop 1000 spectrophotometer.

2.3 General Cell Culture Techniques

2.3.1 Routine Cell Maintenance

2.3.1.1 Suspension Cell Lines

Leukaemic cell lines grown in suspension were cultured in upright flasks and cultured in an incubator providing a humid environment containing 5% CO_2 heated to 37°C. Cells were counted every 2 to 3 days and diluted to a concentration of $0.5x10^6$ cells/ml. Suspension cell lines were cultured in RPMI-1640 media supplemented with 10% Foetal Calf Serum (FCS) and 2mM GlutaMAX.

2.3.1.2 Adherent Cell Lines

The cell line derivatives of HEK293 were cultured in flasks laid flat in an incubator providing a humid environment with 5% CO_2 at 37°C, in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FCS, 4mM GlutaMAX and 2mM sodium pyruvate. Cells were split 3 times a week in a suitable ratio between 1:6 and 1:10 of cell suspension to fresh media. To split cells, old media was aspirated from the cell layer and the cells were gently washed in 10ml PBS. 2ml of a 0.25% Trypsin solution was applied to the cells for 2 minutes and the cells detached by gently tapping the side of the flask. 10ml fresh media was then added to the cell suspension to neutralise the Trypsin and cells were pelleted by centrifugation at 400 x *g* for 5 minutes. The supernatant was aspirated and the cell pellet resuspended in a suitable volume of fresh media before transfer to a new tissue culture flask. Passage numbers were recorded and only cells between passage 5 and passage 20 were used for the production of lentivirus.

2.3.2 Cell Counting

Cell concentration was estimated using a Neubauer Improved Haemocytometer. Cells were mixed in a 1:1 ratio with trypan blue and 10µl of the mixture loaded into the counting chamber. The counting chamber consists of four counting quadrants, each of which has a volume of $0.1mm^3$. Viable cells (which do not take up the trypan blue dye) across all four quadrants were counted and an average count calculated by dividing the total number of cells by 4. This number was then multiplied by 20000 to give a cell concentration in cells per millilitre.

2.3.3 Freezing Cell Stocks

For the long term storage of cell lines, $5x10^6$ cells were resuspended in 0.5ml of media containing 10% FCS. This was then mixed with 0.5ml of freezing media, consisting of 80% FCS and 20% DMSO, and the cell suspension transferred to a cryovial. Cryovials were inserted into a Mr. Frosty freezing container filled with isopropanol to ensure controlled cell cooling at a rate of -1°C per minute when placed into a -80°C freezer. After 24 hours at -80°C, cells were

transferred to storage racks contained within in the gaseous phase of liquid nitrogen for long term storage at -130°C.

Chapter 3 Determining Minimal Regions of Copy Number Alteration

3.1 Introduction

3.1.1 Early Genetic Mapping Techniques

Chromosomal abnormalities were first observed in tissue samples from a range of tumours in the late nineteenth century and were believed to play a role in tumourigenesis (Von Hansemann, 1890). However, techniques for visualising chromosomes were of an extremely low resolution and as a result, for over 30 years it was commonly accepted that humans had 48 chromosomes (Painter *et al.*, 1921). Gradually, techniques were developed and chemicals, such as colchicine which arrests cells in a mitotic state, were introduced which greatly improved the quality of chromosomal preparations. These developments led to the correct human chromosomal number of 46 being recognised in 1956 (Tjio and Levan, 1956), which allowed conditions such as Down syndrome (trisomy 21) and Turner syndrome (-X) to be linked to the gain or loss of specific chromosomes.

At this time, it was believed that most tumours had a normal complement of chromosomes (Bayreuther, 1960), thus there was little interest in studying the chromosomes of tumour samples. However, in 1960, Nowell and Hungerford discovered a recurrent chromosomal abnormality in the neoplastic cells of patients with CML. This small chromosome was named the Philadelphia chromosome after the city in which it was discovered (Nowell and Hungerford, 1960). Since this discovery, thousands of other chromosomal aberrations have been identified and associated with cancer development, and many techniques for their identification have been developed.

Until early in the 1970s, cytogenetic studies were performed on chromosomes using stains that did not reveal a banding pattern, which made identifying individual chromosomes and detecting structural aberrations difficult. Torbjörn Caspersson and his colleagues showed that a banding pattern visible under the microscope could be unveiled using quinacrine fluorescent staining, known as Q-banding (Caspersson *et al.*, 1968; Caspersson *et al.*, 1970). This technique allowed the consistent recognition of each chromosome, but was flawed due to the short period of quinacrine fluorescence, making the technique unsuitable for routine patient studies. This led to the development of a number of novel banding techniques, the most successful and widespread of which has been G-banding.

In the G-banding technique, preparations are stained by an application of trypsin followed by Giemsa (Seabright, 1971). This staining method led to approximately 500 bands per haploid

genome, allowing the identification of many novel structural aberrations such as translocations, inversions, deletions and duplications, in addition to the chromosomal number changes previously known. The resolution of the technique was increased through the synchronisation of cultures and preparation of slides from cells in pro-metaphase or prophase stages of the cell cycle rather than metaphase (Yunis, 1976). These chromosomes appear longer under the microscope and over 1000 bands can be observed in a normal haploid genome, allowing for the more precise identification of chromosomal aberrations. The use of G-banding has allowed for a standardised nomenclature of physical gene locations on a chromosome to be established. An example of this is the alpha tubulin gene, located at 12q13.12. The first 12 refers to the chromosome number, whilst the q refers to the long chromosome arm (as opposed to p which signifies the short chromosome arm). The numbers following this letter represent the position on the chromosomal arm: in this case, region 1, band 3, sub-band 12 (Strachan and Read, 1999).

3.1.2 FISH

Fluorescence in situ hybridisation (FISH) techniques were developed in the early 1980s and involve the use of single stranded DNA probes which bind complementary genomic DNA sequences in both interphase and metaphase cells (Langer-Safer *et al.*, 1982). These probes are labelled with various fluorophores, which fluoresce when excited at appropriate wavelengths and can be viewed using a fluorescence microscope. Due to the flexibility that FISH provides, both in the sequences that can be targeted and the range of available fluorophores, a number of different targets can be probed at the same time, and can be directed towards specific parts of a gene. This is highly useful in a diagnostic setting, where recurrent abnormalities can be quickly screened for in a large number of samples. FISH is however limited in that it will only detect the specific sequences that are being tested for, so finding novel abnormalities can be difficult without some prior understanding of their nature. Although some advanced probing techniques, such as whole chromosome painting and M-FISH, are available (Ried *et al.*, 1998), FISH remains limited in its ability to provide a whole genome screening process and other techniques are typically used for these applications.

3.1.3 Comparative Genomic Hybridisation

Comparative genomic hybridisation (CGH) was developed as a method for screening cancer genomes and comparing them to control DNA metaphase spreads, in order to ascertain genomic copy number at a range of chromosomal locations (Weiss *et al.*, 1999). In this technique, tumour DNA and control DNA are labelled with fluorophores and mixed together. Cot-1 DNA is added to suppress repetitive sequences and the mixture is hybridised to normal human metaphases, where they can be scanned using a fluorescence microscope. The ratio of

tumour to control DNA hybridised to each point of the metaphase spread can then be measured, and regions of gain and loss identified. There are a number of limitations to CGH however, such as its low resolution; only able to detect abnormalities greater than 5MB in length (Weiss *et al.*, 1999).

This technique has since been further developed into what is known as array CGH (aCGH). Metaphases were replaced with bacterial artificial chromosome (BAC) clones, containing regions of the genome that were of particular interest, and spotted onto glass slides (Garnis *et al.*, 2004). The normal and tumour DNA samples are prepared using a similar process to regular CGH and co-hybridised to the spotted BAC clones, where they are scanned and fluorescence levels assessed. As the technology has progressed, BAC clones have been replaced by cDNA clones, PCR products and more recently, synthetic oligonucleotides (Ylstra *et al.*, 2006). The use of synthetic oligonucleotides carries a number of advantages. Their short sequences reduce cross hybridisation from off target sequences. Newer spotting technologies, such as inkjet spotting and photolithography, allow a large number of probes to be spotted in small spaces with high accuracy, further increasing the number of probes for greater depth of genome coverage. These improvements have achieved almost genome wide coverage in the current generation of copy number arrays. Oligonucleotide arrays have also opened up the possibility for genotyping to be carried out using single nucleotide polymorphism (SNP) probe arrays (Wang and Bucan, 2008).

Similar to copy number arrays, SNP arrays contain many probes hybridised to a glass slide (LaFramboise, 2009). However, each probe is also designed to interrogate the identity of various SNPs located within the genome. They were developed from single locus SNP genotyping techniques, which interrogated the identity of a single SNP using a range of methods, from PCR and enzyme based strategies to methods relying on the hybridisation of specific oligonucleotides. Unlike copy number arrays, only sample DNA is hybridised to the oligonucleotides, which is instead compared to reference control DNA samples carried out in previous experiments. Each SNP is interrogated by a 25 base pair sequence which hybridises to both alleles of a SNP. If all 25 base pairs match (known as a perfect match), the signal produced from the probe is brighter than if only 24 base pairs match (known as a mismatch). Each SNP is interrogated by two probes, each of which is assigned as either allele A or allele B. From these, 3 possible genotype outcomes are computed for each SNP depending on which alleles are present – AA, AB or BB (LaFramboise, 2009).

Both copy number and genotyping information can be extracted from SNP arrays. For the study of cancers, this is invaluable in providing information about loss of heterozygosity (LOH). LOH refers to the loss of a chromosomal region, which can result in a net change in copy

number, or be copy neutral (CN-LOH). CN-LOH can drive cancer development by acting as the 'second hit' in the classical two-hit hypothesis of tumourigenesis as proposed by Alfred Knudson (Mao *et al.*, 2007). If one copy of a tumour suppressor gene is mutated, loss of the wild type gene and duplication of the mutated copy leaves a cell without a functional wild type copy of the gene, and thus cancer may progress.

This project has focussed on the use of 3 recent Affymetrix SNP arrays, the Human Mapping 100K SNP Array Set (100K), the Human Mapping 500K Array Set (500K) and the Genome-Wide Human SNP Array 6.0 (SNP 6.0). These were chosen as the majority of publicly available SNP array data for AML makes use of one of these platforms. The main features of these arrays, along with the recently released CytoScan HD Array, are listed in Table 7. The newer SNP 6.0 and CytoScan HD arrays are composed of a mixture of probe types, some of which are solely used in the detection of copy number variation and provide increased levels of coverage.

	Total Probes	**Copy Number Probes**	**SNP Probes**	**Release Year**
100K	100,000	0	100,000	2006
500K	500,000	0	500,000	2006
SNP 6.0	1,800,000	946,000	906,600	2008
CytoScan HD	2,600,000	1,850,000	750,000	2012

Table 7. A comparison of the 3 most recently released Affymetrix SNP arrays for determining copy number and genotyping information (information obtained from Affymetrix: www.affymetrix.com).

With the increasingly widespread use of aCGH on large numbers of samples, researchers are usually encouraged to share their data with other groups who may be able to extract further information from data sets generated. To facilitate this sharing, websites such as ArrayExpress (http://www.ebi.ac.uk/arrayexpress/), and the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) have been established. Datasets can be searched for by their sample type and downloaded for further analysis by other researchers, allowing additional information that may not have been relevant to the original project to be extracted.

3.1.4 Multiplex Ligation-dependent Probe Amplification

To allow for the simultaneous screening of common copy number aberrations in large numbers of samples, the Multiplex Ligation-dependent Probe Amplification (MLPA) technique was devised (Schouten *et al.*, 2002). The technique utilises probe pairs to detect changes in copy number. Each probe contains a universal primer recognition site in conjunction with target specific sequences for the region to be investigated. One of the probes also contains a 'stuffer sequence' to increase the length of the product sequences for electrophoretic

resolution, with each probe pair in a probe mix assigned a different target size. The target sequences in each probe pair are each one half of a complete complementary sequence to the exon or gene of interest, and so are located adjacent to each other when hybridised to the same strand of genomic DNA. A ligation reaction is then able to seal the gap between the two probe sequences, PCR is used to amplify the ligated probe sequences, whilst unligated probe halves are unable to be amplified (Schouten *et al.*, 2002).

The PCR reaction amplifies probes in a manner proportional to their copy number status. The PCR products can be separated using capillary electrophoresis, where relative peak heights can be measured and compared to produce an estimate of copy number for the regions interrogated (Schouten *et al.*, 2002). The technique is summarised in Figure 5.

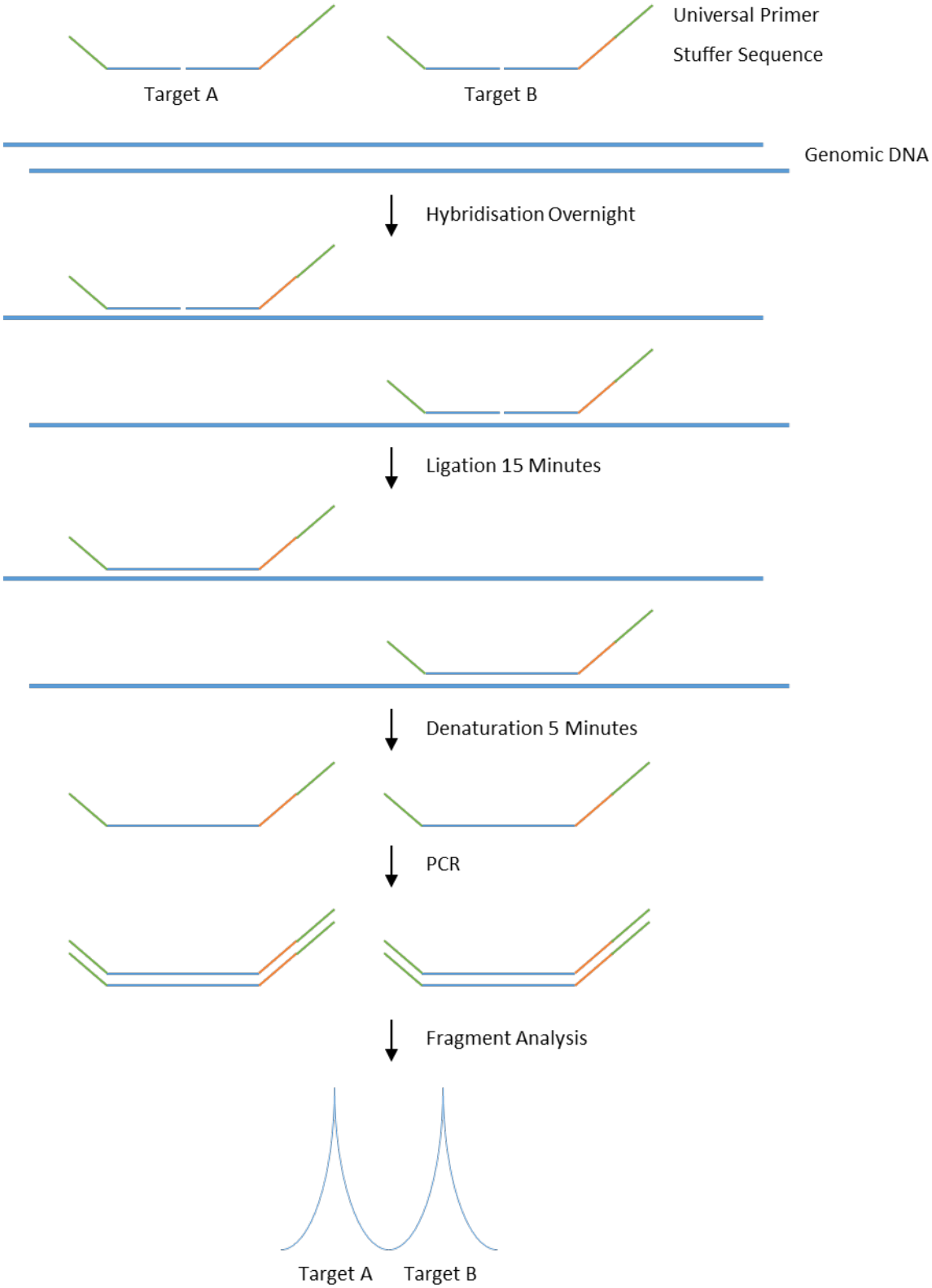


Figure 5. The MLPA technique. Genomic DNA is interrogated with complementary probe pairs that require an exact match to bind. The nick in between a complementary probe pair is ligated during the ligation reaction when bound to DNA. The samples are then denatured to remove bound probes from the DNA, before PCR amplification using universal primers to amplify successfully ligated probes. DNA fragments can then be separated by capillary electrophoresis to identify the relative copy number of each sequence.

3.2 Aims

Despite a number of studies investigating deletions of 12p and 11q amplifications, no clear minimal regions of deletion or amplification have been established. The objectives of the studies described in this chapter were therefore:

- To establish a minimally deleted region on the short arm of chromosome 12 using publicly available copy number array data.
- To refine this minimally deleted region using MLPA on normal karyotype patient samples.
- To establish a minimal region or regions of amplification on the long arm of chromosome 11 using publicly available copy number array data.

3.3 Methods

3.3.1 Affymetrix Copy Number Array Analysis

Publicly available AML datasets were selected from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) website and raw data (.CEL) files were downloaded as compressed .TAR archives and extracted using 7zip. Raw data files were loaded into the Affymetrix Genotyping Console and underwent quality control checks before further analysis took place to ensure the integrity of the data. Genotyping data were compiled for each data file, before copy number and LOH data were calculated using the Birdseed v2 algorithm. The Affymetrix Genotyping Console Browser software was used to align the copy number and LOH calls to a suitable build of the human genome. For data generated from the Affymetrix GeneChip Human Mapping 500K Array this was the hg18 (NCBI36) assembly, whereas data from the Affymetrix Genome-Wide Human SNP Array 6.0 were aligned to the hg19 (GRCh37) assembly of the human genome. Copy Number Segmentation files were also generated using the Affymetrix Genotyping Console, with a minimum limit of 5 consecutive concurrent probe calls required for an altered copy number state to be identified by the software.

Regions of interest containing copy number alterations were recorded by their location on the human genome. Coordinates obtained from 500K analysis were translated from the hg18 to the hg19 build using the LiftOver service provided by the University of California, Santa Cruz (http://genome.ucsc.edu/cgi-bin/hgLiftOver) to allow direct comparison of results across different array platforms. Where matched remission or germline samples were available, these were compared to the samples taken at diagnosis to identify cases where somatic copy number alterations were present.

3.3.2 Multiplex Ligation-dependent Probe Amplification

3.3.2.1 Probe Design

Synthetic MLPA probes were designed following the instructions provided in the "Designing synthetic MLPA probes" manual (MRC Holland) and the MAPD: MLPA Probe Design tool from Stony Brook University, New York (http://bioinform.arcan.stonybrook.edu/mlpa2/cgi-bin/mlpa.cgi). Probes were synthesised by Integrated DNA Technologies and mixed with the MRC Holland SALSA MLPA P200-A1 Human DNA Reference-1 probemix (MRC Holland).

Target Gene	Product Size	Left Probe Sequence	Right Probe Sequence
CREBL2	**104**	GGGTTCCCTAAGGGTTGGATCTGGCATCAGGGAAGCGAACTGTTAGGCGA	GAGGAGGAGGCAGCCAGAACCATATCCCCTTTCTAGATTGGATCTTGCTGGCAC
APOLD1	**108**	GGGTTCCCTAAGGGTTGGATCCTGAGTGGCTCCCAATCCAGTCCTCCAAGTA	CTCAGAGGGGAAGCCCGTGAAGCCGTCATCTAATCTAGATTGGATCTTGCTGGCAC
LRP6	**112**	GGGTTCCCTAAGGGTTGGAGGCAGTCAGTTTGTGGTCACTGCTCAAATTGCCCA	TCCTGATGGTATTGCTGTGGACTGGGTTGCACGAATCTAGATTGGATCTTGCTGGCAC
CDKN1B	**118**	GGGTTCCCTAAGGGTTGGAGGTTTTGTTTTTTTGAGAGTGCGAGAGAGGCGGTCGTG	CAGACCCGGGAGAAAGATGTCAAACGTGCGAGTGTCTATCTAGATTGGATCTTGCTGGCAC
CDKN1B	**126**	GGGTTCCCTAAGGGTTGGACAGAACAGAAGAAAATGTTTCAGACGGTTCCCCAAATGCCGG	TTCTGTGGAGCAGACGCCCAAGAAGCCTGGCCTCAGAAGACGTCTAGATTGGATCTTGCTGGCAC
ETV6	**130**	GGGTTCCCTAAGGGTTGGATCTGGGTTGGGGAGAGGAAAGGAAAGTGGAAAAAACCTGAGAAC	TTCCTGATCTCTCTCGCTGTGAGACATGTCTGAGACTCCTGCTCTCTAGATTGGATCTTGCTGGCAC
ETV6	**136**	GGGTTCCCTAAGGGTTGGAGAAGCCCATCAACCTCTCTCATCGGGAAGACCTGGCTTACATGAACC	ACATCATGGTCTCTGTCTCCCCGCCTGAAGAGCACGCCATGCCCATTTCTAGATTGGATCTTGCTGGCAC
ANO2	**142**	GGGTTCCCTAAGGGTTGGAGCTTTTGTCATTGCGATCACCTCCGACTTTATCCCCCGCCTGGTGTACCA	GTACTCCTACAGTCACAATGGGACTCTGCACGGCTTTGTCAACCACACCCTCTAGATTGGATCTTGCTGGCAC
MANSC1	**146**	GGGTTCCCTAAGGGTTGGAGACCAAAACTAAACTGAAATTTAAAATGTTCTTCGGGGGAGAAGGGAGCTTG	ACTTACACTTTGGTAATAATTTGCTTCCTGACACTAAGGCTGTCTGCTAGTCTCTAGATTGGATCTTGCTGGCAC

Table 8. The MLPA probes designed to assess 12p copy number status. Probes were designed using the MAPD: MLPA Probe Design Tool and synthesised by Integrated DNA Technologies.

3.3.2.2 Multiplex Ligation-dependent Probe Amplification Reaction

50ng of sample DNA was diluted to 10ng/µl using 0.1M Tris-EDTA (TE) buffer (pH 8.2) and placed into individual tubes. TE buffer only was used as a negative control tube. These tubes were then placed into a thermocycler (Applied Biosystems) for the DNA denaturation process to take place, by heating the samples to 98°C for 5 minutes. After cooling to 25°C, the samples were removed from the thermocycler and 3µl of a hybridisation master mix were added. The hybridisation master mix consists of 1µl P200 reference probemix (MRC Holland), 0.5µl synthetic probemix (Integrated DNA Technologies) and 1.5µl MLPA buffer (MRC Holland) per reaction. The tubes were then returned to the thermocycler where they were incubated for 1 minute at 95°C then held at 60°C for between 16 and 20 hours.

After this incubation, samples were cooled to 54°C and 32µl of Ligase-65 master mix was added to each tube. The Ligase-65 master mix consisted of 25µl dH_2O, 3µl Ligase Buffer A (MRC Holland), 3µl Ligase Buffer B (MRC Holland) and 1µl Ligase-65 enzyme (MRC Holland) per reaction. Samples were then incubated at 54°C for 15 minutes to allow ligation to occur before heating to 98°C for 5 minutes in order to heat inactivate the Ligase-65 enzyme. The samples were then cooled to 20°C and 10µl of PCR master mix added to each sample. The PCR master mix consisted of 7.5µl dH_2O, 2µl SALSA PCR primer mix (MRC Holland) and 0.5µl SALSA polymerase (MRC Holland) per reaction. The PCR thermocycler programme was then initiated, with 35 PCR cycles of 95°C for 30 seconds, followed by 60°C 30 seconds and then 72°C for 60

seconds. After the PCR cycles were complete, the samples were incubated at 72°C for 20 minutes before being chilled to 15°C then removed from the thermocycler. This process is summarised in Table 9.

Process	Step	Temperature	Time (mm:ss)
DNA Denaturation	1	98°C	05:00
	2	25°C	*Hold*
Hybridisation Reaction	3	95°C	01:00
	4	60°C	*Hold*
Ligation Reaction	5	54°C	*Hold*
	6	54°C	15:00
	7	98°C	05:00
	8	20°C	*Hold*
PCR Reaction	9 – 35 Cycles:	95°C 60°C 72°C	00:30 00:30 01:00
	10	72°C	20:00
	11	15°C	*Hold*

Table 9. Multiplex Ligation-dependent Probe Amplification One Tube Protocol PCR cycling conditions.

3.3.2.3 Separation by Capillary Electrophoresis

After the PCR reaction was completed, the MLPA products were separated by capillary electrophoresis to identify the relative concentrations of the various products. 0.7µl of the PCR reaction were mixed with 0.2µl of CEQ Size Standard 600 (Beckman Coulter) and 32µl of Beckman Sample Loading Solution (Beckman Coulter) in a 96 well plate. One drop of mineral oil (Beckman Coulter) was added to the top of each sample and the plate loaded into the CEQ-8800 (Beckman Coulter) equipped with 33cm capillaries. The electrophoresis was then performed utilising the settings detailed in Table 10.

Run Parameter	Setting
Run Method	Frag
Capillary Temperature	50°C
Denaturation	90°C for 120 seconds
Injection Voltage	1.6kV
Injection Time	30 seconds
Run Time	60 minutes
Run Voltage	4.8kV

Table 10. Beckman CEQ8800 Run Settings for Multiplex Ligation-dependent Probe Amplification.

The data produced from the capillary electrophoresis run were analysed using the GeneMarker Software (SoftGenetics). For the analysis settings, MLPA Ratio was used as the analysis method and Peak Height used as the quantification method. Ratios of probe intensity between control and sample of less than 0.75 were classed as deletion and more than 1.3 were classed as gain.

3.3.3 Fluorescence *In Situ* Hybridisation

3.3.3.1 Extraction of Bacterial Plasmid

Plasmids were selected to target specific genes from a database of stocks currently held by the Leukaemia Research Cytogenetics Group (LRCG). These stocks consist of a bacterial suspension mixed with glycerol stored at -80°C. To extract plasmid DNA for use in FISH procedures, the Nucleobond Xtra Midi kit (Macherey-Nagel) was used. This kit is based on the NaOH/SDS lysis method of releasing plasmid produced by bacteria. The lysate is then filtered by gravity flow through a column filter and plasmid is bound to a silica resin. Whilst bound to the resin, the plasmid is washed before elution and precipitation, then is finally resuspended in water.

A sterile pipette tip was used to take a pick from the required stock and the bacteria streaked out on an LB-agar plate containing an appropriate selection antibiotic for the bacteria. After incubation at 37°C for 16 hours, a single colony was picked and used to inoculate a starter culture of 5ml LB containing the appropriate selection antibiotic. After 8 hours of shaking incubation at 37°C, 200µl of the starter culture was transferred to 200ml LB containing selection antibiotics for a further 16 hours shaking incubation at 37°C. The bacteria were then harvested by centrifugation at 6000 x *g* at 4°C for 15 minutes.

After centrifugation, the bacterial pellet was resuspended in 8ml resuspension buffer (Macherey-Nagel), and 8ml lysis buffer (Macherey-Nagel) added to the suspension before incubation at room temperature for 5 minutes. During this period, the column and filter were equilibrated using 12ml equilibration buffer (Macherey-Nagel). The buffer was slowly applied to the top rim of the column in an anti-clockwise motion. After the 5 minute incubation of the lysate was complete, the lysis reaction was stopped by the addition of 8ml neutralisation buffer (Macherey-Nagel), and the tube inverted three times to ensure even mixing. The lysate was then applied to the column filter. This was achieved by slowly pipetting the lysate onto the rim of the column in an anti-clockwise motion. After the lysate had flowed through the column by gravity flow, 5ml equilibration buffer was then applied to the column in the same anti-clockwise motion as before, and the column filter discarded. 8ml of wash buffer (Macherey-Nagel) was then applied to the silica membrane and allowed to clear the column by gravity flow. The column was then transferred to a sterile 50ml tube and the plasmid DNA eluted by the addition of 5ml elution buffer (Macherey-Nagel).

Plasmid DNA was then precipitated with the addition of 3.5ml isopropanol, followed by centrifugation at 15,000 x *g* for 5 minutes. The isopropanol was removed by pipetting and the plasmid DNA pellet air dried for 10 minutes at room temperature. The DNA was then reconstituted in dH_2O.

3.3.3.2 Fluorescent Labelling of FISH Probes

The extracted DNA was labelled using the Nick Translation DNA labelling system (Enzo Life Sciences). Nick translation makes use of the actions of DNA Polymerase I and DNase I to replace unmodified nucleotides with fluorophore-labelled nucleotides. DNase I introduces multiple nicks in DNA, whilst DNA Polymerase I catalyses the incorporation of nucleotides to the 3'-OH terminus of the nick and simultaneously removes nucleotides from the 5'-PO_4 terminus. In a clean 1.5ml microcentrifuge tube, the following reagents were added (all Enzo Life Sciences); 1µg DNA template in 25µl of dH_2O; 5µl reaction buffer; 5µl dNTP mix; 2.5µl dTTP mix; 2.5µl Fluorophore-dUTP; 5µl DNA Polymerase I; and 5µl DNase I. The reagents were mixed by flicking and incubated for 120 minutes at 15°C before chilling on ice. The reaction was then terminated by the addition of 5µl Stop Buffer and heating to 65°C for 5 minutes.

The nick translated probes were then precipitated to allow the preparation of a hybridisation mix. 5.5µl of the nick translation reaction was added to a fresh 1.5ml microcentrifuge tube with 3µg Human Cot-1 DNA (Invitrogen) and the mixture made up to a total volume of 16µl with dH_2O. 1.6µl of 3M sodium acetate (Sigma-Aldrich) was added, followed by 40µl of ice-cold 100% ethanol before mixing by pulse vortexing and incubation on ice for 15 minutes. The DNA was pelleted by centrifugation at 13,000 x *g* for 30 minutes in a microcentrifuge chilled to 4°C. The pellet was then resuspended in 6µl dH_2O and 14µl hybridisation buffer (Cytocell, UK).

Probes were then purified using illustra ProbeQuant G-50 Micro Columns (GE Healthcare), which utilise Sephadex G-50 to purify probes by gel filtration in order to remove unincorporated nucleotides left by the nick translation. The column was first prepared by pulse vortexing to resuspend the Sephadex resin. The cap was then removed from the base of the column, and the Sephadex column placed into a collection tube. The tube was centrifuged at 735 x *g* for 1 minute to slightly dry the column. The probe suspension was then immediately applied to the centre of the Sephadex resin and the column placed in a fresh 1.5ml microcentrifuge tube. The column was then centrifuged for 2 minutes at 735 x *g* to elute the purified sample into the microcentrifuge tube and purified probes were stored at -20°C.

3.3.3.3 General FISH Technique

FISH was performed on 76mm x 26mm Superfrost glass microscope slides (Thermo Scientific) cleaned with fresh fixative. 2µl of fixed cell nuclei suspension was applied to the centre of the slide and allowed to air dry. Probe solution was mixed 1:1 with hybridisation buffer (Cytocell) in a 0.5ml microcentrifuge tube and 2µl of the mixture applied to a 13mm diameter No. 1.5 glass coverslip (VWR International). The coverslip was applied to the dried fixed cells present on the microscope slide and sealed using Fixogum (Marabu), before transferring the slides to a HYBrite probe hybridisation platform (Abbott Molecular) for hybridisation. Hybridisation was

performed by heating the samples to 75°C for 5 minutes, followed by incubation at 37°C for 16 hours. After this incubation step, the gel fixative was removed and the slide was soaked in 2xSSC solution until the coverslip fell away. The slides were then transferred to WASH 1 (Chapter 2.1.5) and incubated at 72°C for 2 minutes, followed by incubation at room temperature for 2 minutes in WASH 2 (Chapter 2.1.5). Slides were then removed from the wash and excess solution drained, before 7µl of 4',6-diamidino-2-phenylindole (DAPI) solution (Vector Laboratories) was applied and the microscope slide covered with a 24mm x 50mm No. 1 glass cover slide (VWR International).

FISH slides were viewed using a BX61 fluorescence microscope (Olympus) equipped with a 100x magnification, 1.30 aperture oil-immersion objective lens.

3.4 Results

3.4.1 Copy Number Array Analysis

Publicly available copy number array data from AML patients were downloaded from the Gene Expression Omnibus and analysed as described in Section 3.3.1. This allowed analysis of 866 patients across a total of 1442 data files including matched controls and relapse samples.

3.4.1.1 Cohort Characteristics

The cohorts used each had a different selection criterion as indicated in Table 11. The mean ages of the studies, where this information was available, were generally low when compared to the mean age of diagnosis in AML which is usually around 70 years (Kumar, 2011). The percentage of males included for study was also slightly lower than expected, with 55% of AML cases diagnosed in males (CRUK, 2014).

Study	GEO Accession	Array Type(s)	Number of Patients	Number of Matched Controls	Selection	Mean Age (Years)	% Male
Radtke *et al.* 2009	GSE15347	500K, 100K	111	65	Paediatric AML	9.3	49
Bullinger *et al.* 2010	GSE19101	500K, 100K	157	157	Normal Karyotype	46.1	46
Gupta *et al.* 2012	GSE20672	SNP 6.0	56	56	Adult AML	-	-
Barresi *et al.* 2010	GSE21780	SNP 6.0	19	11	Normal Karyotype	43.2	47
Parkin *et al.* 2010	GSE23452	SNP 6.0	114	112	Non AML-M3	61.3	58
Haferlach *et al.* 2011	GSE27832	SNP 6.0	7	0	12p Deletion	70.7	43
Kühn *et al.* 2010	GSE32462	SNP 6.0	300	175	AML-M3	-	-
Rücker *et al.* 2012	GSE34542	SNP 6.0, 500K	102	0	Complex Karyotype	-	-
Total			**866**	**576**			

Table 11. The AML patient cohorts utilised in this study. Where age and sex information was available, this is also displayed.

3.4.1.2 Deletions of Chromosome 12p

The overall incidence of 12p deletions across those data sets that had not been specifically selected for the detection of 12p deletions was 45 of 859 patients (5.2%), whilst deletions of 12p were detected in 2 of 111 cases (3.6%) from the paediatric AML data set (Radtke *et al.*, 2009, GSE15347). A total of 6 patients had deletions encompassing the entire 12p chromosome arm. An example of a patient with a deletion is shown in Figure 6, where it can be seen that the copy number is reduced to one.

Figure 6. An example of a patient with a deletion of chromosome arm 12p.Sample GSM393735 from series GSE15347 (Radtke et al., 2009). The deleted region is highlighted in red.

Detailed information on each patient found to have a deletion of 12p is displayed in Supplementary Table 34. In all cases except two, the *ETV6* gene was within the minimally deleted region. One of these exceptions was previously reported by Haferlach *et al.* (2011), in which the region of deletion spanned 12,113,639 base pairs from the 12p telomere to 13,763,702. The second patient to exclude *ETV6*, from the Rücker *et al.* (2012) study, had a deleted region from 12,120,652 to 13,495,401, after remapping of genomic coordinates from hg18 to hg19.

Through the alignment of deletions (after conversion of genome coordinates to the same human genome build - hg19), a MDR was derived, which contained 10 genes: *BCL2L14*, *LRP6*, *MANSC1*, *LOH12CR1*, *DUSP16*, *CREBL2*, *GPR19*, *CDKN1B*, *APOLD1* and *DDX47*.

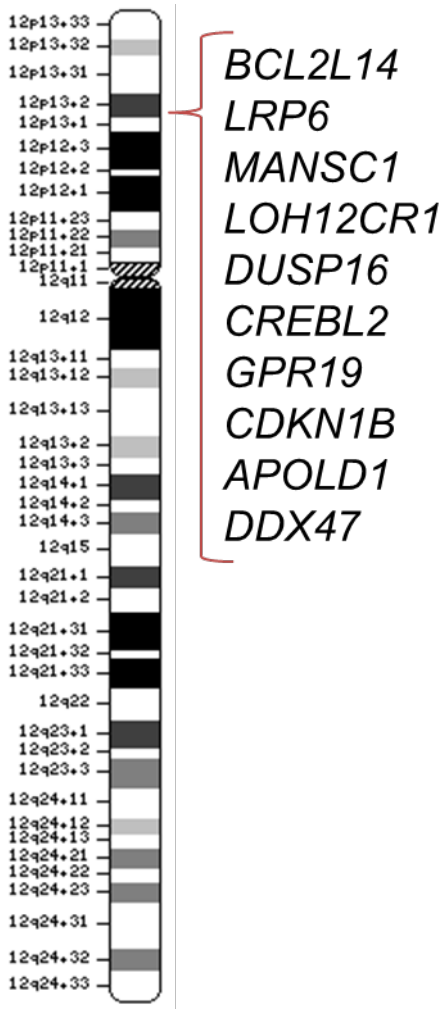


Figure 7. Minimally deleted region determined for chromosome 12p determined by SNP array analysis to derive a region of copy number alteration. The minimally deleted region contained 10 genes.

3.4.1.3 Gains and Amplifications of Chromosome 11q

The 866 patient cohort obtained from public databases was also analysed to identify gains and amplifications of the long arm of chromosome 11. Gains of regions of 11q with a copy number of 3 were found in 22 patients, the sizes of which varied greatly between patients. One patient was found to have gained an entire copy of chromosome 11, whilst two more had gains encompassing the entire 11q chromosome arm. Some small focal gains were observed, for example in sample GSM472689, a relatively small region of 0.2MB was gained, containing 9 genes. Full details of samples with gains are outlined in Supplementary Table 35.

Three patients had gains with a copy number of 4, details of which are shown in Supplementary Table 36. One of these patients, GSM850809 showed an additional gain of most of 11q on a background of what appeared to be trisomy 11.

Five patients were identified with amplifications of 11q displaying a copy number greater than 4. Four of these patients were from the complex karyotype study (Rücker *et al.*, 2012,

GSE34542). Full details of these amplifications are outlined in Supplementary Table 37. The only sample with an amplification of 11q that was not from the Rücker *et al.* (2012) complex karyotype dataset was from the Parkin *et al.* (2010) non-M3 AML study. This sample, GSM631097 showed high levels of amplification between 126,953,865 and 130,834,263 (Figure 8). This region contains a number of genes, including the potential proto-oncogenes *ETS1* and *FLI1*. There was no indication of amplification involving *MLL* or surrounding genes in this sample. Additional information was available from the annotation file for this patient: They were a 76 year old female with no history of previous AML or MDS, classified into the del (9q) prognostic group based on their cytogenetic data (MRC Intermediate risk). The SNP array data showed that the patient had a deletion of 5q and CN-LOH of the 17p region containing *TP53,* in combination with a mutation of *TP53.*

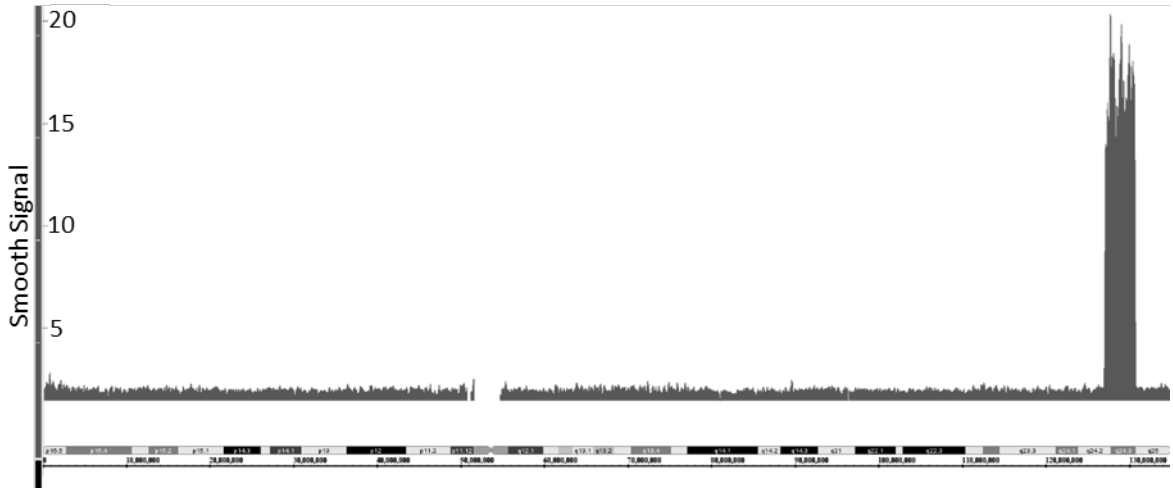


Figure 8. Sample GSM631097 showing amplification of 11q, analysed by SNP 6.0 microarray. This patient was found to have a copy number of 2 for most of chromosome 11 and a small focal region of high level amplification encompassing 11q23.

Sample GSM850743 was of particular interest due to an unusual pattern of amplification. The sample had a baseline copy number of 3 for most of the 11q arm, with a large area amplified that included *ETS1* and *FLI1* but not the *MLL* locus (Figure 9). Unusually, this sample also showed high level focal amplification in a region between 132,168,501 and 132,570,489, containing the 3' end of the *NTM* gene and the 5' end of the *OPCML* gene.

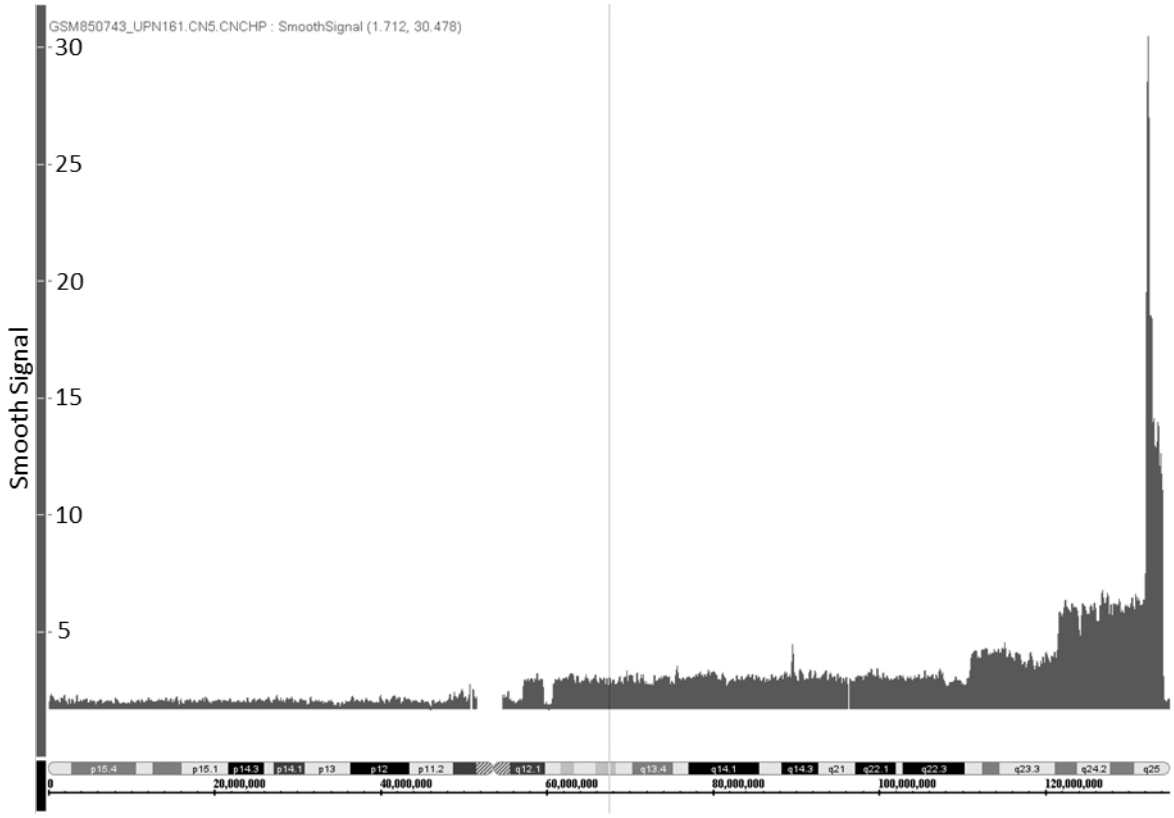


Figure 9. Sample GSM850743 displaying amplification of the long arm of chromosome 11 analysed by SNP 6.0 microarray. This sample showed a copy number of 3 for the majority of the 11q arm, with increasing stepwise copy number towards the terminal region, with a focal amplification in a small region containing parts of the NTM and OPCML genes.

The three remaining 11q amplified samples were from 500K array data in the Rücker complex karyotype dataset. GSM850721 was found to have an amplification spanning from 11q23.3 containing the *MLL* gene to the end of chromosome 11 (Figure 10A). GSM850724 had a small focal amplification containing part of *MLL*, as well as the genes: *TTC36, TMEM25, C11orf60, ARCN1, PHLDB1, TREH, DDX6, CXCR5* and *BCL9L* (Figure 10B). The final case, GSM850821, had two focal regions of amplification. The first contained the *MLL* gene, whilst the second region was distal of the first and contained the *ETS1* and *FLI1* genes, appearing to have a slightly higher copy number for this region compared to the *MLL* containing region (Figure 10C).

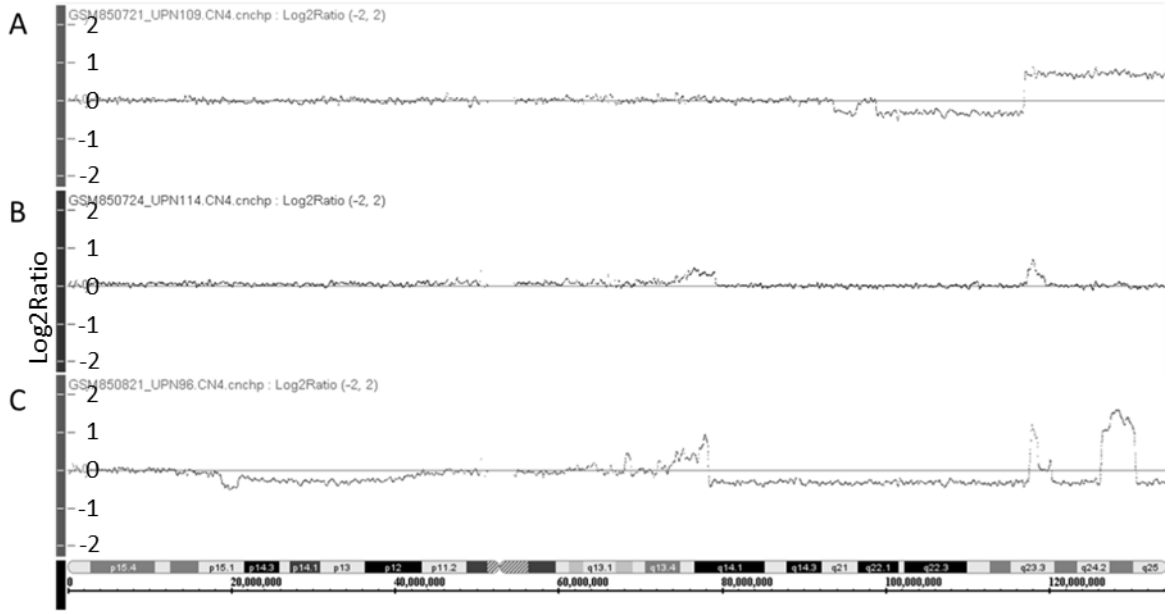


Figure 10. Samples displaying amplification of 11q analysed with 500K SNP microarrays. A. Sample GSM850721, showing amplification of the terminal portion of 11q. B. Sample GSM850724 showing a small focal region of amplification of 11q23. C. Sample GSM850821 showing two focal regions of amplification of 11q23.3 and 11q24.3.

Through the alignment of these amplifications with previously reported examples of amplification (Zatkova *et al.*, 2009), three commonly amplified regions were identified. The genes from within these regions are summarised in Figure 11.

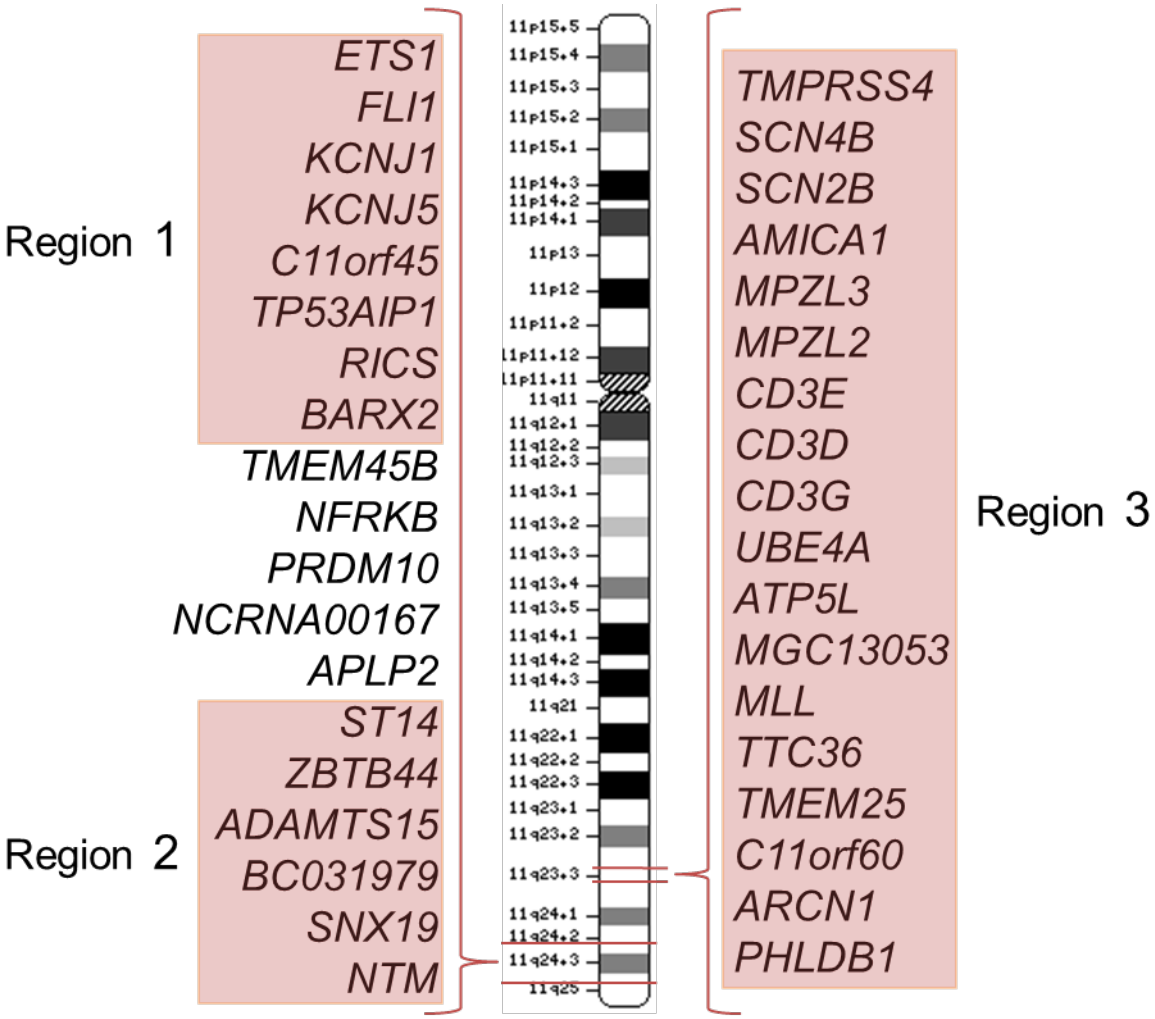


Figure 11. Commonly amplified regions on chromosome 11q derived from analysis of 500K and SNP6.0 microarray data. Three commonly amplified regions were identified, highlighted in red, containing a total of 32 genes.

3.4.2 Multiplex Ligation-dependent Probe Amplification

The MLPA technique was utilised to screen a panel of patients with normal karyotypes to further refine the minimally deleted region defined by the copy number array analysis. Normal karyotype patients were selected based on the findings of the SNP array work described in this chapter in addition to a previous study (Andreasson *et al.*, 1997), which have suggested that 12p deletions were present in some of these patients below the resolution of cytogenetic analysis. These deletions are likely to be smaller, therefore potentially more informative for the identification of a MDR.

Probe pairs were designed to specifically hybridise to genes of interest on 12p. Additionally, these probes were designed to avoid SNPs, which involved designing probes to bind to introns where SNP coverage is well known. Probe pairs were also designed to produce PCR products corresponding to a range of discrete sizes, allowing their individual resolution through capillary electrophoresis. The gene targets and probe sizes used are listed in Table 12.

Target Gene	Product Size (base pairs)
CREBL2	104
APOLD1	108
LRP6	112
CDKN1B	118
CDKN1B	126
ETV6	130
ETV6	136
ANO2	142
MANSC1	146

Table 12. Custom designed MLPA Probe sizes used for the screening of 12p deletions in normal karyotype AML DNA samples.

The probeset was mixed with the SALSA MLPA P200-A1 Human DNA Reference Probemix (MRC Holland) and reaction conditions optimised using control male and female DNA. Probe peaks were of a satisfactory height (between 2000 and 10000 units) and the reference probemix was able to correctly identify the presence of the Y chromosome in the male reference sample (Figure 12A and C), whilst also correctly identifying the presence of 2 X chromosomes in the female control sample (Figure 12B and D).

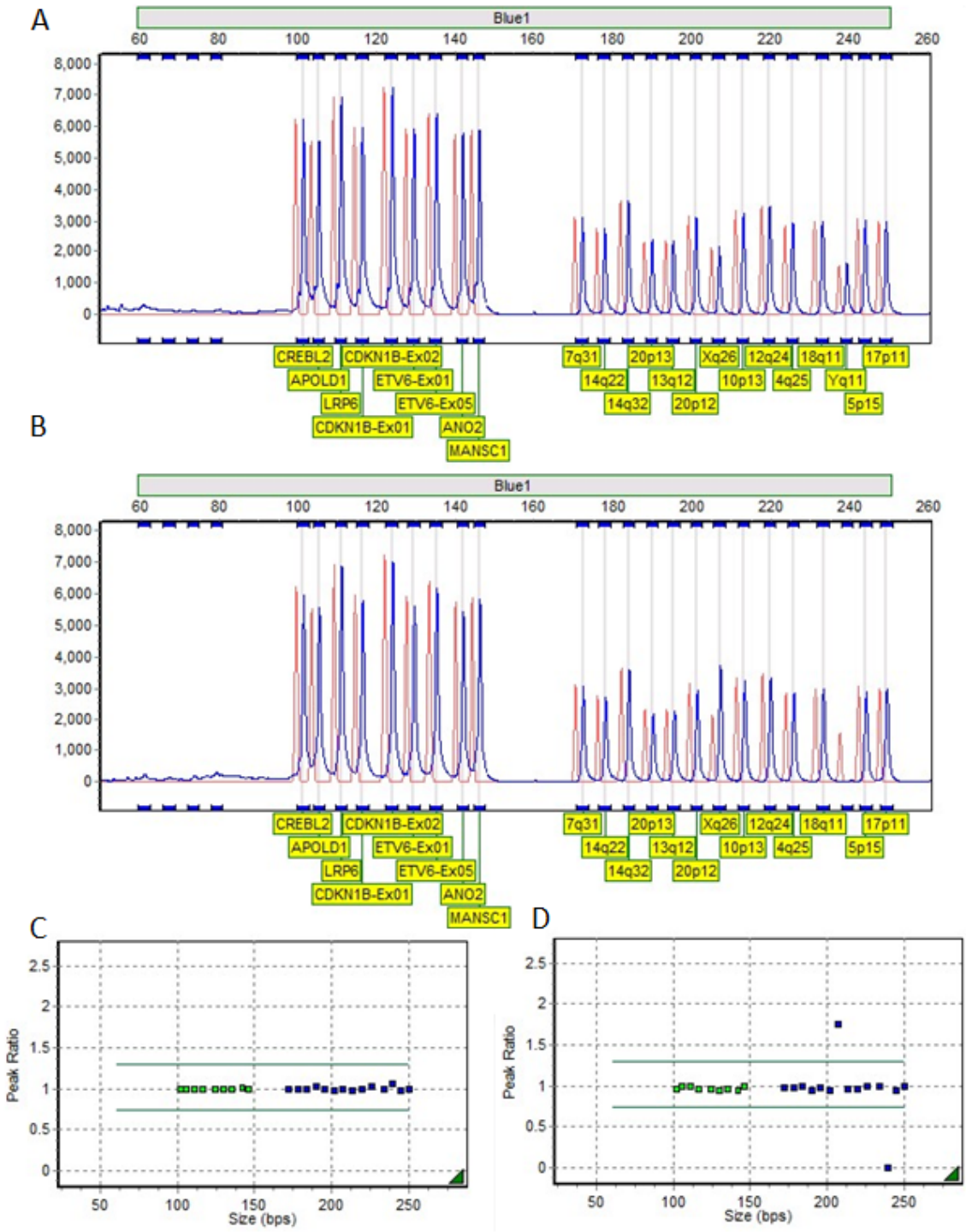


Figure 12. Male and female controls analysed using MLPA in combination with a custom designed 12p probeset. A. Peak heights for male control. B. Peak heights for female control. C. Relative peak ratio for male control sample. The green squares represent the custom 12p probes, while the blue squares represent the control probes included in the P200 kit (MRC Holland). D. Relative peak ratio for female control sample. The increased peak ratio point represents the X chromosome, whilst the value of 0 corresponds to the Y chromosome probe pair.

The 12p probeset was then tested with a selection of cell line DNA. The cell lines KG-1 and GDM-1 were selected as these cell lines have a monoallelic deletion of 12p. Additionally, a patient sample with a 12p deletion detected by karyotyping was included (8070).

The probes used were able to correctly identify deletions in the 12p deleted cell lines (Figure 13A and B). The retained probe at 142 base pairs in both samples represents the *ANO2* gene, selected for its proximity to the 12p telomere. As it was retained in these samples, it confirmed that the deletion of 12p in these cell lines was interstitial. Furthermore, the reference probes confirmed the gain of chromosome 13 in the GDM-1 cell line, known from karyotyping and SNP array data. Interstitial deletion of 12p was also identified in patient 8070 (Figure 13C). The quality control fragments (visible in the top left) were amplified in this patient sample suggesting that the DNA was poor quality, however the expected result was still visible.

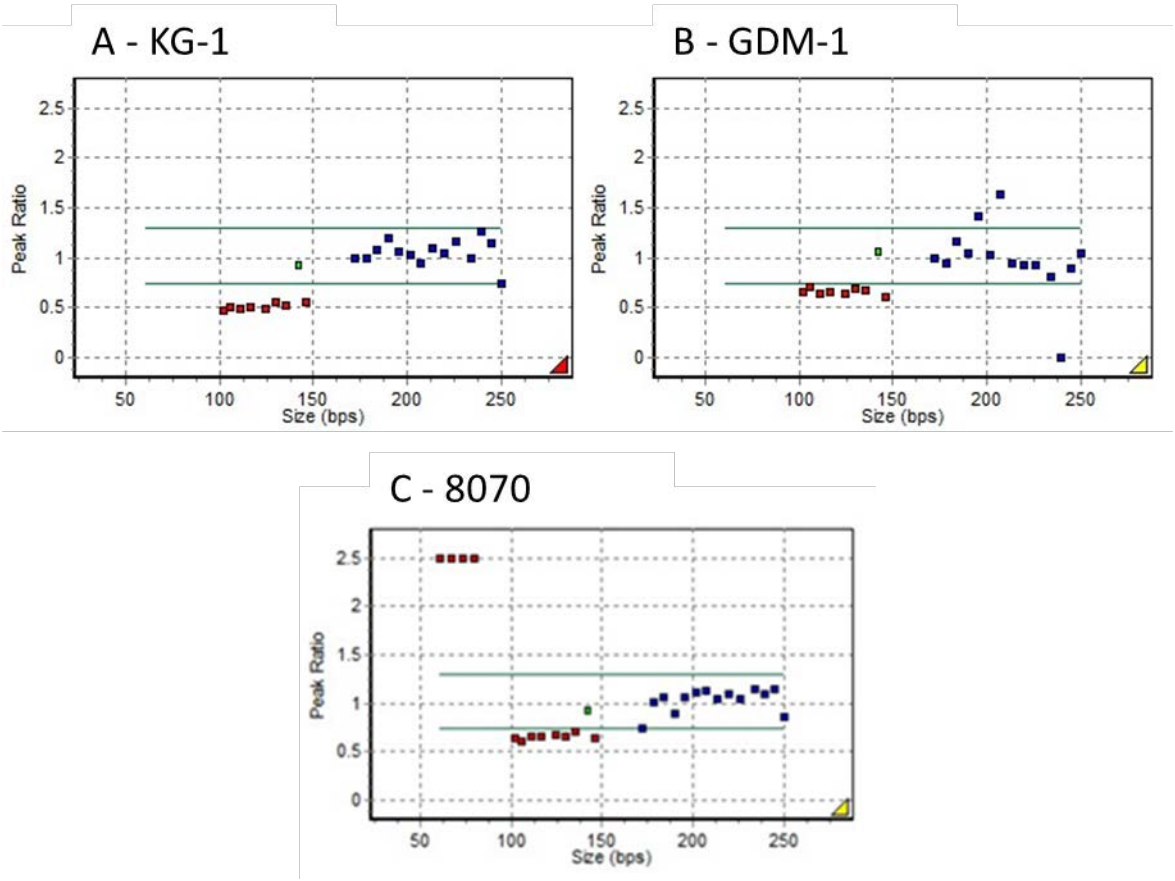


Figure 13. Control samples analysed using MLPA in combination with a custom designed 12p probeset. A. KG-1 cell line with 12p deletion. B. GDM-1 cell line with 12p deletion. C. Patient 8070 with 12p deletion detected by karyotyping. The red squares show probes from the 12p kit with a signal ratio below 0.75, and were therefore determined to be deleted. The retained probe in this area represents the 12p terminal probe for the ANO2 gene, suggesting an interstitial deletion of 12p in this cell line.

After optimisation, the probeset was then used to screen normal karyotype AML patients for the presence of 12p deletions. Two examples of patients with identified deletions of 12p are shown in Figure 14. Among 75 patients, five were found to have deletions of 12p (6.7%). Two of these patients had deletions that included the *ANO2* gene close to the 12p telomere, suggesting a large deletion of the p arm of at least 7MB in size. Fixed cell nuclei were available for four of these five patients, and FISH using a probe targeting *ETV6* within the deleted region

confirmed 12p deletions. These deletions were shown to be at least the same size as the minimally deleted region established from the SNP array data, so it was not possible to further refine the 12p candidate gene list from these data.

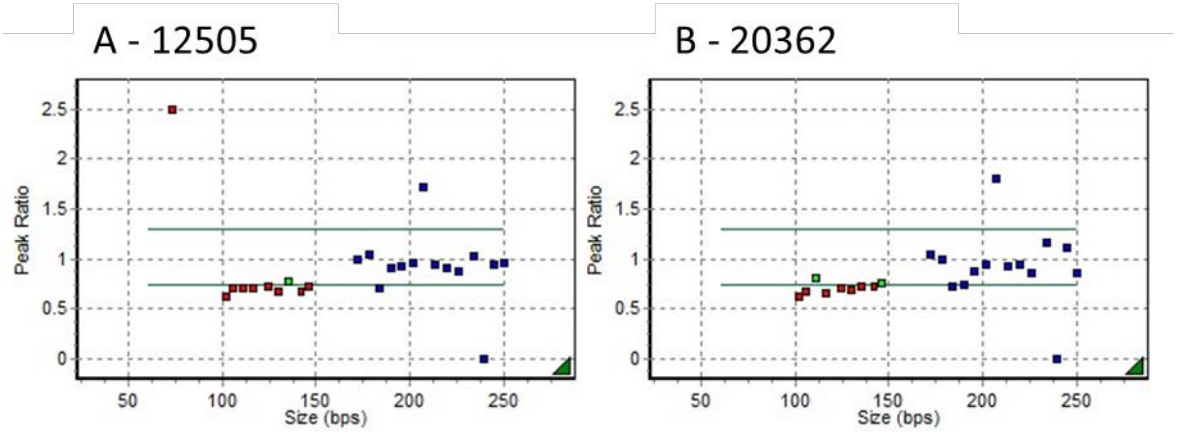


Figure 14. Normal karyotype patients analysed with the 12p MLPA kit, found to have deletions of 12p through MLPA screening. These are shown by the red squares on the left hand side of the peak ratio plots, which sit below the 0.75 threshold.

3.4.3 Cell Line Selection

3.4.3.1 12p Deleted AML Cell Lines

The Sanger COSMIC Cell Line database was screened for publicly available SNP 6.0 data to find deletions of chromosome arm 12p in AML cell lines. Three suitable cell lines were identified: GDM-1, KG-1 and NKM-1. These lines were selected as they had interstitial deletions of 12p including all genes established through SNP array analysis, in conjunction with relatively few other genetic aberrations.

Cell Line	AML Subtype	FAB Subtype	Gender	Age (Years)	Reference
GDM-1	Myelomonoblastic	M4	F	66	Ben-Bassat *et al.*, 1982
KG-1	Myeloblastic	M1	M	59	Furley *et al.*, 1986
NKM-1	Myeloblastic with granulocytic maturation	M2	M	33	Kataoka *et al.*, 1990

Table 13. AML Cell Lines with deletion of chromosome arm 12p, found using the Sanger COSMIC Cell Line database.

3.4.3.2 11q Amplified AML Cell Lines

Three AML cell lines reported to have amplification of the *MLL* gene have been described in the literature.

Cell Line	AML Subtype	FAB Subtype	Gender	Age (Years)	Reference
2L1	Myeloblastic	M5	M	6	Felix *et al.* 1998
UoC-M1	Megakaryocytic	M7	M	68	Allen *et al.* 1998
SAML-2	Myelomonocytic	M4	M	42	Knutsen *et al.* 2003

Table 14. AML Cell Lines with amplification of chromosome 11q, found through literature review.

The cell line 2L1 was established from the bone marrow of a child who developed t-AML after treatment for neuroma and rhabdomyosarcoma. In this cell line, multiple signals were detected with a probe targeted to the *MLL* gene on a der(5)t(5;11) chromosome, as well as on a der(11) chromosome (Felix *et al.*, 1998). This cell line is no longer available, due to contamination of cell stocks (Correspondence with Dr. Peter Aplan, National Cancer Institute, USA). The cell line SAML-2 was also established from a t-AML, in this case from a patient receiving treatment for Hodgkin disease. This cell line has two der(8) chromosomes, formed from translocations between chromosomes 8 and 11, resulting in both *MYC* and *MLL* amplification (Allen *et al.*, 1998). There has been some difficulty in culturing this cell line, possibly due to its slow growth and fragile nature. This has resulted in extremely low stocks of the line remaining worldwide and as such is not available (Correspondence with Dr. Turid Knutsen, National Cancer Institute, USA and Prof. Hans Drexler, DSMZ, Germany). The UoC-M1 cell line was established at the University of Chicago in 1998 from a patient with AML-M1. The cell line carries a dicentric chromosome 11, containing alternating segments of chromosomes 9 and 11, with a total of four copies of *MLL* present (Allen *et al.*, 1998; Knutsen *et al.*, 2003). Although the presence of four copies is not regarded by all as true amplification (Myllykangas *et al.*, 2006) the expression of genes from the 11q region has been shown to be increased in this cell line relative to other AML cell lines (Allen *et al.*, 1998), therefore it is a suitable model for amplifications of 11q.

The UoC-M1 cell line was received as a kind gift from Michelle LeBeau, University of Chicago. As there were no publicly available SNP array data for the line, SNP 6.0 and FISH analyses were

performed to gain more information about the genetic aberrations present. FISH using a *MLL* probe confirmed the presence of 4 copies of this gene, as shown in Figure 15. The SNP 6.0 data also showed 4 copies of this region, stretching from 11q22.3 to the 11q terminus, as shown in Figure 16. Interestingly, the SNP analysis also showed a number of aberrations of chromosome 21 that had not previously been reported in this cell line. These data showed amplification of multiple genes on chromosome 21, including the putative oncogenes *ERG* and *ETS2*. Another putative oncogene, *RUNX1* was not amplified to a high level, although there were 4 copies. The pattern of amplification on chromosome 21 is shown in Figure 17. Multiple copies of *ERG* were confirmed by FISH (Figure 18).

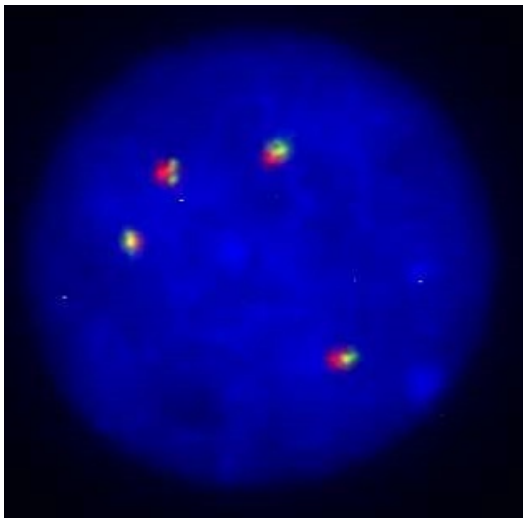


Figure 15. FISH performed on the UoC-M1 cell line using a MLL break-apart probe. As four fusion signals are observed, it is inferred that four copies of MLL are present in the cell line.

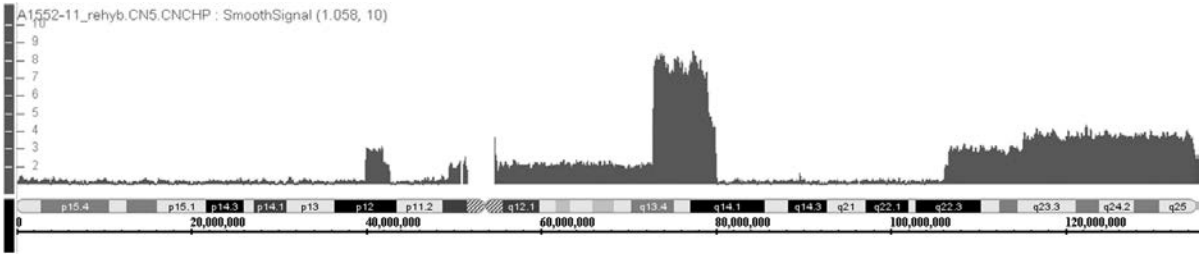


Figure 16. SNP 6.0 Array data from UoC-M1 Chromosome 11. The terminal region containing the minimally amplified regions found earlier in the project has a copy number of 4.

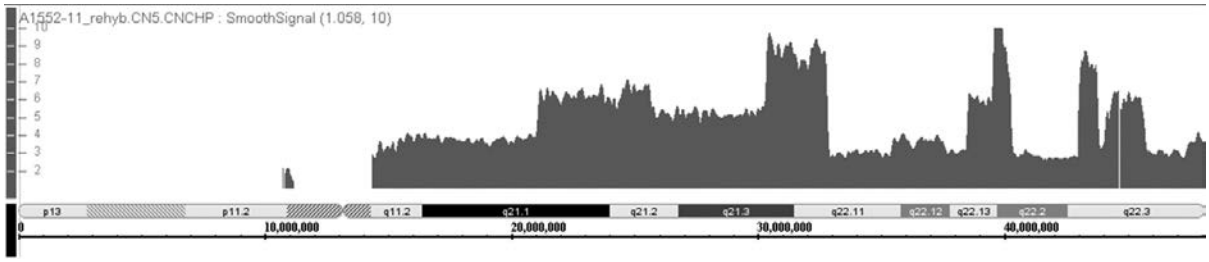


Figure 17. SNP 6.0 Array data from UoC-M1 Chromosome 21, showing variable copy number across the long arm of the chromosome, with some regions of amplification.

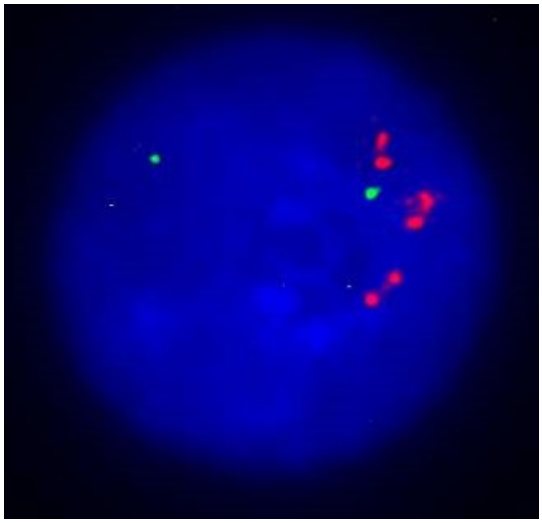


Figure 18. FISH performed on the UoC-M1 cell line using probes for ERG (Red) and a Chromosome 21q22.13 probe (Green). The presence of multiple red signals signifies amplification of the chromosome region containing ERG.

3.5 Discussion

As discussed in Chapter 1, deletions of 12p and amplifications of 11q are recurrent cytogenetic abnormalities in AML that correlate with a poor prognosis. The aim of this chapter was to establish minimal regions of deletion and amplification that could be used to produce a candidate gene list for further investigation in subsequent work. The availability of public databases were instrumental in completing this line of work, and such a study was only possible due to the large amount of studies that have submitted their data to GEO, as well as the Sanger database of copy number alterations and mutations in cancer cell lines.

3.5.1 Deletions of Chromosome 12p

Deletions of 12p were detected in 2 of the 111 cases (1.8%) in the Radtke *et al.* dataset, which contained only paediatric cases, in line with previously reported incidences of this abnormality in this age group (Harrison *et al.*, 2010). In the absence of specific age information for the other studies, only the overall incidence across all samples can be studied. The 5.2% incidence (excluding the selected 12p deleted cases) is as expected taking into account the 1-5% previously reported (Grimwade *et al.*, 2010; Harrison *et al.*, 2010). The detection rate may have been slightly higher than traditional cytogenetic techniques due to the use of sensitive SNP array technology, as it has been previously shown that deletions of 12p may be below the resolution of cytogenetic techniques. Andreasson *et al.* (1997) reported that up to 10% of 12p deletions were missed by cytogenetics but were subsequently detected using specific FISH probes (Andreasson *et al.*, 1997). In this study, 5.1% of patients across the two studies focussing specifically on normal karyotype patients had deletions of 12p, lending further support to this notion.

The MLPA probeset developed for this section of work was shown to be an accurate method of screening patient DNA for copy number aberrations of the 12p region. Four of the five normal karyotype cases shown by MLPA to have deletions of 12p were successfully validated using FISH. Although the MLPA was unable to identify cases to further refine the minimally deleted region established by the SNP array study, it provided information about 12p deletions in normal karyotype patients: 6.7% of these patients were found to have a deletion of this region. This portion of the study may have benefitted from a larger cohort of patient material to screen, however this was not available at the time. Future studies utilising these techniques could easily be performed as the probeset has now been designed, optimised and validated.

The minimally deleted 12p region contains 10 genes (*BCL2L14*, *LRP6*, *MANSC1*, *LOH12CR1*, *DUSP16*, *CREBL2*, *GPR19*, *CDKN1B*, *APOLD1* and *DDX47*). *ETV6* was excluded based upon copy number data obtained from two patients. One of these patients was from the previously

discussed Haferlach *et al.* (2011) study, and the other from the Rücker *et al.* (2012) study. Despite these findings, it was decided to include *ETV6* in the candidate gene list for functional follow up. As only copy number SNP arrays were used to determine the MDR, it is unknown if other chromosomal rearrangements that do not affect copy number may alter the expression of *ETV6*. The 12p13 chromosomal region has been shown to have an increased degree of instability, with multiple break points possible that lead to complex rearrangements (Sato *et al.*, 2001). It would be ideal to further assess the prevalence of rearrangements in this region using sequence data obtained from AML patients both with and without deletions of 12p, that have been analysed on high throughput sequencing platforms. Analysis in this way would have the added benefit of providing mutation data which may be more informative in implicating certain genes in a tumour suppressor role.

In addition to this lack of genomic information, there is a body of evidence implicating *ETV6* in having a functional tumour suppressor role. *ETV6* can form fusion genes when translocated in AML, but these often do not result in a functional transcript. Instead it has been suggested that they lead to altered wild type *ETV6* dosage through haploinsufficiency (Silva *et al.*, 2008). Mutations in *ETV6* have also been reported in AML, some of which are believed to have dominant negative effects disrupting *ETV6* function in the absence of a second hit (van Doorn-Khosrovani *et al.*, 2005). Finally, research by Feurstein *et al.* (2014) has found that *ETV6* expression is effectively halved in patients with deletions of 12p13 compared to those without 12p abnormalities, providing evidence that decreased dosage of this gene may be a driver of leukaemogenesis in these patients (Feurstein *et al.*, 2014).

Three AML cell lines have been identified with deletions of 12p that will be utilised in future work – GDM-1, KG-1 and NKM-1. These cell lines were selected as they carry interstitial deletions encompassing the MDR, and have relatively few other chromosomal aberrations (Ben-Bassat *et al.*, 1982; Furley *et al.*, 1986; Kataoka *et al.*, 1990; Catalogue of Somatic Mutations in Cancer).

3.5.2 Amplifications of Chromosome 11q

Overall, 31 patients (3.6%) had gain of material on chromosome 11q. From the complex karyotype group, 12 of 102 patients (11.8%) were found to have gains. This is a little lower than the previously reported 15-40% of patients in the complex karyotype group (Schoch *et al.*, 2002; Rücker *et al.*, 2006).

Twenty four patients had gains that resulted in a copy number of three or four on 11q. Of these, 15 had large regions of gain that included *MLL*, *ETS1* and *FLI1*. A further four patients had gains including *MLL* but not *ETS1* and *FLI1*, whilst 3 patients had gains of *ETS1* and *FLI1* but

not *MLL*. Finally, two patients had focal gains located outside the established common regions of gain. As these two focal regions were not recurrent and have not previously been reported, it is likely that they are passenger mutations that do not result in a leukaemogenic effect.

Of the five patients identified with amplifications of 11q, four of them also had a complex karyotype. As most amplifications of 11q are found in patients with complex karyotypes, this observation was expected. The overall incidence of 11q amplifications across all patients sampled was low (0.6%), although the large numbers of normal karyotype and paediatric AML samples used in this study is likely to have contributed as 11q amplifications are rare in these groups.

Through combining array data with previously published data, three common regions of amplification have been identified that will be investigated in the next stages of the project. These regions contain a total of 30 genes, including the putative proto-oncogenes *MLL*, *ETS1* and *FLI1*. For functional investigation, the cell line UoC-M1 will be used. Although not carrying true amplification of the region of interest, the cell line has four copies of the candidate genes, and it has been demonstrated to have significantly increased expression of genes from this region when compared to control samples (Allen *et al.*, 1998).

3.5.3 Amplifications of Chromosome 21q

Although not the focus of this project, an interesting observation was the amplification of the long arm of chromosome 21 found in the UoC-M1 cell line. Amplifications of chromosome 21 have previously been reported in AML, with one study finding amplifications of this region to be more common than amplifications of chromosome 8q, but less frequent than amplifications of chromosome 11q (Roller *et al.*, 2013). Another study has linked amplification of *APP, ETS2* and *ERG* from within the chromosome 21q region with an increased level of expression (Baldus *et al.*, 2004), suggesting a functional relevance for these genes. The amplified region identified in the UoC-M1 cell line overlaps with the amplified regions identified in these two studies, although a separate study has suggested that the putative oncogenes *ERG* and *ETS2* are not the target of chromosome 21 amplification (Canzonetta *et al.*, 2011). This conflicting evidence combined with the observation of recurrent amplifications of this region make future functional investigations of this region an interesting prospect that would be facilitated using the UoC-M1 cell line.

Chapter 4 Identification of Driver Genes by Negative Selection

4.1 Introduction

4.1.1 HIV Lentivectors

As the cause of one of the most prevalent diseases affecting modern humans, human immunodeficiency virus (HIV) has been continuously investigated over the last three decades. This research has provided a great depth of knowledge about a number of aspects of HIV, including its genome, particle structure and viral life cycle. As a member of the retrovirus family, HIV has the ability to integrate its genome with that of its host cell, resulting in long term infection and continuous transcription of viral products. Furthermore, HIV is able to replicate in slowly dividing and non-dividing cells, hence its classification as a lentivirus. This unique set of attributes has resulted in significant interest in developing HIV based gene delivery vector systems for research purposes.

4.1.1.1 The HIV Life Cycle

The first step in the HIV life cycle involves the entry of the viral particle into the target host cell, which is mediated through the interaction of HIV envelope proteins (Env) with the surface membrane of the target cell (Gomez and Hope, 2005). This interaction leads to the formation of a six helix protein bundle that fuses the virion membrane with the target cell membrane, allowing delivery of the viral core to the target cell cytoplasm (Melikyan *et al.*, 2000).

The viral core is made up of proteins cleaved from the polyprotein transcribed from the viral *gag* (Group specific antigen) gene transcripts. When the viral core is exposed to the host cell's cytoplasm, the viral RNA genome is released from its capsid (Arhel, 2010). Reverse transcription of the genome is mediated by reverse transcriptase encoded by the HIV *pol* gene, at first forming an RNA-DNA complex, from which the RNA strand is degraded and a double stranded DNA copy of the viral genome is formed (Moore *et al.*, 2009; Simon-Loriere *et al.*, 2011). Integrase, another product of the *pol* gene, then integrates the double stranded DNA helix into the host cell genome. The integrase enzyme first acts by cleaving both 3' ends of the DNA molecule, creating sticky ends amenable to integration (Fanales-Belasio *et al.*, 2010). It then transfers the DNA into the host cell nucleus and facilitates its integration into the genome.

After integration with the host genome, synthesis of HIV transcripts and their translation to proteins occurs, allowing new viral particles to be produced which push through the host cell membrane, budding into complete virions that can go on to infect other cells (Gomez and Hope, 2005). The entire HIV replication process is summarised in Figure 19, and a schematic of the complete virus particle is shown in Figure 20.

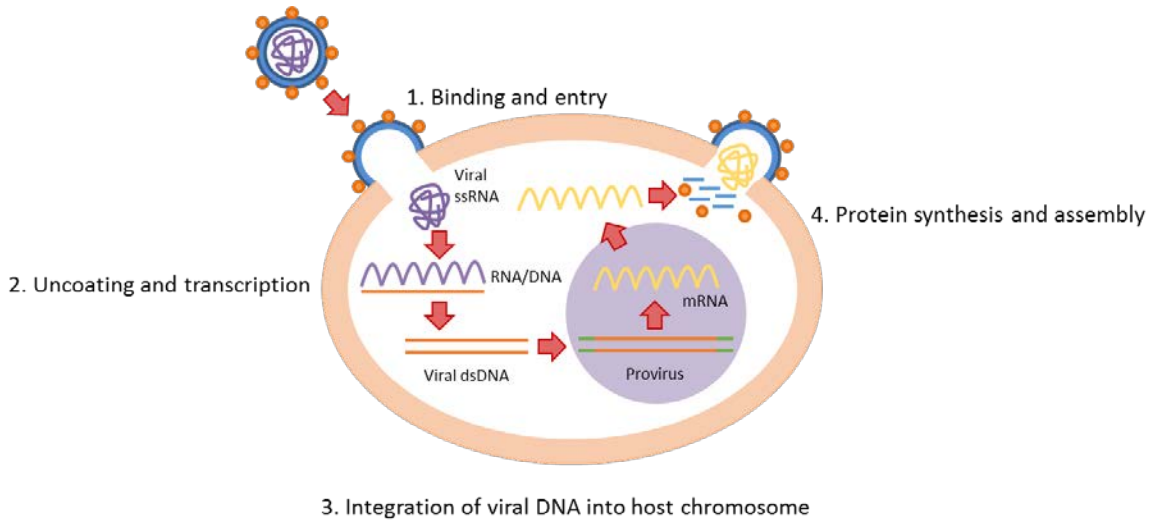


Figure 19. The HIV Life Cycle. After binding the host cell membrane, the viral core is exposed and uncoated before reverse transcription into double stranded DNA takes place. Integrase mediates the integration of viral DNA into the host cell genome, where viral mRNA sequences are transcribed leading to the production of new viral proteins. These proteins are assembled and complete viral particles push through the host cell membrane and bud, forming complete virions capable of infecting additional host cells (Rambaut et al., 2004; Engelman and Cherepanov, 2012).

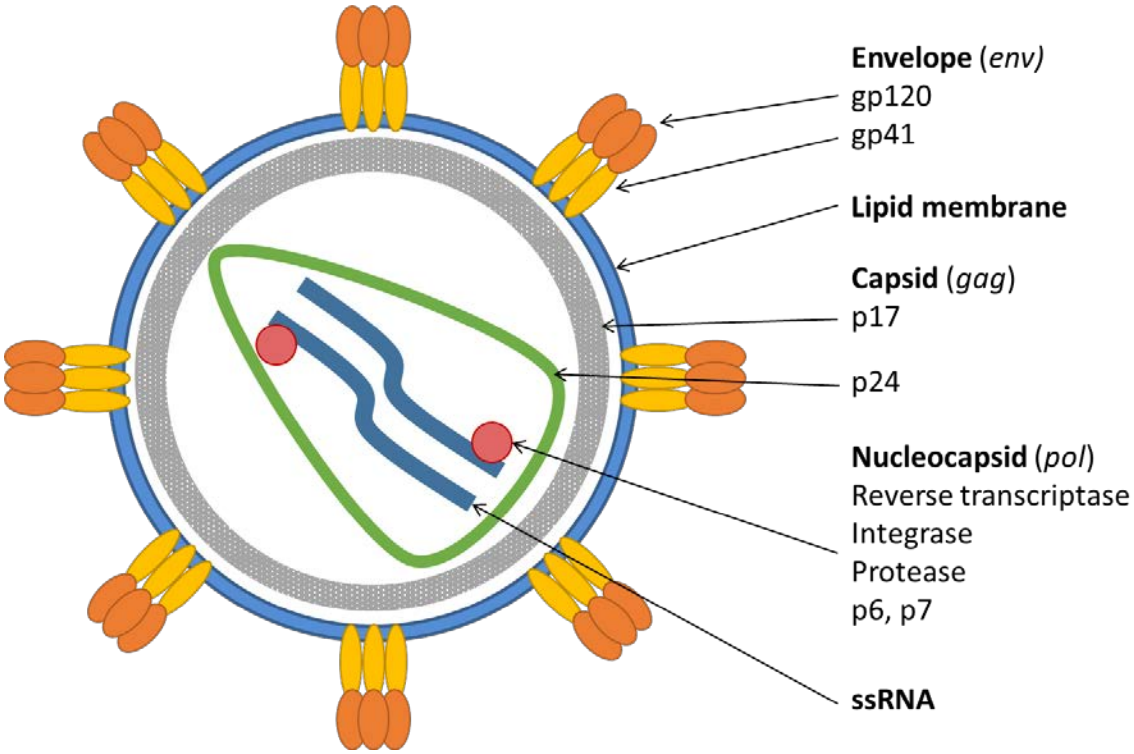


Figure 20. The Fully Assembled HIV Particle. The main structures are listed in bold, along with the gene that encodes the related proteins, listed underneath. The viral genome is encased within a core made up of products of the gag (group specific antigen) gene. Products of the pol (polymerase) gene are also present, providing the enzymes required to integrate into new host cells. The products of the env (envelope) gene, provide the surface proteins responsible for binding target cells. The lipid membrane is provided by the host cell when the completed viral particle pushes through the host membrane in the budding process (Robinson, 2002; Fanales-Belasio et al., 2010).

4.1.1.2 Using HIV as a Gene Transfer Vector

The primary concern when considering the use of HIV to deliver gene expression constructs is safety. This is achieved by ensuring that replication competent virus particles cannot be produced, but are still able to achieve the goal of integrating constructs capable of driving long term expression of target genes. With this goal in mind, Naldini *et al.* (1996) designed a HIV packaging system consisting of three plasmids. The genes required to produce functional lentiviral particles are transiently expressed by plasmids in order to produce viral particles that are not themselves capable of reproducing, but are able to integrate a viral sequence containing one or more genes to be expressed in the target cell (Naldini *et al.*, 1996).

The packaging construct contains the viral *gag* and *pol* genes, with the regulatory genes *tat* and *rev*, as well as the HIV accessory genes *vif, vpr* and *nef*. All of these genes are under the control of a cytomegalovirus (CMV) promoter to induce gene expression in a wide range of cell types. More significantly, the *env* gene is disrupted, with the RRE (*rev* response element) from within the gene remaining as the only functional *env* feature. Retention of the RRE ensures exportation of full length mRNAs coding for Gag and Pol proteins from the packaging cell nucleus (Naldini *et al.*, 1996).

As the *env* gene is disrupted on the packaging plasmid, without another source of envelope proteins provided, it would not be possible to produce mature lentiviral particles. The envelope plasmid instead contains the *vsv-g* gene, again under the control of a CMV promoter. This drives the production of the Vesicular Stomatitis Virus (VSV) Glycoprotein-G coat protein, which is used instead of the HIV Env proteins to produce a pseudotyped virus consisting of the HIV core surrounded by a VSV-G coat. Viral particles pseudotyped in this way are no longer specific to one cell type (for example CD4+ cells as with HIV viruses) and are instead able to transduce all cell types (Naldini *et al.*, 1996). For other applications, such as gene therapy vectors, viral coat proteins from other viruses can be substituted to create pseudotyped viruses that are specific to certain cell types.

The transfer vector provides the sequences to be integrated into the target cell genome. By including truncated forms of the *gag* and *env* genes, two copies of full length mRNAs transcribed from this plasmid are able to form complexes with replication enzymes and form the nucleocapsid structure of mature viral particles. The RRE from within *env* ensures nuclear exportation, and sequences within *gag* ensure correct localisation to the viral membrane (Gomez and Hope, 2005). The transfer vector can include full length cDNA sequences in combination with a promoter of choice, to drive gene expression in infected cells where the transfer cassette is successfully integrated.

This combination of packaging plasmids is transfected into the 293T cell line, or one of its derivatives. The 293T line was derived from 293 cells through the introduction of the SV40 large T-antigen, and allows the episomal replication of transfected plasmids that contain the SV40 origin of replication which improves lentiviral particle production (Oka *et al.*, 2010). The SV40 large T-antigen also inhibits phosphorylated Rb protein, leading to an increased growth rate in 293T cells compared to 293 which improves their ease of use in culturing and the production of lentiviral particles (Meng *et al.*, 2010).

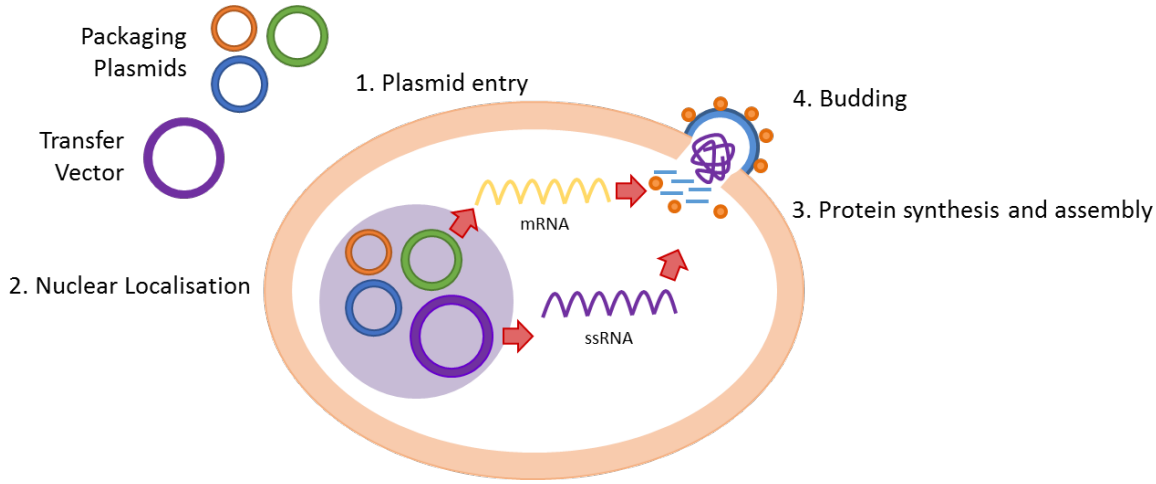


Figure 21. Production of lentiviral particles in the 293T cell line. Plasmids are transfected through the endocytosis of cationic liposomes, cationic polymers or phagocytosis of calcium phosphate precipitates. Packaging plasmids encode for viral protein components, whilst the transfer vector provides the ssRNA to be contained within the lentiviral particle for delivery into target cells.

This basic three plasmid system has subsequently been modified to further reduce the risk of producing replication competent retrovirus particles. In the second generation of these vectors described by Zufferey *et al.* (1997), the packaging construct was reduced by removing the accessory genes *vif, vpr* and *nef*, which were found to be surplus to requirements for viral production *in vitro*. In the third generation of vectors, *tat* is replaced by constitutively active promoter sequences in the packaging construct. In addition to this, the third generation of vectors have a self-inactivating (SIN) deletion of the 3' Long Terminal Repeat (LTR). This deletion is transferred to the 5' LTR of integrated proviral DNA, and abolishes viral promoter activity (Zufferey *et al.*, 1998), preventing the synthesis of full length vector RNA in target cells. This further ensures the safety of produced lentiviral particles and has the added effect of ensuring that host genes near to the integration site are not aberrantly expressed as a result of viral promoter activity.

A number of other features can be included to tailor the produced lentivirus particles for specific applications. For example in this project, transfer vectors based on the SIN-SIEW vector will be utilised. In this transfer vector, the internal spleen focus forming virus (SFFV)

promoter is selected to drive the expression of integrated cDNAs, for its high activity in haematopoietic cell types, including haematopoietic stem cells (Demaison *et al.*, 2002). Other elements such as the Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) further enhance expression and increase the amount of produced lentivirus. The inclusion of an internal ribosome entry site (IRES) allows for translation initiation in the middle of transcribed mRNA sequences (Hellen and Sarnow, 2001), and can therefore drive the translation of a second coding sequence present on the construct. In regards to this project, this means that enhanced green fluorescent protein (eGFP) can be expressed alongside a cDNA of choice from a single integrated construct.

4.1.2 Murine *In Vivo* Screening Techniques

In vivo screening techniques have been incredibly useful in determining the functions of many genes. Multiple different approaches have been taken across a wide range of species to elucidate the underlying mutations causing varying phenotypes. The earliest studies focussed on lower organisms such as *E. coli*, *C. elegans*, and *D. melanogaster*, and used varied genetic techniques to determine whether underlying genetic mutations causing certain observed phenotypes were caused by either natural or induced mutagenesis (Justice, 2000). Whilst these proved to be successful in determining the function of some human genes and their roles in disease, the evolutionary span between humans and these species is often so large that useful information is lost when bridging the genetic distance between model organisms.

Mice have been shown to have an extremely large number of genes (over 90% sequence homology) in common with humans (Mouse Genome Sequencing Consortium, 2002), as well as sharing similar physiology. Advances in genetic techniques and the completion of the mouse genome project (Waterston *et al.*, 2002) now means that murine models are highly suitable for application in the study of human diseases.

4.1.2.1 Early Insertional Mutagenesis Screening

Early murine screen approaches were focussed on chemical mutagenesis, in particular N-ethyl-N-nitrosourea (ENU), pioneered by Bill Russell of the Oak Ridge National Laboratories (Russell *et al.*, 1979). These were coupled with positional cloning to identify the phenotype produced by exposure to the chemical mutagen (Justice, 2000).

More recent advances have moved on from chemical mutagenesis, and approaches such as the Sleeping Beauty (SB) transposon system have been developed. The SB system is based on a synthetic transposon and transposase created from studying the phylogeny of inactive Tc1/mariner-type transposases isolated from a number of fish species. For its use in *in vivo* screening, the SB system is based upon the use of two transgenic mouse lines. One of these

lines contains the SB DNA transposase vector, while the other line carries the SB transposon transgene. When these two lines are crossed, the transposon is mobilised, allowing its integration elsewhere in the mouse genome. Should the transposon relocate within a gene, the production of functional protein from that gene locus is disrupted. O'Donnell *et al.* (2012) used this to identify potential tumour suppressor genes contributing to MYC-driven liver cancer. In this study, mice carrying the SB transposon vector and SB transposase inserts were crossed with an existing model of MYC-induced liver cancer, in which mice overexpress MYC in response to tetracycline within their liver cells. In this case, the SB transposon contained a gene trap element (T2/Onc), allowing identification of integration sites easily. They found that in those mice containing the SB transposase and transposon, liver tumour development was significantly accelerated and they identified 16 disease associated genes using this method (O'Donnell *et al.*, 2012).

4.1.2.2 Retroviral Based Insertional Mutagenesis

Retroviral based insertional mutagenesis screens have primarily been used to identify oncogenic mutations in an *in vivo* setting. These systems typically make use of slowly transforming retroviruses, which induce tumours in mice with a latency of between 3 and 12 months (Uren *et al.*, 2005). The mutation of different genes is facilitated by the integration of proviral elements, which themselves drive and regulate the transcription of sequences required to produce further retroviral particles, infecting more cells. The integration of the proviral sequences can lead to the dysregulation of nearby genes with a number of differing effects. These include the activation of proto-oncogenes or the disruption of tumour suppressor genes, which then confer a selective growth advantage to that particular cell, encouraging clonal outgrowth (Uren *et al.*, 2005).

To identify genes driving the accelerated clonal expansion of cells, viral insertion sites are mapped to their location on the genome. Where insertion sites are found to occur in a particular region at a rate higher than expected by chance, the nearby genes are implicated in providing the growth advantage held by these cells. Within a single clonal tumour there are often three or four insertions near different oncogenes and tumour suppressor genes (Uren *et al.*, 2005; Ranzani *et al.*, 2013), although one recent screen identified an average of 20 insertions per gene (Uren *et al.*, 2008).

4.1.2.3 Gene Trapping

Gene trapping is a technique that has been applied to generate many knockout stores of mouse embryonic stem cells, and has been the most widely used technique applied to produce a library of embryonic stem cell lines, each lacking a different gene of the mouse genome. The International Knockout Mouse Consortium (IKMC) is the umbrella organisation encompassing multiple smaller organisations with the collective aim of completing this mouse embryonic stem cell library. As of 2011, the consortium had achieved knockout of around 17,000 genes (Dolgin, 2011). As with ENU mutagenized ES cells, the stored cells can be thawed and injected into blastocysts to establish a mouse line carrying the particular conditional knockout of interest (Limaye *et al.*, 2009).

Gene-trap vectors are functionally similar to the other mutagenesis techniques discussed above, and are introduced on to a cell by either retroviral delivery or electroporation. Upon integration, they lead to the disruption of endogenous gene expression, as well as expression of a reporter gene. The advantage of gene trap vectors is that they contain a splice acceptor site at the 5'-end of the reporter gene, which allows a fusion transcript between the reporter and endogenous gene to be formed (Lee *et al.*, 2007). By using a gene trap vector to disrupt the expression of a gene, the integration site can be quickly identified using 5' or 3' rapid amplification of cDNA ends (RACE). This allows gene-trap vectors to be used in high throughput approaches to generate knockout libraries. Although many different vectors have been used in combination with both delivery methods, a small subset of genes have been difficult to target with this method, because integration is biased, with some sites being favoured over others (Stanford *et al.*, 2001).

4.1.3 Functional Screening

Functional screening using murine systems typically involves the introduction of genetic material through the use of retroviral systems to drive the production of various genetic molecules, including shRNAs, cDNAs and miRNAs. The wide variety of vectors and target cells available to researchers provide an extremely flexible set of systems, a subsection of which will be discussed here.

Functional screens have been extremely useful in identifying genes important in a range of cancers. A number of considerations have to be made when designing a functional screen to identify putative oncogenes and tumour suppressors. The first concerns delivery of the vector. shRNAs are useful for a number of reasons. They are strong activators of the RNAi machinery within the host cell able to efficiently repress gene expression even when expressed from only a single genomic copy (Fellmann *et al.*, 2013). The widespread availability of shRNAs already

integrated into lentiviral constructs and available in library formats further facilities their ease of use in large scale screening projects. cDNA screens can require more effort in terms of preparing lentiviral constructs containing the cDNA of choice, but can have the advantage of providing a more informative representation of the functions of a particular gene. miRNA screens have been useful in uncovering gene functions, as they have the potential to interact with a number of sites across the genome and have implicated previously unrelated genes in certain cellular functions (Hausser and Zavolan, 2014).

The screening process can be applied to a range of cell types, often determined by the model being studied. A recent study combined RNAi screening techniques using shRNAs with a previously well-defined mouse model of liver cancer, in which hepatocellular carcinomas were induced by genetically manipulating cultured liver progenitor cells before transplanting them into the livers of recipient mice (Zender *et al.*, 2008). To this already established model, a pool of 48 shRNAs, targeting various genes from within commonly deleted regions of liver carcinoma, were introduced. By quantifying the shRNAs in carcinoma samples a number of potential tumour suppressor genes were identified.

In some screens, the target cell has been the limiting factor to the size of the pool being investigated. For example, a study by Adams *et al.* (2012) investigating haematopoietic injury response, used mouse haematopoietic stem progenitor cells (HSPCs). They calculated that their barcoded miRNA pool would be limited by the HSPCs to be used, in terms of total number of HSPCs available from one source, and in the limited number of engraftment events occurring per mouse. As they found this factor to be a bottleneck in their experiment, they limited the number of miRNAs to be studied to 135, creating a library which focussed specifically on haematopoietically-expressed miRNAs (Adams *et al.*, 2012). More recent studies of shRNAs have successfully used much larger libraries in HSPCs, with a study by Wang *et al.* (2012) utilising a library containing shRNAs targeted to 947 cancer-related genes. In this study, they achieved much higher rates of viral infection compared to the Adams *et al.* (2012) study, as well as high levels of cell engraftment in the mouse models used (Wang *et al.*, 2012).

Quantification and validation of cancer related genes identified in functional screens can take a number of different forms. Some studies have isolated integrated events before cloning them into a recipient vector for quantification using sequencing (Zender *et al.*, 2008). This approach has an added advantage in that the recipient vector can then itself be used for subsequent validation techniques. Another approach is to extract all the genomic DNA, and then perform deep sequencing (Wang *et al.*, 2012). This method relies upon widespread integration of events of interest, as well as a large depth of sequencing to avoid missing potentially interesting results. Barcoding approaches have also been successful and allow for the

detection of specific barcodes using PCR in combination with a bead based detection system (Adams *et al.*, 2012). qPCR approaches have also successfully been utilised, especially in smaller scale screens (Hope *et al.*, 2010), where the labour intensity of the technique is less noticeable.

Construct quantification is expected to identify a number of hits, which are typically validated using a range of approaches. In large scale screens, hits are usually confirmed in smaller experiments focussing only on the gene of interest. Different shRNA sequences can be used to ensure that the results seen are not due to off target effects. Where shRNAs or cDNAs have been used in the initial screen, the reciprocal experiment can be performed. In samples where shRNAs targeted to 3' untranslated regions (UTRs) have been used, these can be rescued through expression of cDNAs that do not contain this region and therefore cannot be targeted by these shRNAs. Researchers may also wish to investigate the expression levels of the locus in question in a selection of human disease samples. Functional analysis can then be performed to analyse the role of the gene, although exactly what this entails varies greatly depending on the known functions of the candidate gene and disease being studied.

4.1.4 Negative Selection Assay

The selection assay used in this project is based on the concept of negative selection. cDNAs representing potential tumour suppressor genes from chromosome 12p or shRNAs targeted to potential oncogenes from chromosome 11q were delivered by lentiviral expression vectors into cells containing the aberration of interest. The cells were then expanded *in vitro* and *in vivo* in immunocompromised mice and any detrimental effects on growth by vector expression were measured by sequencing the relative proportions of integrated vectors before and after leukaemic cell expansion, as well as at a number of time points during *in vitro* culturing. An overview of the negative selection strategy is shown in Figure 22.

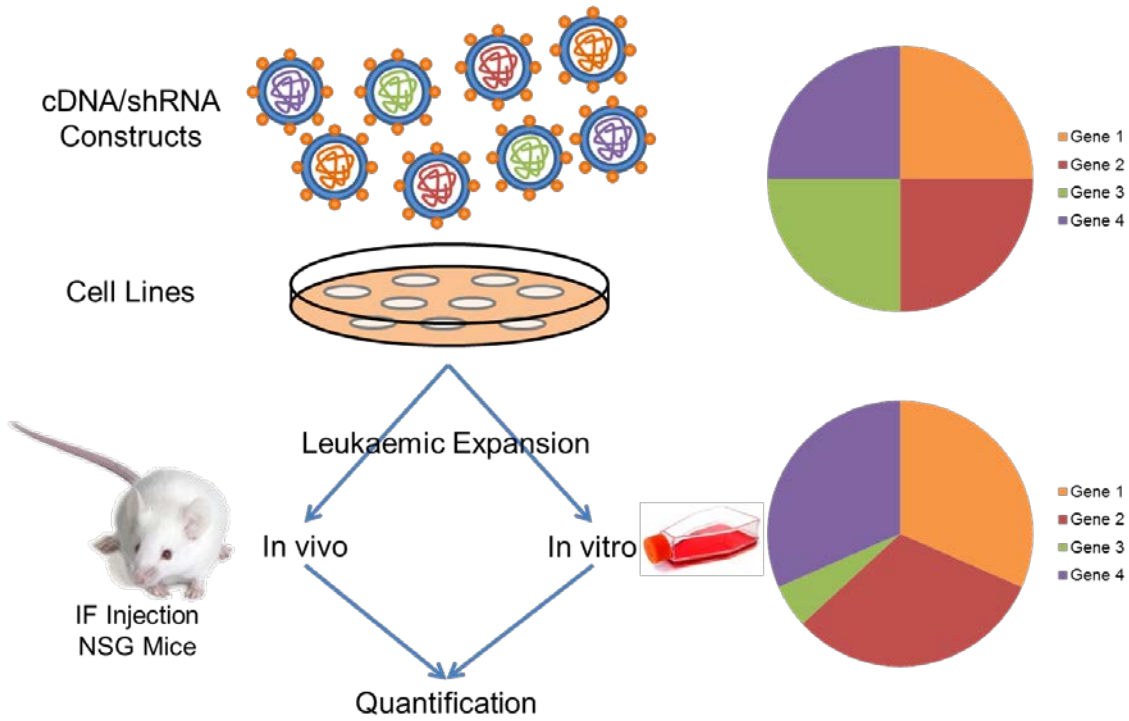


Figure 22. The negative selection assay to be used in this project. Lentiviral particles, each containing an individual expression construct for a different shRNA or gene are pooled and used to transduce target cells containing the genetic aberration of interest. The cells are then cultured in vitro or engrafted into immunocompromised Nod-Scid Gamma mice. The relative abundance of each construct is then measured before and after engraftment into the mice, and at multiple time points during the in vitro expansion. In this example, gene 3 is negatively selected against and would therefore be considered a hit.

4.1.5 Immune-Deficient Mouse Models

The vast majority of *in vivo* experimentation using mouse models to study leukaemia is centred upon immune-deficient strains. These strains are unable to produce the full complement of haematopoietic lineages and accept transplants of human HSCs, which localise to the BM and populate the blood. The first major breakthrough in developing an immune-deficient mouse model was with the severe combined immune-deficient (*Scid)* strain of mouse, which lacked both B and T cells (Bosma *et al.*, 1983). Experimentation with the *Scid* mouse successfully demonstrated that the strain could be populated with human haematopoietic cells. Moisier *et al.* (1988, 1991) infused leukocytes isolated from peripheral blood (PB), and demonstrated the mouse was able to produce human B and T cells that could mount an immune response when exposed to tetanus toxin (Moisier *et al.* 1988, 1991). McCune *et al.* (1988) also demonstrated the ability of cells to engraft and produce B and T cells, but did so by transplanting human foetal tissues and foetal liver cells. (McCune *et al.*, 1988) The fact that human lymphocytes could survive in this mouse strain meant that they could be infected with HIV-1, subsequently establishing the first humanised models of AIDS.

Despite these successes, the *Scid* model was limited, as the strain still displayed high levels of innate immune function, as well as the spontaneous emergence of B and T cells with age. Shultz *et al.* (1995) backcrossed *Scid* mice with nonobese diabetic (NOD) mice, which harboured defects in their innate immune systems. The resulting NOD-*scid* mice supported high levels of human cell engraftment (Shultz *et al.*, 1995). Further investigation into this strain using long positional cloning identified the *Sirpa* gene to be responsible for enabling NOD mice to support human haematopoietic stem cells (Takenaka *et al.*, 2007). Human *CD47*, expressed on all haematopoietic cells, is tightly bound by NOD Sirpa protein, present on the surfaces of NOD mouse myeloid cells. It induces host macrophage tolerance after the transplant of human HSCs, allowing the human cells to grow without a macrophage response from the host animal. In other mouse strains, Sirpa does not bind CD47 and macrophages phagocytose the foreign cells (Jaiswal *et al.*, 2009).

Whilst the NOD-*scid* mouse model addressed human haematopoietic transplantation issues identified in the *Scid* strain, other drawbacks were identified. The NOD-*scid* model suffers from a high incidence of thymic lymphoma, preventing many long-term studies. Furthermore, they still have an active NK population that provide some resistance to engraftment. For this reason, two derivative strains of NOD-*scid* mice were developed. Both of these strains carry altered IL-2R common γ chains, a critical component in IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21 signalling. In NOG (NOD/Shi-*scid*/IL-2Rγ^{null}) mice, the chain is truncated, whilst in NOD *scid* gamma (NSG) mice, the chain carries a mutation (Ito *et al.*, 2002, Shultz *et al.*, 2005). As a result of these changes, these mice have a complete loss of B, T and NK cells, and they subsequently demonstrate a 5-fold higher rate of engraftment of CD34+ human haematopoietic cells. The defects in cytokine signalling implemented by these changes also prevent lymphomagenesis (Ito *et al.*, 2002; Shultz *et al.*, 2005), allowing for long term studies to be successfully undertaken after engraftment of human HSCs.

These strains are currently being further developed to produce more humanised mice that express human cytokines such as thromobopoietic, IL-3 and GM-CSF (Willinger *et al.*, 2011). These models are designed to further boost the rate at which human HSCs are successfully engrafted.

4.1.6 Next Generation Sequencing

The Sanger sequencing method has been at the forefront of DNA sequencing technologies for much of the last 25 years, highlighted by its role in completing the only finished-grade human genome sequence (Human Genome Consortium, 2004). This method is based upon the use of dideoxynucleotide triphosphates (ddNTPs), which terminate chain elongation by DNA

polymerase, and are either radioactively or fluorescently labelled for detection by eye or in an automated machine. Although revolutionary when first introduced, the Sanger method is an ageing technology with a number of unresolved drawbacks, which have resulted in a reduction in its use over the last few years. Instead, preference for a number of new technologies, referred to as Next Generation Sequencing (NGS) methods, has emerged. Addressing issues such as high cost, low speed and developing the ability to sequence longer DNA fragments has been the focus of developments in NGS technologies. This project has made use of the Illumina MiSeq system for the quantification of lentiviral construct integration, in conjunction with a PCR-based method of target enrichment.

4.1.6.1 Illumina Sequencing by Synthesis

The Solexa sequencing platform, available since 2006, was acquired by Illumina in 2007. The Illumina platforms are based upon the principles of sequencing-by-synthesis in combination with reversible terminator nucleotides for each of the four bases (Mardis, 2008). A fragmented DNA library is prepared and special adaptors are ligated to both ends. One end of these adaptors is immobilised to a solid flow cell. The DNA then bends over and attaches to the flow cell at its other end, resulting in the formation of a bridge. This bridge is then amplified multiple times, using Bst polymerase for the synthesis of new strands and formamide to denature complete DNA strands. This amplficiation process results in the formation of clusters containing approximately 1000 single stranded amplicons formed from the same template (termed polonies), which are located in distinguishable locations within the flow cell (Mardis, 2008).

At this stage, a sequencing primer is added which hybridises to universal primer sites present on the adaptors flanking the DNA sequence. Specially modified reversible terminator nucleotides are also added, which carry a fluorophore specific to the base it is attached to (A, C, T or G). These nucleotides are blocked at the 3' end, and require the cleavage of two chemical bonds to remove the fluorophores from the base. Once a nucleotide has been incorporated, the slide is scanned with two lasers to detect which nucleotide has been added to the sequence. Chemical cleavage removes the block from the 3' end of the nucleotide, restoring the 3' –OH group and allowing further extension of the sequence. The read length and data output of the MiSeq machine used in this project are flexible, depending on which reagent kit is used and the settings applied to the machine. Read lengths available range from 25 nucleotides with a single direction of coverage, up to a maximum of 300 nucleotides with forward and reverse coverage. In a paired end run, the same molecule is read in one direction, and then in the reverse direction on the opposite strand. This provides information about both ends of the molecule, as well as the physical distance between them in the genome, helping in

both the resolution of structural rearrangements and the assembly of genomes across repetitive regions (Fullwood *et al.*, 2009). Each run on the machine can produce up to 15 gigabases of data, with run times of around 4 hours for a single-read run, and 1-2 days for a paired end run (Fullwood *et al.*, 2009).

For the quantification of integrated viral vectors in this project, two rounds of PCR are used to prepare libraries before sequencing. The first round of PCR is performed on DNA extracted from cell populations transduced with lentiviral pools, which amplifies either the whole shRNA or part of the cDNA, depending on which vectors are integrated. The primers in this first round of PCR add adaptor sequences that are complementary to the second round PCR primers. In this second round, viral vector sequences are further amplified and barcode sequences are added that label specific samples. At this stage, multiple samples can be pooled and primers or primer-dimers are removed by bead purification in preparation for the sequencing process.

4.2 Aims

After the identification of a commonly deleted region on chromosome arm 12p and three commonly amplified regions on chromosome arm 11q described in the previous chapter, the objectives of the studies in this chapter were therefore:

- To establish an optimised pipeline to perform functional screening using a negative selection assay. This will include:
    - Optimisation of lentiviral particle production.
    - Characterisation of AML cell line growth in an *in vivo* model.
    - Design of primers and optimisation of PCR conditions to produce amplicons suitable for the quantification of integrated lentiviral constructs by next generation sequencing.
- To identify tumour suppressor genes from within the commonly deleted 12p region through the expression of cDNAs corresponding to the identified candidate genes in a negative selection assay.
- To identify oncogenes from within the commonly amplified 11q regions through the expression of shRNAs targeted to the identified candidate genes in a negative selection assay.

4.3 Methods

4.3.1 Cloning of cDNAs into SIN-SIEW

4.3.1.1 Gateway Expression Clone Production

Open reading frames (ORFs) were purchased from Source BioScience and Genecopoeia in the Gateway PLUS Shuttle Clone format. These are plasmids that consist of the desired ORF flanked by attL recombination sites. In the presence of Gateway Plus Recombination enzyme and a suitable destination vector containing the complementary attR recombination sites, the ORF can be recombined into the destination vector of choice. For this project, the ORFs were cloned into the SIN-SIEW lentiviral transfer vector (a kind gift from Prof. Olaf Heidenreich, Newcastle University) This process is summarised in Figure 23.

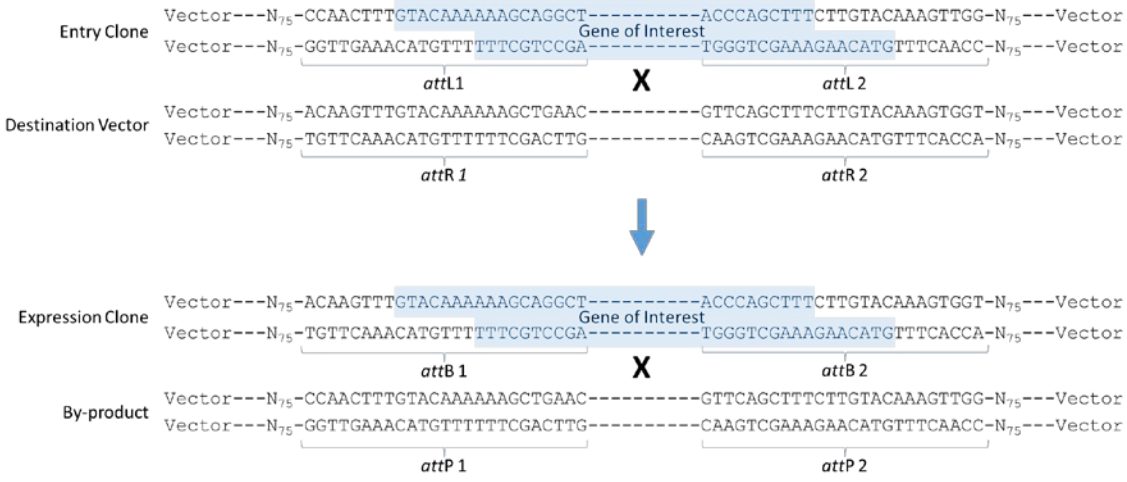


Figure 23. The Gateway recombination process and associated sequences.

Shuttle clones that were received as agar stabs were spread onto LB agar plates containing chloramphenicol using a sterile pipette tip. The plates were cultured at 37°C for 16 hours, and a single colony was taken to establish an overnight culture in 5ml of LB agar containing chloramphenicol. A QIAprep Spin Miniprep Kit (Qiagen) was then used to extract the plasmid DNA from these preparations as described previously.

To perform the LR recombination, 100ng of entry clone (Genecopoiea) was mixed with 150ng of destination vector in a total reaction volume of 8µl, diluted in TE buffer (pH 8.0). 2µl LR clonase II enzyme (an Invitrogen proprietary mixture of Integrase, integration host factor and Excisionase) was then added to each sample and the solution mixed briefly by pulse vortexing. The mixture was then incubated at 25°C for 16 hours. The ligase was then cleaved by the addition of 1µl proteinase K (Sigma-Aldrich) solution to each reaction and incubation at 37°C for 30 minutes.

Competent One Shot Stbl3 *E. coli* cells (Life Technologies) were thawed on ice and split into aliquots of 250µl each. To each aliquot, 1µl of recombination reaction was added and the suspension mixed by flicking before incubation on ice for 30 minutes. After this period, samples were heat shocked by placing them into a 42°C heat block for 45 seconds before being returned to ice for a further 2 minutes. 125µl Super Optimal Broth with Catabolite repression (SOC) medium was added to each sample and samples were placed into a shaking incubator for 1 hour. The sample was then spread onto an LB-agar plate containing ampicillin using an angled glass pipette that had been flame sterilised. The plates were then cultured at 37°C for 16 hours and inspected for colony formation. Growth indicated that the LR recombination was successful and single colonies were isolated for the preparation of glycerol stocks and subsequent plasmid extraction by either Qiagen Mini or Maxi prep as described in Chapter 2.2.1.

4.3.1.2 Sequencing of Lentiviral Inserts

To sequence the ORFs inserted into SIN-SIEW, universal primers were used. A forward primer was used to prime a PCR reaction from the SFFV promoter present on SIN-SIEW (SFFV F), whilst a complementary reverse primer primed the PCR reaction from the IRES present after the inserted ORF. The reagents and conditions for these reactions are listed in Table 15 and Table 16. A list of primers used is supplied in Supplementary Table 33.

Reagent	**Volume**	**Final Concentration**
5 x Hf Buffer	20µl	1x
10mM dNTPs	5µl	200µM
Phusion Hot Start II Polymerase	1.25µl	1 unit
H_2O	63.75µl	-
SFFV F 10µM Primer	5µl	500pM
IRES R 10µM Primer	5µl	500pM
1ng SIN-SIEW Plasmid	1µl	-

Table 15. Reaction mixture for lentiviral insert sequencing

Step	Temperature	Time (mm:ss)	Cycles
Denaturation	98°C	00:30	
Denaturation	98°C	00:20	35
Annealing	63°C	00:40	
Elongation	72°C	02:00	
Elongation	72°C	10:00	
	4°C	*Hold*	

Table 16. Reaction conditions for lentiviral insert sequencing.

4.3.1.3 PCR Purification

To purify the products of the sequencing PCR, the QIAquick PCR Purification Kit (Qiagen) was used. This kit binds DNA to a silica membrane within a spin column in the presence of a high salt concentration, where it is washed before elution in a suitable volume of water. The binding buffer in the kit is designed to allow the efficient binding of single or double stranded PCR products larger than 100 base pairs, and to remove over 99% of contaminating primers and unused nucleotides.

To the 100µl of PCR product, 500µl of binding buffer PB was added and the solution mixed by pipetting. The pH indicator present turned yellow if the pH of the solution was correct. If the solution was orange or violet, 10µl of 3M sodium acetate (pH 5.0) was added and the solution mixed, which resulted in the mixture turning yellow. The sample solution was applied to a QIAquick spin column and centrifuged at 17,900 x *g* for 60 seconds. The resulting flow through was discarded and 750µl of wash buffer PE applied to the column and centrifuged at 17,900 x *g* for 60 seconds. The flow through was discarded again and the centrifugation step repeated to ensure the removal of residual ethanol present in buffer PE. The QIAquick column was then transferred to a 1.5ml microcentrifuge tube and 50µl elution buffer EB applied to the silica membrane. The column was incubated at room temperature for 1 minute before centrifugation at 17,900 x *g* for 60 seconds. The eluted DNA was measured for concentration and purity using a NanoDrop 1000 Spectrophotometer (NanoDrop), and sent to DBS Genomics (Durham University) for sequence analysis using Sanger sequencing.

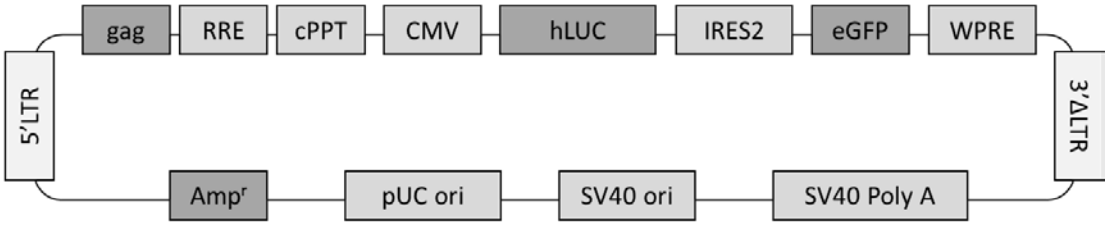
4.3.2 Production of Lentiviral Particles

In order to safely produce lentiviral particles capable of transducing human cells, the genes required to code for the various viral components have been split between multiple packaging plasmids. These packaging plasmids are transiently co-transfected with a transfer vector into a packaging cell line, where the viral genes are expressed and the components packaged into a complete virus particle. Schematics for plasmids used in this project are shown in Figure 24.

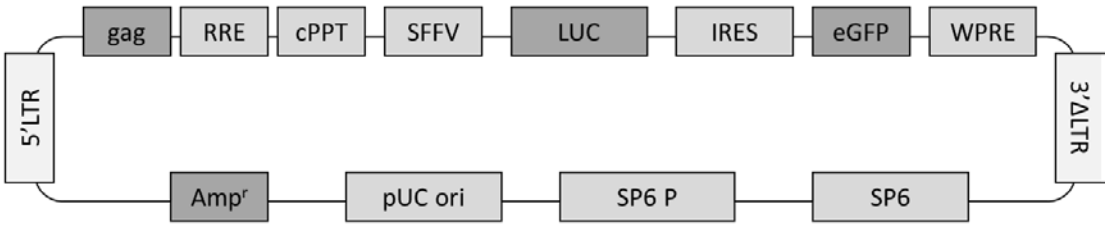
Lentiviral particles were produced in the transformed human embryonic kidney cell line 293T, or variants derived from this cell line. The 293FT cell line was produced by Life Technlogies and is a fast growing derivative of the 293T cell line, selected for its superior ability to produce lentiviral particles when compared to 293T. Lenti-X 293T is another variant of the 293T cell line which was developed by Clontech for use with the Clontech Lenti-X HTX Packaging System.

4.3.2.1 Plasmids used for Lentiviral Production

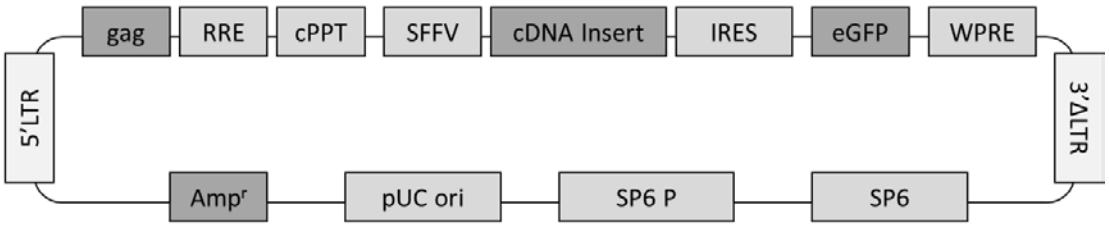
pLv81



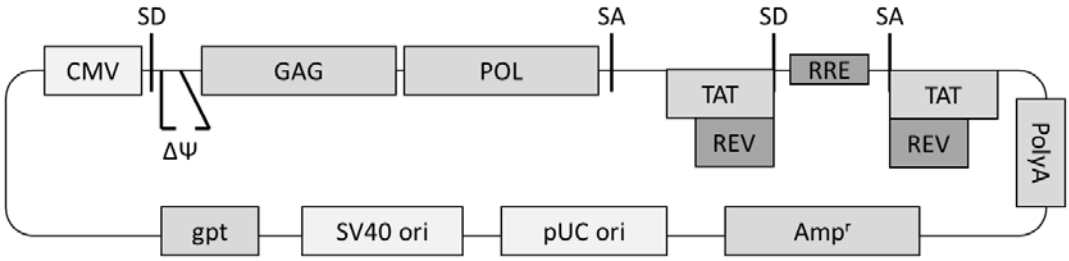
pSLIEW



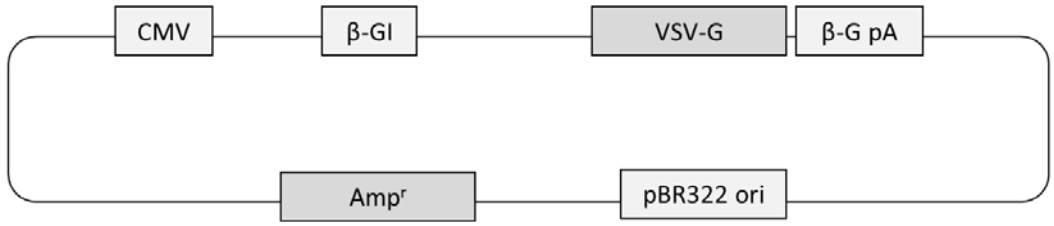
SIN-SIEW



pCMVΔR8.91



pMD2.G



pGIPZ

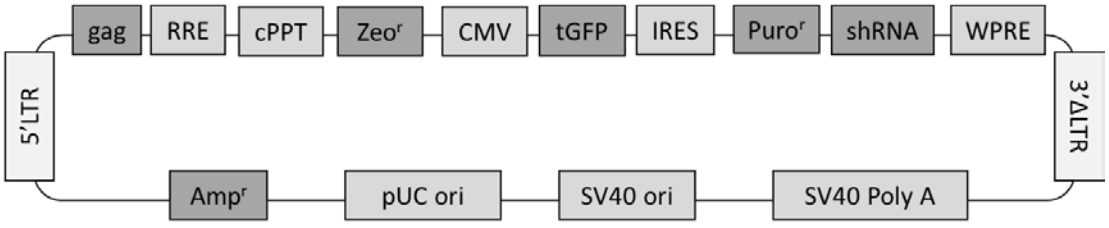


Figure 24. Maps of plasmids used for the production of lentiviral particles. RRE: Rev Response Element, cPPT: Central Polypurine Tract, eGFP: Enhanced Green Fluorescent Protein, IRES: Internal Ribosome Entry Site, Puro^r: Puromycin Resistance, shRNA: shRNA entry site, WPRE: Woodchuck hepatitis Posttranscriptional Regulatory Element, 3'ΔLTR: Truncated HIV/FIV 3' Long Terminal Repeat retroviral promoter, SV40 Poly A: Simian Virus 40 poly(A) site, SV40 ori: Simian Virus 40 origin of replication, pUC ori: Plasmid origin of replication, Amp^r: Ampicillin Resistance, 5' LTR: HIV/FIV 5' Long Terminal Repeat retroviral promoter, Zeo^r: Zeocin Resistance, CMV: Cytomegalovirus Promoter, tGFP: Turbo Green Fluorescent Protein,

Three different methods were used for the transfection of the lentiviral packaging plasmid mixture, and these are listed below.

4.3.2.2 Calcium Phosphate Precipitation

For the production of pGIPZ based lentiviral particles, the Trans-Lentiviral Packaging System was used (GE Dharmacon). This system utilises five separate packaging plasmids to package pGIPZ based transfer vectors with a high efficiency. Calcium phosphate forms complexes with DNA that co-precipitate and are absorbed by cells through phagocytosis. 24 hours before transfection, 1.2 x 10^6 packaging cells were plated into wells of a 6-well tissue culture plate in 2ml DMEM. Before transfection, 6µg of transfer vector was mixed with 4.3µl Trans-Lentiviral Packaging mix and the mixture made up to 135µl using dH_2O. 15µl of $CaCl_2$ (concentration not defined by manufacturer) was added to the DNA mixture and the tube mixed by pulse vortexing. Whilst vortexing, 150µl 2x Hank's Balanced Salt Solution (HBSS) was added dropwise to the mixture. The mixture was then incubated at room temperature for 3 minutes before addition to the packaging cells. The cells were then incubated at 37°C, 5% CO_2 for 16 hours, when the media was exchanged for DMEM containing 5% FCS and incubated for a further 48 hours before harvesting of viral supernatant.

4.3.2.3 EndoFectin

EndoFectin (GeneCopoeia) allows the transfection of plasmids as it forms cationic polymer complexes with plasmid DNA that are incorporated into cells through endocytosis. 3 x 10^6 packaging cells were plated into 10cm tissue culture plates 24 hours before transfection in 10ml DMEM. On the day of transfection, two solutions were prepared in 1.5ml microcentrifuge tubes. The first is the packaging mix, which consists of 200µl Opti-MEM I (Gibco), 10µg transfer construct, 2.5µg pCMVΔR8.91 and 7.5µg pMD2.G. An EndoFectin dilution was also prepared by mixing 200µl Opti-MEM I with 15µl EndoFectin. The EndoFectin dilution was then added to the packaging mix by dropwise pipetting, whilst the packaging mix was constantly agitated by flicking of the tube. The resulting mixture was then incubated at room temperature for 20 minutes. This incubation allows the EndoFectin (a cationic polymer) to form a complex with the nucleic acids present in the packaging mixture, increasing transfection efficiency. The

entire mixture was added dropwise to the packaging cells and the plate gently shaken, not swirled to ensure even distribution of nucleic acid complexes.

After incubation at 37°C, 5% CO_2 for 16 hours, the media was exchanged for 10ml DMEM containing 5% FCS and cells were cultured for a further 48 hours before harvesting of viral supernatant.

4.3.2.4 Lenti-X Polymer Packaging

The Lenti-X HTX Packaging System is designed by Clontech and consists of a specially formulated packaging mix in conjunction with a transfection polymer to produce high titre lentivirus. 24 hours prior to transfection, 4 x 10^6 packaging cells were seeded onto 10cm tissue cultures plates in 10ml DMEM. Immediately before transfection, a packaging mix and a polymer mix were prepared in 1.5ml microcentrifuge tubes. The packaging mix was made up of 557µl Xfect reaction buffer, 36µl Lenti-X HTX packaging mix and 7µl of transfer vector diluted to 1µg/µl. The polymer mix consisted of 592.5µl Xfect Reaction Buffer and 7.5µl Xfect Polymer. It was critical at this point that the Xfect Polymer did not remain in aqueous solution for longer than 30 minutes so the subsequent steps were completed quickly. The polymer solution was added to the packaging solution and mixed by pulse vortexing for 10 seconds. The DNA-polymer mixture was then incubated at room temperature for 10 minutes to allow the polymer to form complexes with the DNA, before the entire mixture was added dropwise to the packaging cells. After incubation at 37°C, 5% CO_2 for 16 hours, the media was exchanged for fresh DMEM containing 10% FCS and the cells incubated for a further 48 hours before viral supernatant was harvested.

4.3.3 Lentiviral Particle Harvesting

The viral particles produced are released into the cell culture supernatant. This was carefully removed by pipetting and centrifuged for 5 minutes at 400 x *g* to pellet cellular debris. The supernatant from this step was then syringe filtered through a low protein binding 0.45µm Millex Syringe Polyvinylidene Difluoride (PVDF) filter (Merck Millipore). If a second harvest was required, the media was replaced with fresh DMEM and the harvesting procedure repeated 24 hours later. If no concentration step was to take place after harvesting, the supernatant was divided into aliquots and stored at -80°C.

4.3.4 Lentiviral Particle Concentration

4.3.4.1 Lenti-X Concentrator

Lenti-X Concentrator (Clontech) is a reagent which allows for concentration of lentiviral particles without the high centrifugation speeds of ultracentrifugation. The clarified viral supernatant, harvested as described above was mixed in a 3:1 ratio with Lenti-X Concentrator.

The mixture was then incubated at 4°C for between 30 minutes and 16 hours, and the sample subsequently centrifuged at 1500 x *g* for 45 minutes at 4°C. The supernatant was carefully removed and the viral pellet resuspended in a volume of DMEM equivalent to 1% of the initial volume of viral supernatant used. The sample was then split into aliquots and stored at -80°C.

4.3.4.2 Ultracentrifugation

For the concentration of lentiviral particles by ultracentrifugation, the ultracentrifuge buckets, lids adaptors and disposable conical centrifuge inserts were sterilised using 70% ethanol and air dried in a laminar flow hood. 30ml of clarified viral supernatant was transferred to the dry centrifuge inserts and paired centrifuge buckets were carefully balanced to within 0.2g of each other using extra medium. The ultracentrifugation was performed at 86,242 x *g* for 2 hours, in a Beckman Optima ultracentrifuge fitted with a SW 32Ti rotor cooled to 4°C. A constant vacuum was applied to the samples for the duration of the centrifugation step. After the centrifugation, the supernatant was carefully removed by decanting and pipetting, and the viral pellet resuspended in 1% of the starting volume of viral supernatant loaded into the centrifuge. Resuspended viral particles were aliquoted and stored at -80°C.

4.3.5 Lentiviral Transduction

Suspension cell lines were transduced by the addition of lentivirus containing media to the suspension culture in the presence of 8µg/ml polybrene. Polybrene increases transduction efficiency by neutralising the charge repulsion that occurs between virions and sialic acid present on cell surfaces (Davis *et al.*, 2004). Spinfection was also used to increase transduction efficiency. Cells were plated into wells of either 6-well or 24-well plates at a concentration of 0.5 x 10^6 cells/ml, with a total volume per well of 3ml for 6-well plates and 1ml for 24-well plates. Media containing viral particles was added to each well in addition to polybrene, and the plates were wrapped in Parafilm. The plates were then placed into a centrifuge warmed to 32°C and spinfection performed at 1500 x *g* for 2 hours. Following spinfection, the parafilm was removed and the cells incubated for 16 hours at 37°C.

After this incubation period, the media containing viral particles and polybrene was exchanged for fresh culture media to allow ongoing culturing of transduced cells. Larger volume cultures were transferred to a suitably sized flask.

4.3.6 Lentiviral Titre Determination

Lentiviral stocks were titrated in 293T or 293FT cells. 1 x 10^5 cells in 1ml media were plated in each well of a 24-well tissue culture plate. Between 1µl and 100µl of viral supernatant was added to each well and the spinfection protocol followed as detailed in 4.3.5. The percentage

of cells expressing GFP was measured as detailed in 4.3.10.1and the viral titre calculating using the following formula:

$$eGFP\ titer, TUml^{-1} = \frac{F \times N}{100 \times V \times D_f}$$

Where F = % of eGFP-positive cells as determined by FACS, N = estimated number of target cells per well, V = volume (in ml) of the lentivirus-containing supernatant that was added per well and D_f = the dilution factor for lentivirus-containing supernatant.

4.3.7 *In Vivo* Experimentation

4.3.7.1 Project Approval

Animal experiments were conducted under the Home Office Project License (PPL 60/4552), by trained researchers who had completed Home Office training courses and held Personal Licenses under the Animals (Scientific Procedures) Act 1986 (Personal License I2B4BD9C7).

4.3.7.2 General Handling and Monitoring of Mice

Mice were kept with littermates in individually ventilated cages (IVCs) provided with a sterile filtered air supply. The mice were regularly checked for condition and the growth of tumours was carefully monitored. Mouse weights were recorded weekly and animals were immediately culled if their weight dropped below 90% of their previously recorded maximum weight, or tumours were measured to be larger than 1cm in diameter. Mice were also culled if their condition fell to unacceptable levels, or if disease burden was affecting their quality of life by restricting movement or causing the animals to experience excessive levels of pain.

4.3.7.3 Murine Anaesthesia

Mice were anaesthetised by placing them into an induction chamber filled with 95% oxygen and 5% isoflurane, set to a flow rate of 1L/min. Mice were monitored for signs of anaesthesia including loss of movement, loss of self-righting ability, reduction in breathing rate and failure to respond to physical stimuli (‘pinch test’).

4.3.7.4 Intrafemoral Injection

After the induction of anaesthesia, isoflurane concentration was reduced to 2.5% and the mouse transferred to a face mask equipped with an anaesthetic scavenger mechanism to prevent leakage of anaesthetic gas. The legs of the mouse were shaved using a small electric clipper and the skin sterilised with antiseptic solution. An insulin needle was used to draw up 50µl cell suspension per mouse, containing the cells to be engrafted suspended in full media. The mouse leg was flexed to expose the head of the femur, and an empty insulin needle used to bore through the head of the femur and form a canal within the bone marrow for engrafted

material to be injected. The empty needle was removed and the needle containing the cell suspension inserted into the canal bored in the femur, where 50µl of cell suspension was then injected. The mouse was ear punched whilst under anaesthetic to provide identification (LN – left ear notch, RN – right ear notch, BN – both ear notch, NN – no ear notch, 2LN – two left ear notches, 2RN – two right ear notches) and subcutaneously injected with 10µl of carprofen (Rimadyl) diluted 1:10 as post-surgery analgesia.

4.3.7.5 In Vivo Imaging System

Before introducing immunocompromised mice to the *in vivo* imaging system (IVIS) Spectrum, the machine was thoroughly sterilised with 70% ethanol and allowed to air dry. 20 minutes prior to imaging taking place, mice were intraperitoneally injected with 100µl of luciferin reagent. The mice were then placed under anaesthesia as described in Section 4.3.7.3, and after the induction of anaesthesia the isoflurane concentration was reduced to 2.5% and the mice transferred to anaesthetic nose cones present within the IVIS, supplied with the same gas supply used for the induction of anaesthesia. Mice were orientated to the required position for imaging and bioluminescence was measured over a period of time automatically determined by the system. The luminescent signal was quantified in radiance photons/second.

4.3.7.6 Harvesting of Organs

A post mortem dissection was performed on dead mice and organs examined for signs of engraftment. The spleen was always removed from mice and the liver and kidneys were taken if they appeared to be enlarged or had other signs of engraftment such as white spots caused by excess blast accumulation. Organs were weighed and suspended in PBS. The mouse legs were dissected to excise the femurs and tibias, which were flushed using a 25 gauge needle attached to a 5ml syringe filled with PBS to release cells contained within the bone marrow.

The organs were transferred to individual 10cm sterile petri dishes and minced by fine slicing with a scalpel, before passing them through a 25 gauge needle to break up remaining pieces. The cell suspension was then passed through a 40µm cell strainer to remove debris and centrifuged at 400 x *g* for 10 minutes. The cells were resuspended in RPMI and cell concentration estimated using a Neubauer improved haemocytometer.

If required, human cells were isolated from the cell suspension using Ficoll-Paque PREMIUM 1.077 reagent. Ficoll-Paque makes use of density gradient centrifugation to isolate human mononuclear cells from other blood cells, including those found in rodents which have a higher average density than human mononuclear cells. The required volume of Ficoll-Paque was drawn up into a syringe by inserting a 25 gauge needle through the septum of the Ficoll-Paque bottle. 3ml of Ficoll-Paque was then added to a 15ml centrifuge tube and 4ml of the cell

suspension from the processed organ was carefully layered on top using a glass Pasteur pipette equipped with a rubber bulb. The layered sample was then centrifuged at 400 x *g* for 40 minutes at room temperature. After centrifugation, the mononuclear cells were removed using a glass Pasteur pipette and washed by the addition of 10ml fresh media and centrifugation at 400 x *g* for 10 minutes. The washed cell pellet was then resuspended in 10ml fresh media and cell concentration determined using a Neubaeuer improved haemocytometer. Cells were pelleted and either used immediately to extract genomic DNA or stored at -80°C. Additional viable cells were frozen and stored in liquid nitrogen as described previously.

4.3.8 Preparation of Sequencing Libraries

To produce a library for quantification of integrated lentiviral constructs, primers were designed to amplify a small section of integrated ORFs using a universal forward primer targeted to the SFFV promoter in conjunction with individual reverse primers targeting each integrated ORF. This allowed for a multiplex PCR approach to be taken to enrich the proportion of integrated construct sequences relative to background genomic DNA, whilst maintaining relative quantities of each construct in the pool. The 5' ends of the primers consisted of universal CS1 sequences that were the targets for a second round PCR with primers including Illumina adaptors and barcode sequences. 2µl of each product was analysed by agarose gel electrophoresis to ensure there was no contamination of the sample by unwanted PCR products, and to estimate the quantity of DNA produced by comparison with DNA standards of known concentration. If bands were especially faint, the round 2 PCR was repeated with an increased number of cycles to provide sufficient material for the sequencing reaction. Full details of reagents and conditions used are listed in Table 17, Table 18, Table 19 and Table 20. Primers used are detailed in Supplementary Table 33.

Reagent	**Volume**	**Final Concentration**
10 x Reaction Buffer	2µl	1x
25mM $MgCl_2$	3.6µl	4.5mM
DMSO	1µl	
10mM dNTPs	0.4µl	200µM
Fast Start Enzyme	0.2µl	1 unit
H_2O	6.8µl	-
CS1-SFFV F Primer 10µM	1µl	500pM
CS1 Pooled Reverse Primers 10µM	1µl	500pM
Sample DNA (50ng/µl)	4µl	10ng/µl

Table 17. Reaction mixture for round 1 PCR amplification

Step	Temperature	Time (Minutes)	Cycles
Denaturation	95°C	10:00	
Denaturation	95°C	00:15	25
Annealing	60°C	00:30	
Elongation	72°C	01:00	
Elongation	72°C	03:00	
	4°C	*Hold*	

Table 18. Reaction conditions for round 1 PCR amplification

Reagent	Volume	Final Concentration
10 x Reaction Buffer	2µl	1x
25mM $MgCl_2$	3.6µl	4.5mM
DMSO	1µl	
10mM dNTPs	0.4µl	200µM
Fast Start Enzyme	0.2µl	1 unit
H_2O	4.8µl	-
Barcode Primers	4µl	-
Round 1 DNA (1:100 Dilution)	1µl	-

Table 19. Reaction mixture for round 2 PCR amplification

Step	Temperature	Time (Minutes)	Cycles
Denaturation	95°C	10:00	
Denaturation	95°C	00:15	15-25
Annealing	60°C	00:30	
Elongation	72°C	01:00	
Elongation	72°C	03:00	
	4°C	*Hold*	

Table 20. Reaction conditions for round 2 PCR amplification

4.3.9 Illumina MiSeq Sequencing

Sequencing libraries were created by pooling PCR products, using 1ng of product per target construct for each sample. Library purification was then performed to remove primers and excess dNTPs. To achieve this, AMPure XP beads (Agilent Technlogies) were used, which bind DNA of specific size ranges and can be isolated using a magnet, allowing small DNA fragments to be removed. The AMPure XP beads were incubated at room temperature for 30 minutes after removal from the fridge. The bottle was then shaken until homogenous in colour, and 180µl of beads were added to each 100µl of library, and mixed by pipetting. The mixture was

then incubated for five minutes at room temperature and the tube placed into the magnetic separation device. When the solution became clear, it was removed by pipetting whilst still on the rack. The beads were then washed with 200µl of 70% ethanol, allowed to settle and the ethanol was removed. This wash step was repeated again and the beads allowed to air dry for 5 minutes whilst heated to 37°C. To the tube, 22µl of nuclease free water was added and the tube mixed by briefly pulse vortexing, before incubation at room temperature for two minutes. The tube was then returned to the magnetic separation device and allowed to stand until the beads had collected. The supernatant was then removed by pipetting and stored for subsequent use.

After purification, sample clean-up was confirmed using a 2100 Bioanalyser Instrument (Agilent Technologies) to ensure removal of primers and dNTPs, and retention of the correct library fragment size. Prepared libraries were then sent to Edinburgh Genomics (formerly The GenePool) for analysis on the Illumina MiSeq platform. Raw data files were quality controlled and number of reads associated with each construct calculated using the SeqNext software (JSI Medical Systems).

4.3.10 Flow Cytometry

4.3.10.1 Assessment of GFP Expression

The transduction efficiency was measured by assessing the proportion of cells expressing enhanced green fluorescent protein (eGFP) from SIN-SIEW constructs or turbo green fluorescent protein (tGFP) from pGIPZ constructs.

Approximately 5 x 10^5 cells were pelleted by centrifugation at 400 x *g* for 5 minutes, and resuspended in 1ml PBS. Samples were then analysed using a FACS Calibur (BD Biosciences) equipped with a 488nm laser and a 530/30nm filter. Using forward and side scatter profiles, live cells were selected for, and dead cells and cell debris were eliminated from the analysis. A minimum threshold for GFP expression was set using a non-transduced population in order to correct for auto-fluorescence. GFP positive cells were defined as those cells with a signal greater than the threshold.

4.3.10.2 Cell Surface Phenotype Analysis

Engrafted cell populations were assessed using a FACS Canto II (BD Biosciences). Frozen cell aliquots were diluted in PBS and cells pelleted by centrifugation at 400 x *g* for 5 minutes. Cell pellets were resuspended in 1ml PBS and counted. 5 x 10^5 cells were taken from each aliquot and PBS added to a volume of 100µl. To each sample, the following stained antibodies were added: 10µl anti-human CD13, 10µl anti-human CD33, 2.5µl anti-murine CD45 and 2.5µl anti-murine Ter119. Tubes were then incubated in the dark at room temperature for 30 minutes.

The samples were subsequently washed twice by pelleting in a centrifuge at 400 x *g* for 5 minutes and were finally resuspended in 500µl PBS ahead of analysis using a FACS Canto II flow cytometer. Murine cell populations were excluded by gating against cells positive on the Phycoerythrin-Cy7 (PE, anti-murine CD45 and anti-murine Ter119) channel, whilst engrafted human cells were identified as those positive on the PE (anti-human CD13) and Fluorescein Isothiocyanate 1 (FITC, anti-human CD33) fluorescence channels.

4.4 Results

4.4.1 Lentiviral Transfer Vector Selection

Initial work focussed on the GDM-1 and KG-1 cell lines, and two lentiviral delivery systems were tested. The HIV-based pSLIEW transfer vector packaged in combination with pMD2.G and pCMV packaging vectors was compared with the Feline Immunodeficiency Virus (FIV)-based commercially available pReceiver-Lv81 (pLv81) transfer vector and Lenti-Pac packaging system (GeneCopoeia). By comparison with 293T, neither cell line showed evidence of efficient transduction. However, slightly higher levels of GFP positive cells were achieved with KG-1 than with GDM-1. Of the two constructs, pSLIEW performed better and was used in future experiments (Figure 25).

At this time, the NKM-1 cell line was obtained and tested with the pSLIEW transfer construct. Using an equal number of infectious viral particles as with the GDM-1 and KG-1 cell lines, a much higher level of GFP positivity was obtained. For this reason, the NKM-1 cell line was selected for use in the negative selection assay in combination with the pSLIEW and SIN-SIEW-Gateway transfer vectors.

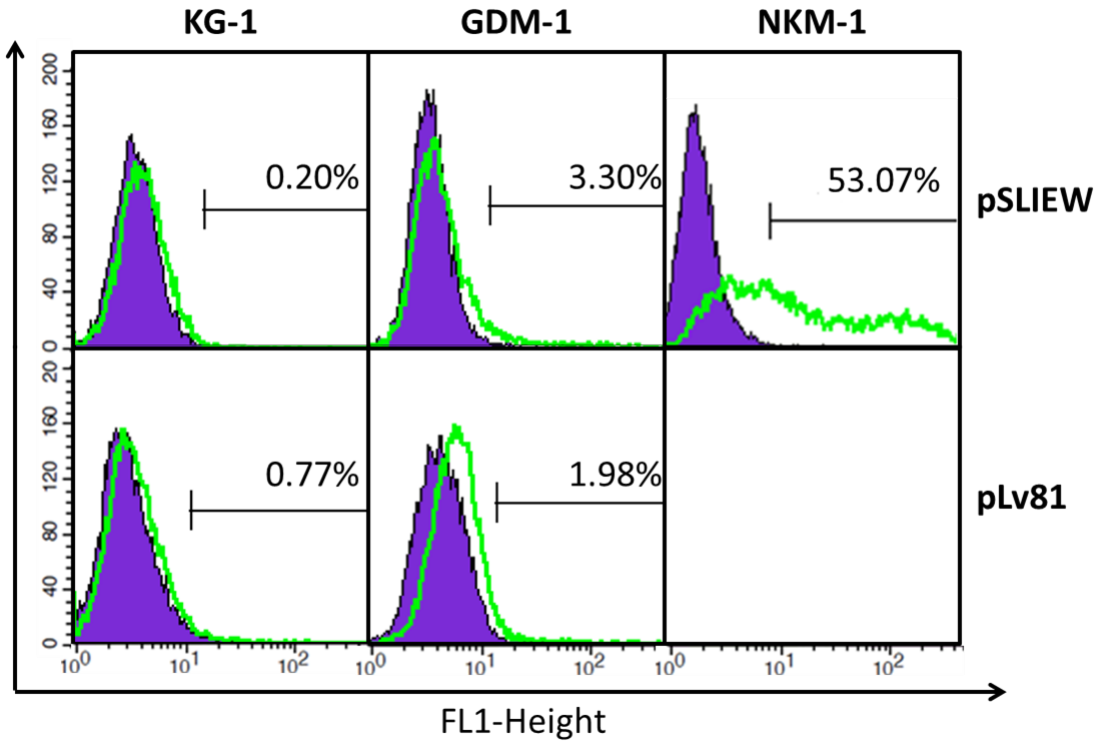


Figure 25. FACS analysis of 12p deleted cell lines transduced with equal units of pSLIEW and pLv81 lentiviral particles, as titred in 293T. Histograms for the negative control are shown as purple infill and for transduced cells as green traces. The x-axis represents FL1-Height, proportional to the amount of GFP present in each cell, whilst the y-axis shows the number of cells detected at each FL1 intensity point. The right hand marker was set using control

non-transduced cells, and the percentage indicates the proportion of cells determined to be GFP positive that are present within this region.

As only one cell line was available for the 11q project, the UoC-M1 cell line was transduced with a population of pGIPZ non silencing control lentiviral particles. Analysis of GFP positivity demonstrated that sufficient proportions (>10%) of positive cells were reached with ease (Figure 26). For this reason, the pGIPZ shRNA lentiviral vector system was selected for use in the 11q amplification project.

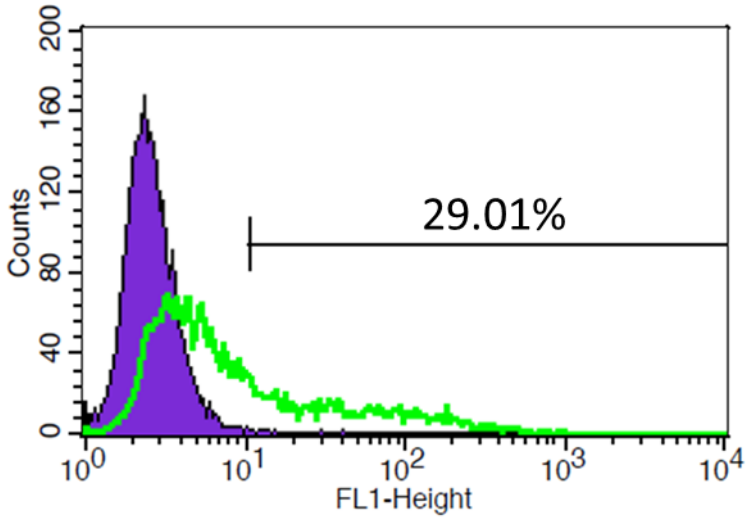


Figure 26. FACS analysis of UoC-M1 cells transduced with lentiviral particles produced using the pGIPZ non silencing control transfer vector. Histograms for the negative control are shown as purple infill and for transduced cells as green traces. The x-axis represents FL1-Height, proportional to the amount of GFP present in each cell, whilst the y-axis shows the number of cells detected at each FL1 intensity point. The right hand marker was set using control non-transduced cells, and the percentage indicates the proportion of cells determined to be GFP positive that are present within this region.

4.4.2 Optimisation of Lentiviral Particle Production

As the SIN-SIEW lentiviral vector system is not commercially packaged in an optimised kit, various lentiviral packaging conditions were tested. Three cell lines (293T, LX293T and 293FT) were assessed for their ability to produce lentiviral particles using the pSLIEW transfer vector with pCMV and pMD2.G packaging plasmids using the EndoFectin Lenti reagent. Highest titres of lentivirus were achieved with the 293FT cell line (Figure 27) which was selected for use in further optimisation.

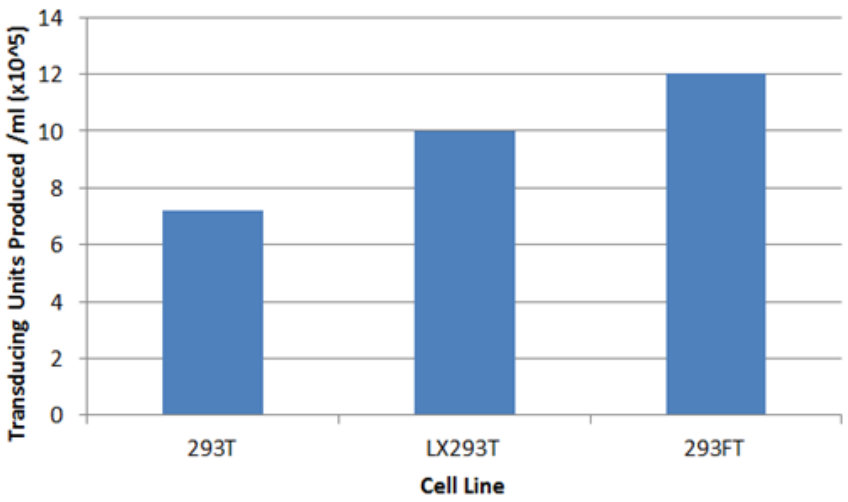


Figure 27. Comparison of packaging cell lines for lentiviral production in the 293T cell line with 2 of its derivatives. pSLIEW transfer vectors were packaged using the same amount of pCMV and pMD2.G packaging plasmids, transfected into the cells using EndoFectin Lenti packaging reagent. Lentiviral particles were harvested 72 hours after transfection and titred using the 293T cell line and subsequent flow cytometry, allowing the number of transducing units produced per millilitre of supernatant to be calculated.

Three reagents that enable plasmid uptake into the packaging cell lines were tested, including two cationic lipid formulations (Lipofectamine LTX at a range of concentrations and EndoFectin-Lenti at a standard concentration) and one polymer based formulation (Lenti-X Xfect) (Figure 28). Whilst there was a small difference observed between the optimal concentration of Lipofectamine LTX and EndoFectin Lenti, the EndoFectin lenti was selected as the reagent of choice for future experiments based on its relatively lower cost.

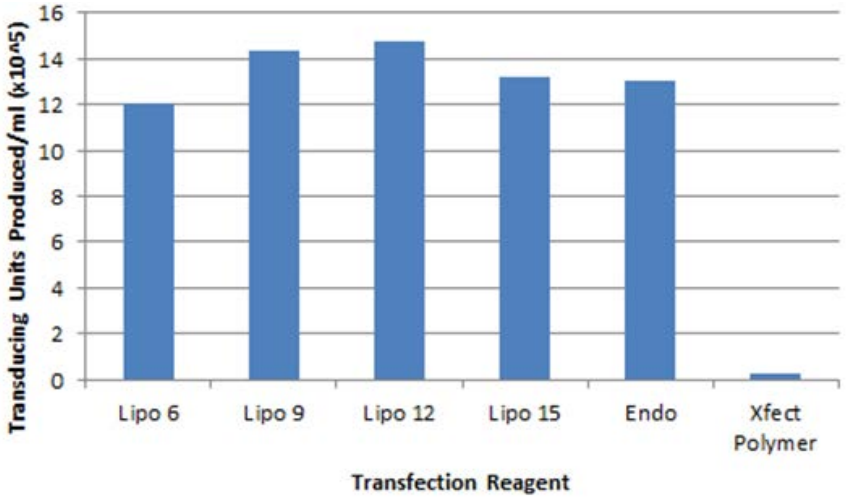


Figure 28. Comparison of transfection reagents for lentiviral production in 293FT. pSLIEW transfer vectors were packaged using the same amount of pCMV and pMD2.G packaging plasmids, transfected into the cells using a range of packaging reagents, with Lipofectamine used at a range of concentrations, following the manufacturer's

protocol. Lentiviral particles were harvested 72 hours after transfection and titred using the 293FT cell line with subsequent flow cytometry, allowing the number of transducing units produced per millilitre of supernatant to be calculated.

A range of additives were tested that have been suggested to improve lentiviral yield when present in the post transfection media. These included caffeine (2mM, Ellis *et al.*, 2011), chloroquine (25µM, Kutner *et al.*, 2009), sodium butyrate (10mM, Cribbs *et al.*, 2013) and Titer Boost (1µl/ml, GeneCopoeia) reagent, as well as exchanging regular FCS with heat inactivated FCS. Many of the additives had a detrimental effect, most significantly caffeine which saw a decrease in yield of approximately 60%. However, the presence of chloroquine was found to increase the lentiviral yield by over 30%, as shown in Figure 29, and was used in subsequent experiments.

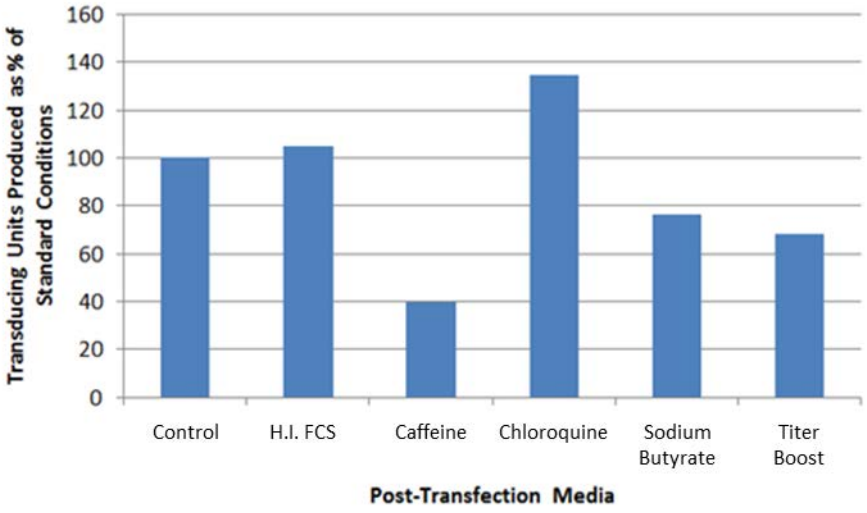


Figure 29. Comparison of post-transfection media additives for lentiviral production in 293FT. pSLIEW transfer vectors were packaged using the same amount of pCMV and pMD2.G packaging plasmids, transfected into the cells using EndoFectin Lenti. A range of post transfection additives were tested as follows – Foetal Calf Serum (FCS) exchanged for Heat Inactivated FCS (H.I. FCS), the addition of 2mM caffeine, 25µM chloroquine, 10mM Sodium Butyrate, or 1µl/ml Titer Boost reagent. Lentiviral particles were harvested 72 hours after transfection and titred using the 293FT cell line with subsequent flow cytometry, allowing the number of transducing units produced per millilitre of supernatant to be calculated.

In addition, the effects of polybrene and spinfection on lentiviral transduction efficiency were assessed. The presence of polybrene (8µg/ml) increased transduction efficiency by approximately 20% in NKM-1 cells and almost 300% in UoC-M1 cells. In conjunction with spinfection (centrifugation at 400 x *g* for 2 hours 30 minutes), transduction efficiency was further increased by an additional 20% in NKM-1 cells and over 100% in UoC-M1 cells, as demonstrated in Figure 30.

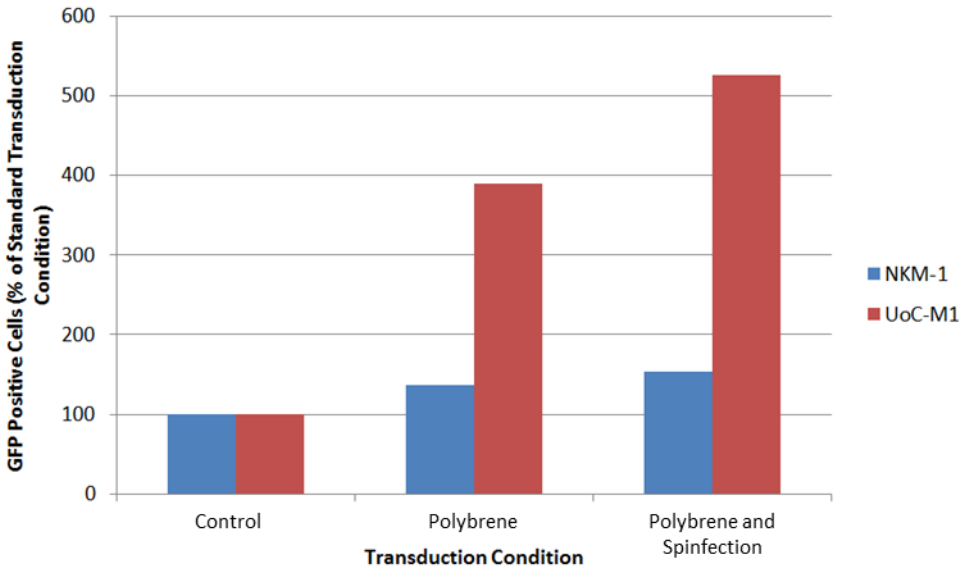


Figure 30. Comparison of transduction conditions for the lentiviral transduction of NKM-1 and UoC-M1 cell lines. Lentiviral particles containing the pSLIEW transfer vector produced in the 293FT cell line were used to transduce 5 x 10^5 NKM-1 or UoC-M1 cells. Some cells were exposed to polybrene (1µg/ml), or a combination of polybrene and spinfection (centrifugation at 400 x g for 2 hours 30 minutes). After 72 hours of incubation, GFP expression was assessed using flow cytometry.

4.4.3 *In Vivo* Models of AML Engraftment

To investigate the engraftment of the selected cell lines in the NSG mouse model, a population of each cell line was transduced with pSLIEW lentiviral particles. The expression of luciferase from this lentiviral vector allowed the growth of leukaemia to be monitored *in vivo*. $1x10^5$ cells were injected intrafemorally into female NSG mice (this procedure was performed by Dr. Helen Blair, Newcastle University, subsequent intrafemoral injections were performed by myself) and the mice were imaged at multiple time points (Figure 31). Injected cells formed solid tumours around the injection site. After sacrificing the mice, cells were recovered from the dissected tumours. In three mice, spleens were 0.1g or above in weight, suggesting that engraftment had occurred at this site (Table 21).

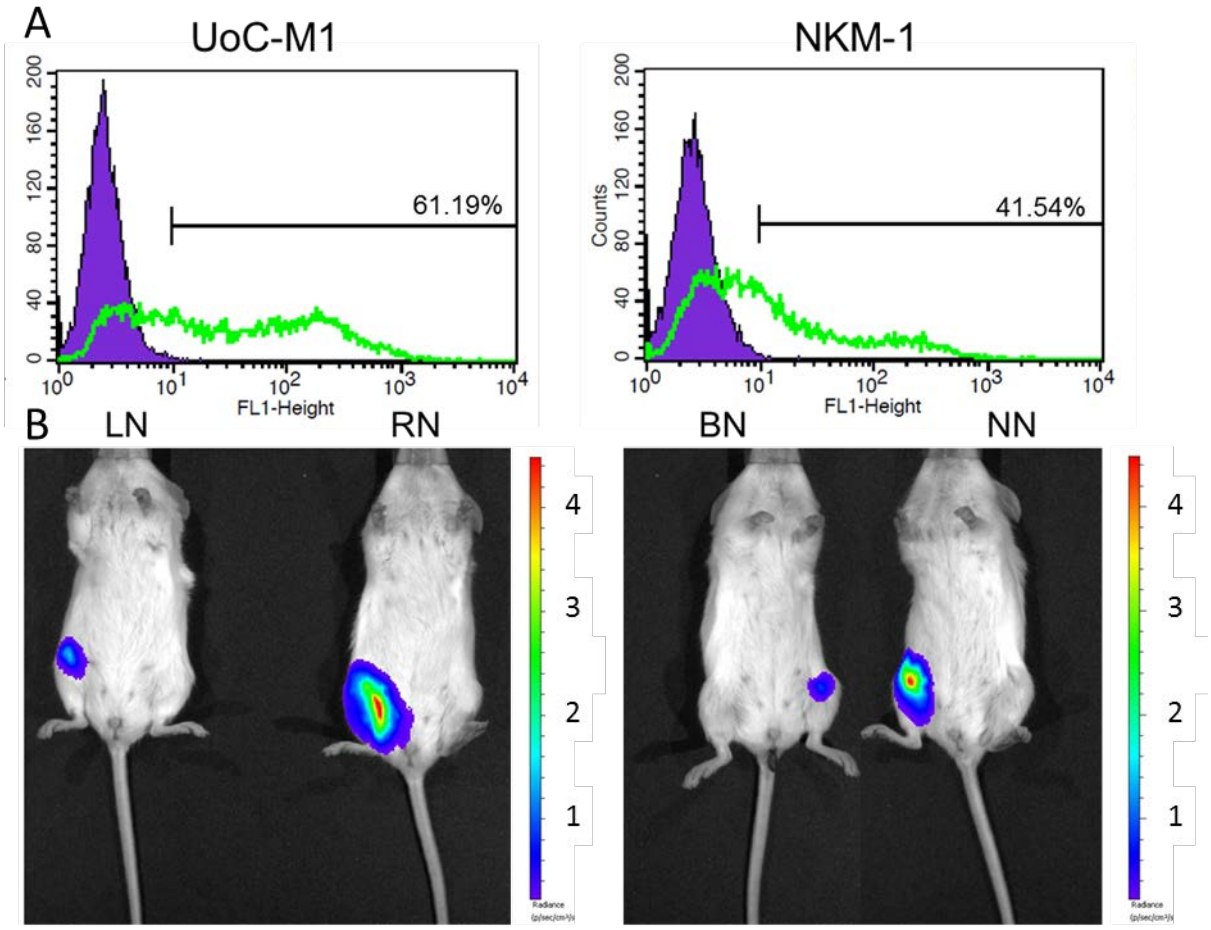


Figure 31. Engraftment of AML cell lines in NSG mice. A. GFP positivity determined by flow cytometry, 3 days after transduction before intrafemoral injection into female NSG mice. These injections were performed by Dr. Helen Blair, Newcastle University, whereas all subsequent injections were performed by myself. The x-axis represents FL1-Height, proportional to the amount of GFP present in each cell, whilst the y-axis shows the number of cells detected at each FL1 intensity point. The right hand marker was set using control non-transduced cells, and the percentage indicates the proportion of cells determined to be GFP positive that are present within this region. B. Bioluminescence imaging 75 days after injection, following intraperitoneal injection of luciferin reagent (100µl). Units shown are radiance x 10^{-4} (p/sec/cm^3/sr).

Mouse	**LN**	**RN**	**BN**	**NN**
Cell Line	UoC-M1	UoC-M1	NKM-1	NKM-1
Culled on day	100	76	104	95
Tumour Weight	0.67g	4.17g	0.02g	0.96g
Spleen Weight	0.07g	0.13g	0.6g	0.10g

Table 21. Time between transplant and cull, tumour and spleen weight for NSG mice injected with AML cell lines to investigate engraftment.

4.4.4 Optimisation of PCR Conditions for Quantification of Integrated Constructs by Next Generation Sequencing

Various combinations of first and second PCR cycle numbers were tested. Ultimately 25 first round cycles were selected to be combined with 15 second round cycles as this combination produced PCR products of sufficient yield that had not yet reached the plateau phase of PCR amplification. Detailed figures are shown in Supplementary Figure 55.

4.4.5 12p Negative Selection Pilot Study

ORF sequences corresponding to each of the 11 genes to be investigated from chromosome 12p were inserted in the SIN-SIEW lentiviral vector using the Gateway recombination system as detailed in Section 4.3.1. Lentiviral particles were produced on an individual basis and titred in the NKM-1 cell line (as selected in Section 4.4.1). This allowed for the preparation of a mixture containing equal numbers of each lentiviral construct to be prepared, and this pool was used to transduce $12x10^6$ NKM-1 cells. The majority of these cells were cultured *in vitro*, whilst some cells were injected intrafemorally into NSG mice ($5x10^5$ cells per mouse). DNA was extracted from cells cultured *in vitro* every 3-4 days, and from the organs of culled mice after separation over Ficoll-Paque by density centrifugation.

4.4.5.1 In Vitro Results

Copy number of individual integrated constructs relative to the entire pool at each time point was calculated from the number of sequence reads, as determined by Illumina sequencing. Relative copy numbers at each time point were then normalised to the starting time point level (Figure 32). Of the 11 constructs tested, the expression construct for *CDKN1B* was strongly and rapidly selected against in the first 21 days after transduction. The changes in *CDKN1B* construct proportion were highly significant when tested using Multivariate Analysis of Variance (MANOVA) (Table 22, p=<0.0001). The expression construct for *CREBL2* was significantly reduced as a proportion of the pool between the first and second time points, but no more effects were seen after this time, suggesting that this may be an artefact (p=<0.001). The expression constructs for *ETV6* were weakly selected against, with a p-value of 0.027 when analysed using MANOVA. Although *APOLD1* showed some reduction in construct copy number, this was not found to be significant (p=0.344).

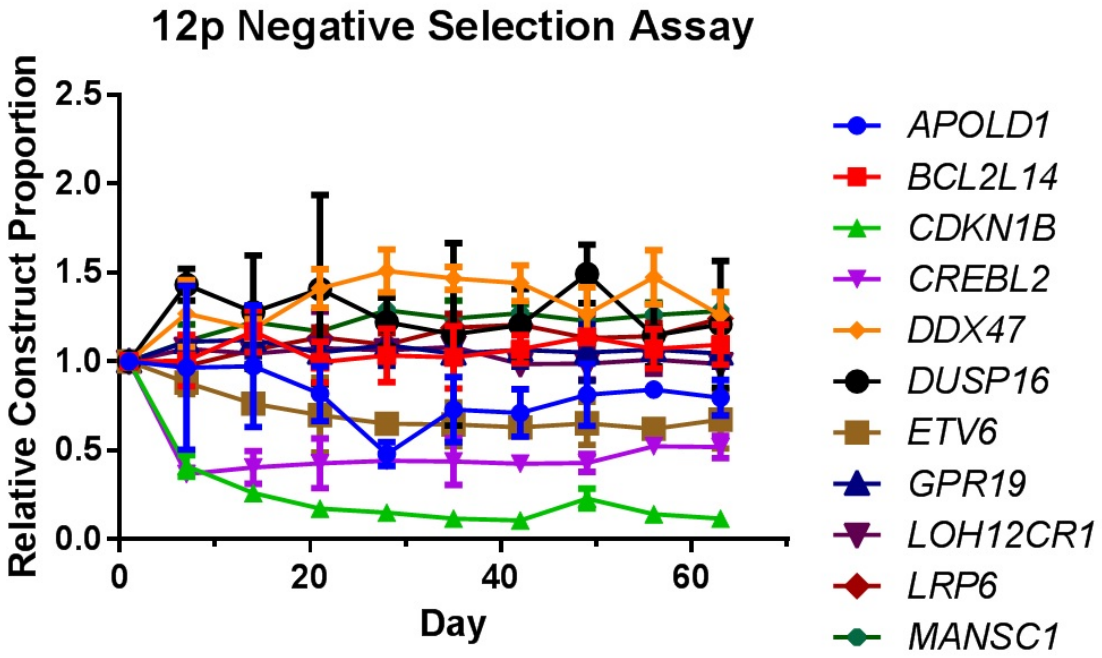


Figure 32. Relative construct proportions during the in vitro portion of the 12p negative selection assay pilot study in the NKM-1 cell line. Error bars represent standard deviation between sequence reads obtained for library generation PCRs performed in triplicate. The expression construct for CDKN1B is strongly and rapidly selected against.

Gene	***CDKN1B***	***CREBL2***	***ETV6***
p-value	< 0.001	0.001	0.027

Table 22. Significant p-values calculated for changes in construct copy number during the in vitro portion of the 12p negative selection assay pilot study in the NKM-1 cell line. Significance and associated p-values were calculated using MANOVA.

4.4.5.2 In Vivo Results

Six mice were intrafemorally injected with NKM-1 cells transduced with the 12p lentiviral particle pool. Two of the injected mice showed no evidence of tumour formation around the bone of the injected femur. The first of these mice (RN) was sacrificed at 32 days after showing signs of difficulty in movement, whilst the second of these mice (2LN) unexpectedly died at 53 days. The remaining 4 mice showed signs of tumour formation, and were culled between 54 and 60 days after injection with cells. After the mice were sacrificed, cells were extracted from the bone marrow, tumours, livers and spleens, and separated using density centrifugation over Ficoll-Paque, before DNA was extracted.

In contrast to the *in vitro* data, relative copy number of expression clones determined by Illumina sequencing was highly variable between samples, as shown in Figure 33. For some samples, a single clone became highly overrepresented, for example in the tumour sample from the LN mouse *DUSP16* appears to be highly overrepresented. In some mice, these results were consistent across all organs sampled, for example the expression clone for *DDX47* was

overrepresented in all tissues of the 2LN mouse. Overall, there were no samples in which the expression clone for *ETV6* was overrepresented when compared to the starting cell population in any of the 22 mouse samples tested. Similarly, the expression clone for *LOH12CR1* was only overrepresented once, and the expression clone for *CDKN1B* was overrepresented twice. All other expression clones were overrepresented in at least 3 of the samples tested.

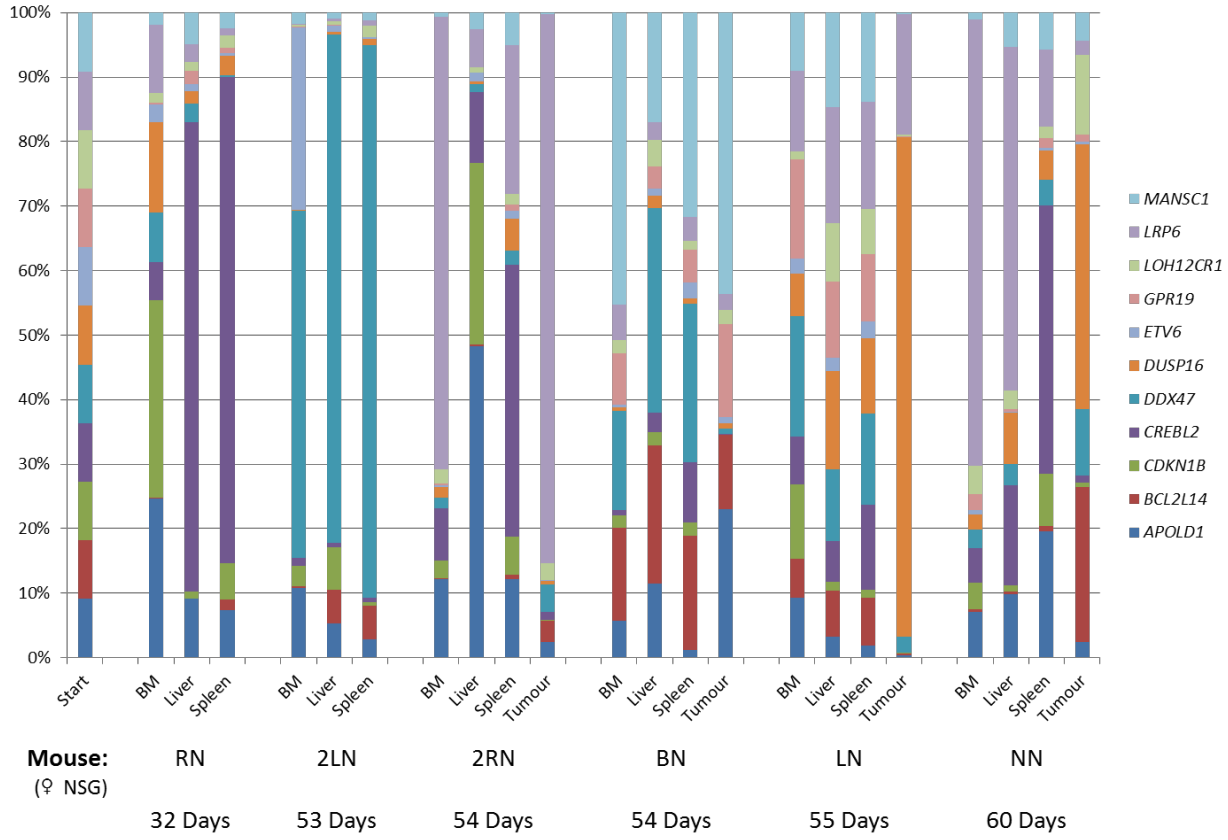


Figure 33. Relative copy number of cDNA constructs integrated in NKM-1 cells before and after expansion in vivo. The number of days refers to the time point at which the mouse was culled.

4.4.6 12p Negative Selection Repeat Studies

The *in vitro* negative selection assay was repeated in triplicate with the NKM-1 cell line. Across all three experiments, the expression clone for *CDKN1B* was strongly and rapidly selected against (p=<0.001, Table 23). The selection seen in the pilot study against expression clones for *APOLD1* and *CREBL2* was not observed in the repeat studies, a weak selection remained against the expression of *ETV6* observed across all three studies (p=0.010).

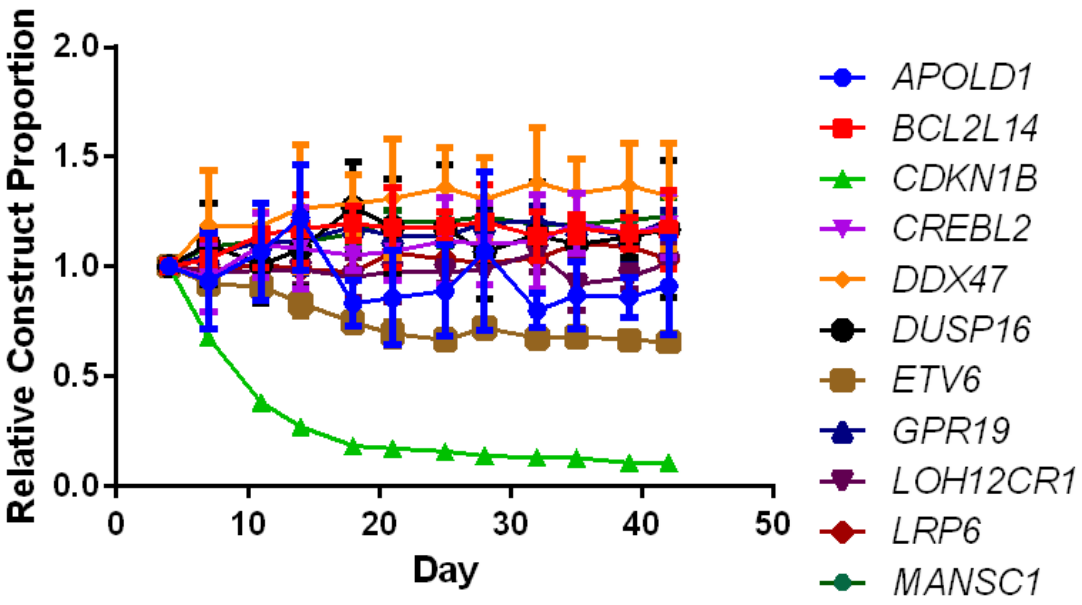


Figure 34. Variation in construct copy number during repeat studies performed in the NKM-1 cell line for the 12p negative selection assay. Figure shown is representative of three experiments. Error bars represent standard deviation between experimental repeats.

Gene	***CDKN1B***	***ETV6***
p-value	< 0.001	0.010

Table 23. Significant changes in construct copy number observed across all experiments in the NKM-1 cell line for the 12p negative selection assay. Significance and associated p-values were calculated using MANOVA.

Triplicate studies were also performed in another AML cell line with a deletion encompassing the minimally deleted region on chromosome 12p, GDM-1. The effect of selection against *CDKN1B* appeared to be weaker in this cell line, although the result was still highly significant (Table 24, p=<0.0001) and there were no significant effects observed in selection against other expression clones (Figure 35).

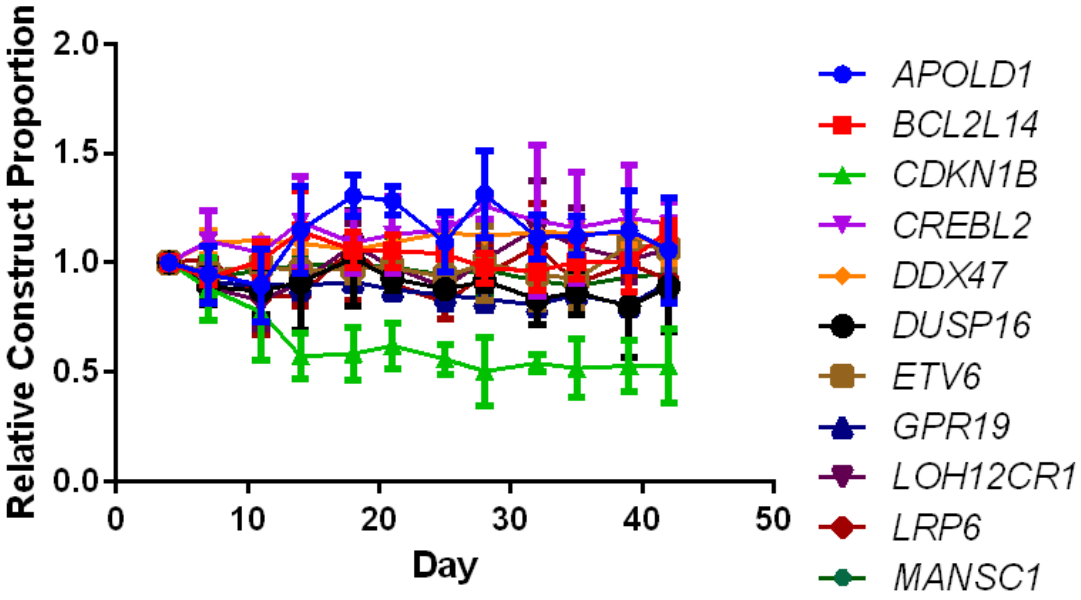


Figure 35. Variation in construct copy number during repeat studies performed in the GDM-1 cell line for the 12p negative selection assay. Figure shown is representative of three experiments. Error bars represent standard deviation between experimental repeats.

Gene	***CDKN1B***
p-value	< 0.001

Table 24 Significant changes in construct copy number observed across all experiments in the GDM-1 cell line for the 12p negative selection assay. Significance and associated p-values were calculated using MANOVA.

4.4.7 11q Negative Selection Studies

4.4.7.1 Assessment of CMV Promoter Silencing

As silencing of the CMV promoter over time is a concern for long term studies using lentiviral transduced populations, a population of UoC-M1 cells was transduced with a non-silencing control shRNA and GFP positivity was measured over a period of 80 days. A high level of GFP positivity in the cell population was reached after 7 days (Figure 36), and this was maintained for over 60 days, suggesting that silencing of the CMV promoter would not be a problem for *in vitro* cultures using the negative selection assay with the GIPZ lentiviral shRNA system.

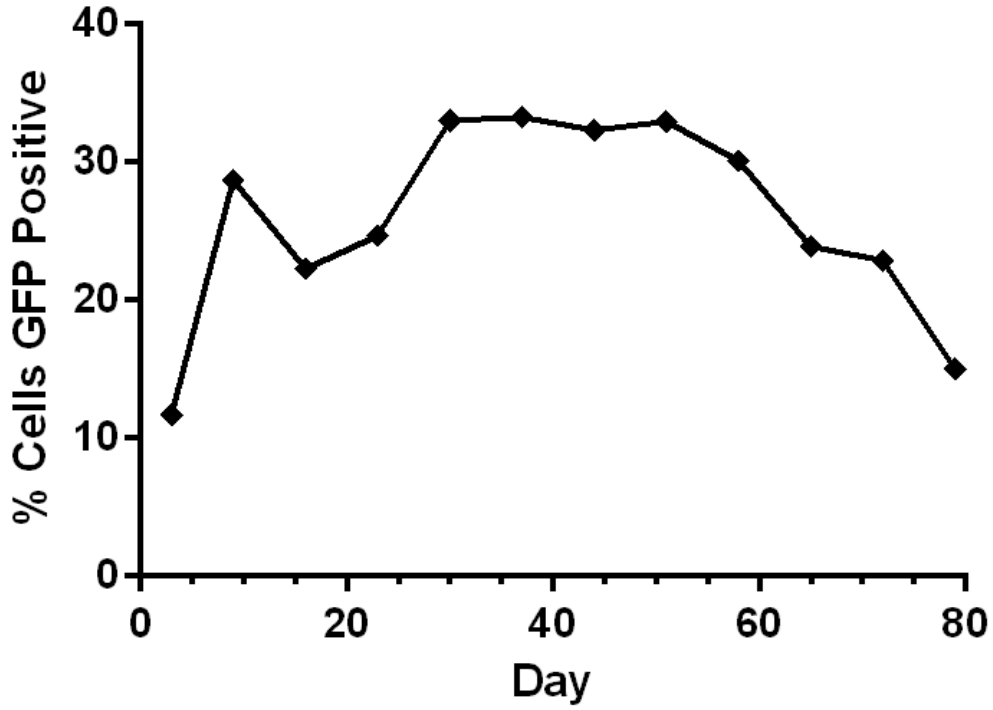


Figure 36. Proportions of GFP positivity over time determined by FACS analysis in a population of UoC-M1 cells transduced with pGIPZ non silencing control lentiviral particles

4.4.7.2 11q Negative Selection Assay

Two methods of GIPZ lentiviral particle library preparation were used. In the first method, each shRNA-containing transfer plasmid was prepared on an individual basis, through culturing picks taken from glycerol stocks overnight on a small scale before extraction with the Plasmid Mini Kit (Qiagen). In the second method, bacterial stocks were cultured in pools containing 10 shRNAs each, across a total of nine pools. Each bacterial pool was cultured on a large scale and extracted using the EndoFree Plasmid Maxi Kit (Qiagen). For both methods, lentiviral particles were produced in pools of 10 shRNAs each and used to transduce $5x10^6$ UoC-M1 cells. All experiments were then repeated twice. Henceforward, the individually prepared clones will be referred to as MinP (Mini prepared), whilst the clones prepared on the large scale will be referred to as MaxP (Maxi prepared).

At multiple time points during the *in vitro* culture, DNA samples were collected and the relative proportions of shRNAs was assessed using high throughput sequencing methods as described above (Section 4.3.8). A list of significantly affected shRNAs was compiled through the use of MANOVA, and two strong hits were identified with significance in both MinP and MaxP samples. In the MaxP samples, *MPZL3* shRNA 1 (p=<0.001) and *MPZL3* shRNA 3 (p=0.035) were selected against. In the MinP samples, shRNA 1 was again significantly selected against (p=<0.001), whilst shRNA 3 was selected against (p=0.050)(Figure 37 and Table 25). The second identified hit was *UBE4A*. In the MaxP samples, shRNA 1 was significantly selected against

(p=0.033), whilst in the MinP samples, shRNAs 1 (p=<0.001) and 3 (p=0.050) were significantly selected against (Figure 38 and Table 26).

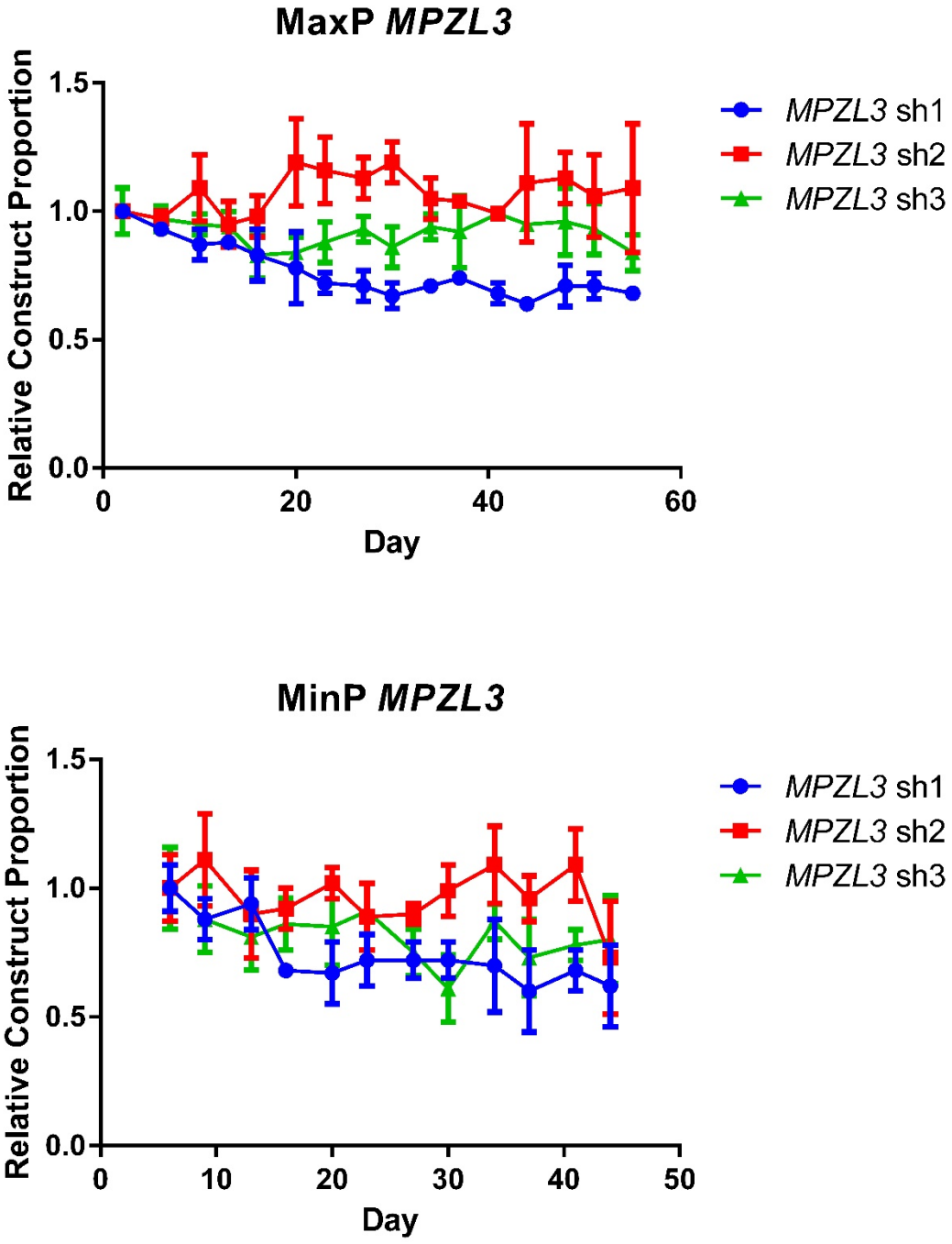


Figure 37. Mean individual results for negative selection using shRNAs targeted to MPZL3. sh1 and sh3 in both the MaxP and MinP samples were identified as hits. Error bars represent standard deviation between experimental repeats.

shRNA	MaxP *MPZL3* sh1	MaxP *MPZL3* sh3	MinP *MPZL3* sh1	MinP *MPZL3* sh3
p-value	< 0.001	0.035	< 0.001	0.050

Table 25. Significant changes in MPZL3 shRNA construct copy number observed in the UoC-M1 cell line for the 11q negative selection assay. Significance and associated p-values were calculated using MANOVA.

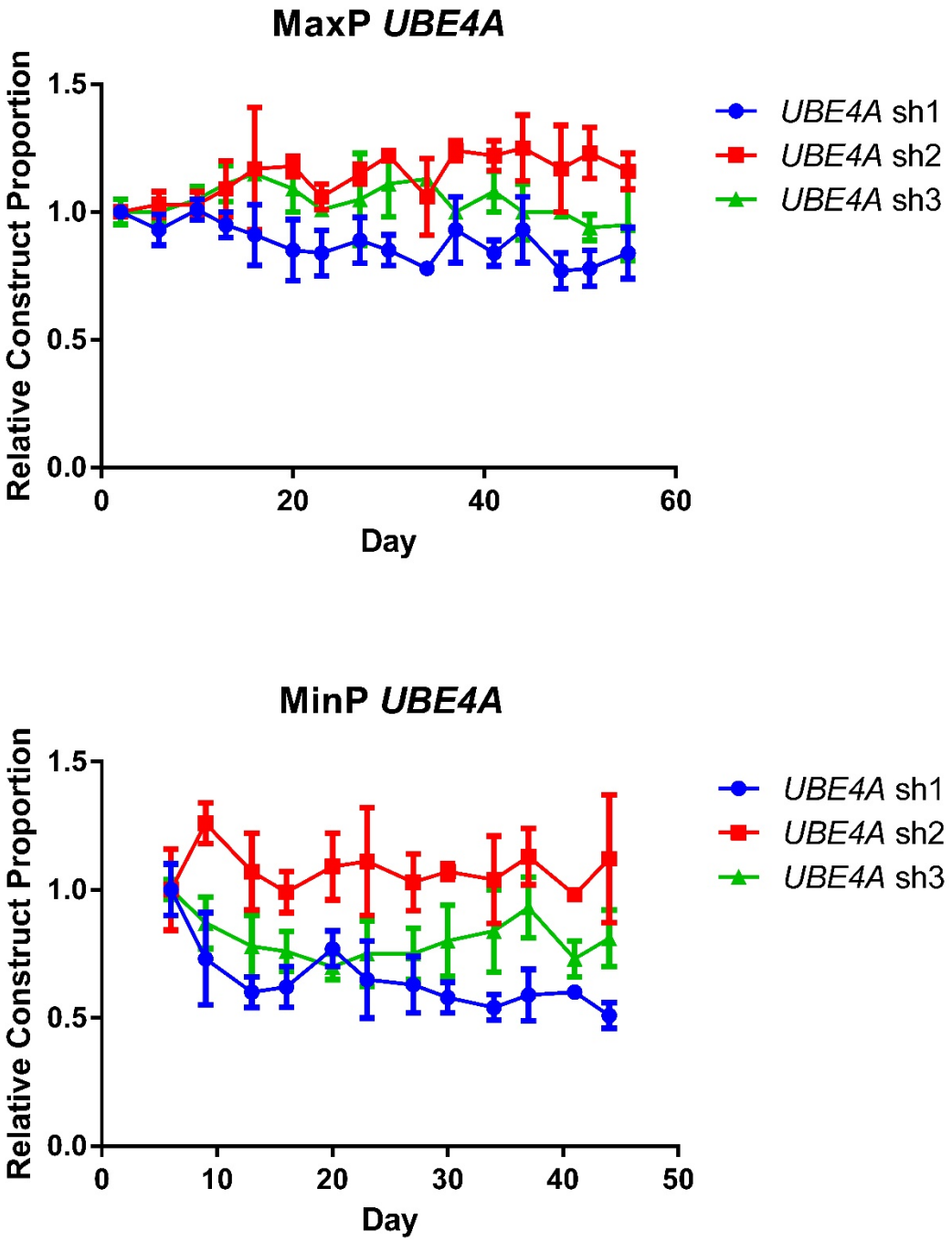


Figure 38. Mean individual results for negative selection using shRNAs targeted to UBE4A. sh1 in both MaxP and MinP samples was identified as a hit, as well as sh3 in the MinP sample.

shRNA	MaxP *UBE4A* sh1	MinP *UBE4A* sh1	MinP *UBE4A* sh3
p-value	0.033	<0.001	0.050

Table 26. Significant changes in UBE4A shRNA construct copy number observed in the UoC-M1 cell line for the 11q negative selection assay. Significance and associated p-values were calculated using MANOVA.

4.5 Discussion

4.5.1 12p Negative Selection Assay

In this section of work, two lentiviral systems for the expression of cDNAs were tested; the FIV-based pReceiver-Lv81 and the HIV-based SIN-SIEW lentivector system. Superior transduction efficiency was demonstrated using a combination of the SIN-SIEW vector with the NKM-1 cell line, therefore these were chosen for use in the first round of negative selection assays to identify tumour suppressor genes from the commonly deleted 12p region. cDNA sequences corresponding to 11 genes from the minimally deleted region established for chromosome 12p were cloned into the SIN-SIEW lentivector using the Gateway cloning method, and inserts were sequenced to confirm their integration and correct alignment.

An optimal method of producing lentiviral particles was formulated and transduction conditions optimised within this system. Through the methodical testing of various conditions, a combination of the 293FT cell line with the EndoFectin reagent was selected. Chloroquine was added post transfection to enhance lentiviral yield, and the transduction protocol was improved with the addition of polybrene and a spinfection step. Next, the PCR conditions for preparing sequencing libraries were carefully optimised to ensure that amplicons were obtained in the logarithmic expansion phase of PCR. Collection in this phase preserves variation in amplicon copy number, which would be diminished if PCR proceeds to the plateau phase of amplification.

In the pilot *in vitro* study for the negative selection assay, expression vectors containing *CDKN1B* were strongly and rapidly selected against. A less obvious negative selection effect was also observed for expression of *ETV6* and *CREBL2* and at this stage, it was hoped that repeating the studies would either clearly eliminate or substantiate these uncertain findings, the results of which are discussed below. The *in vivo* data was largely unclear, with a highly variable pattern representing each clone.

In this project, leukaemic cells were introduced into mice by injection of cell suspension through the head of the femur into the bone marrow cavity. This route of injection was selected as it had been proven to be highly successful for ALL cell lines and primary patient material used in a related project (Bomken *et al.*, 2012). In NSG mice injected with ALL cells, cells engraft in the bone marrow early, before spreading to the spleen and liver as the disease

develops. The spleens and livers of these mice are generally much larger and visible signs of leukaemic engraftment can be observed on the surface of these organs in the form of white spots. In some of these mice, there is also evidence of kidney involvement or central nervous system engraftment (Bomken *et al.*, 2012).

In the model of AML engraftment used in this project, cells appeared to be largely localised to the site of injection. Instead of developing in the bone marrow and then spreading throughout the body, they formed solid tumours around the femur of the mice. Density centrifugation performed by layering organ suspension over Ficoll-Paque demonstrated that there were AML cells present in the liver and spleen of culled mice, and that these cells could be collected and cultured *in vitro*. Despite the relatively low numbers of engrafted cells obtained from these organs, quantification of integrated lentiviral expression clones was possible. However, the data obtained across all mouse samples demonstrated a highly variable pattern of clone representation. In some mice one clone became greatly overrepresented across all organs as well as the tumour sample, whereas in others there were large inter-sample differences. These may have arisen due to a low numbers of cells engrafted into the host mice at the injection site, as well as when spreading to secondary sites. As *in vivo* studies have not, knowingly to myself, been performed using these cell lines, it is unclear whether the time taken for tumours to develop was within a normal range.

Some brain and CNS samples from the mice shown in Figure 31, used to investigate the pattern of engraftment using the IVIS, were sent to Dr. Christina Halsey (University of Glasgow) for investigation as part of a collaborative project. As shown in Figure 39, AML cells were found engrafted in the CNS meningeal tissue, as well as in the brain tissue from the dissected mouse heads. This would suggest that the level of engraftment at the time of sacrifice was greater than expected, and that there was CNS engraftment that was unable to be detected using the IVIS.

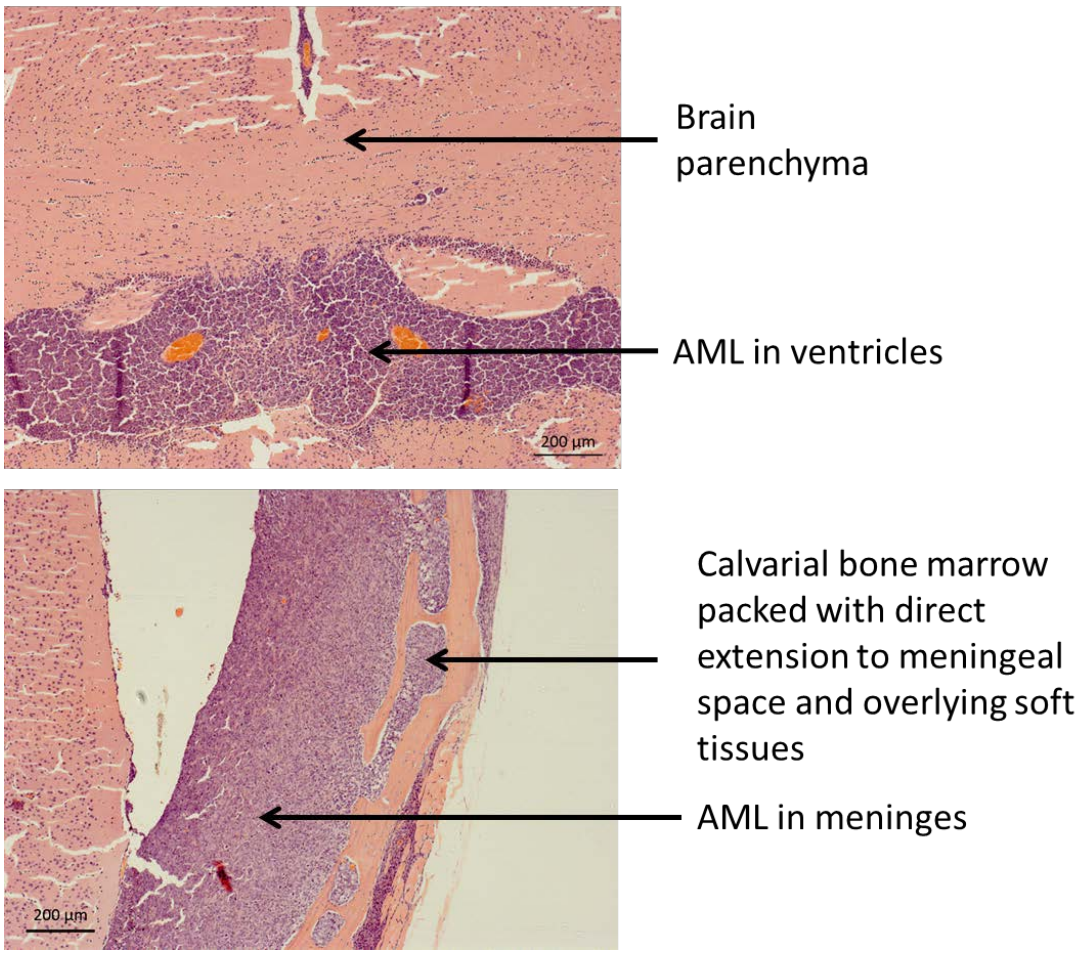


Figure 39. Immunohistochemical analysis of brain and CNS mouse tissues. Sacrificed mouse heads were analysed by Dr. Christina Halsey (University of Glasgow). Pictures shown are from NKM-1 cell line engrafted mouse NN (Figure 31).

Another route of injection may be considered for future studies. Similar selection assays have been performed using intravenous infusion of leukaemic cells, and these have shown success with lentiviral shRNA libraries containing over 1500 shRNAs (Wang *et al.*, 2012; Miller *et al.*, 2013). The study performed by Wang *et al.* (2012) used HSC populations isolated with a cocktail of cell surface antigen markers. A positive selection approach was taken in looking for the enrichment of shRNAs. The second study (Miller *et al.*, 2013) transduced a population of leukaemic stem cells originally generated by the introduction of a lentiviral vector driving the expression of *MLL-AF9* in granulocyte-macrophage progenitor cells isolated from mice, in conjunction with negative selection of shRNAs. Although the cell types used in these examples were different to those used in this project, similar success may be achieved by using intravenous delivery of leukaemic cell lines. A concern was raised that human AML cell lines injected intravenously would form systemic tumours throughout the mouse, which may be difficult to recover (Meyer and Debatin, 2011; Theocharides *et al.*, 2012).

Due to the difficulties experienced with the *in vivo* portion of the pilot study, experimental results were consolidated by performing additional *in vitro* studies in the same and a second cell line. In the NKM-1 cell line, *CDKN1B* was again strongly selected against across all three repeats. The weak negative selection effects that were found in the pilot study against *CREBL2* were not observed, suggesting that these were false positive results eliminated by the introduction of experimental replication. However, the weak selection against *ETV6* in this cell line was reproduced. It is possible that expression of *ETV6* does have a negative selection effect, but that in this model of AML expansion only highly expressing clones are selected against, whilst low and medium expression of *ETV6* is tolerated. The role of *ETV6* in haematopoiesis may also be highly relevant for this investigation. In primary AML samples, the loss of *ETV6* may block differentiation and contribute to leukaemogenesis through this mechanism. However this may not be as relevant in established cell lines that have been cultured for many years and may have acquired additional blocks to differentiation. For this reason, it would be of interest to assess the functional consequences of overexpression of *ETV6* in a range of cell lines and primary patient samples.

In the experiments performed using GDM-1, selection against *CDKN1B* was less pronounced but remained highly significant ($p=<0.001$), with the relative proportion of integrated *CDKN1B* construct dropping by around 50% compared to the starting time point. Perhaps due to the smaller effects observed in these samples, it was also not possible to observe any clear weak effects from other expression constructs. The reduced degree of negative selection observed in this cell line, by comparison with NKM-1 is likely due to a relatively lower overall expression from integrated SIN-SIEW constructs in GDM-1 cells, as determined by analysis of GFP levels in early optimisation experiments using the pSLIEW construct (Figure 25). Despite these lower expression levels, clear confirmation of the negative selection effect of *CDKN1B* expression in another cell line was important, supporting the hypothesis that *CDKN1B* is the main target of 12p deletions in AML.

Within the minimally deleted region identified on 12p, there are two miRNAs – miR-613 and miR-1244-1. As the function of these two miRNAs is unknown, it would be interesting to study their function and possible roles in leukaemogenesis.

4.5.2 11q Negative Selection Assay

In this section of work, two methods for producing large pools of lentiviral particles containing expression constructs for shRNAs were assessed. Both methods produced lentiviral particle pools containing 90 shRNAs, however individual constructs were either pooled at the lentiviral particle production stage (MinP) or at the bacterial culture stage (MaxP). For the assessment of

potential hits, results from both sets of triplicate experiments were analysed together. Two novel candidate oncogenes from this region were identified, *MPZL3* and *UBE4A*. Two out of the three shRNAs targeted to *MPZL3* were identified as significant hits in both the MinP and MaxP experiments, whilst one of the shRNAs targeted to *UBE4A* was identified as a significant hit in both experiments, and an additional shRNA was identified as a hit in the MinP shRNAs only. The potential roles of these genes in leukaemogenesis will be discussed further in Chapter 5.

The changes in relative construct proportion compared to the start point of this experiment were smaller compared to those seen in the 12p negative selection assay performed in the NKM-1 cell line. The largest change observed was for a shRNA targeted to *UBE4A* in the MiniP samples, which was reduced to 51% of its starting copy number proportion. Expression levels of shRNAs may have contributed to this observation, as the mean levels of GFP positivity attained in the UoC-M1 cell line with GIPZ lentiviral vectors were relatively low compared to those obtained in the NKM-1 cell line for the 12p project. The negative effects on growth mostly took place in the first 21 days of *in vitro* culturing (Figure 37, Figure 38), supporting the idea that a proportion of integrated shRNA constructs, with expression below a certain threshold, were tolerated by the cells in culture and not subject to negative selection.

Perhaps unexpectedly, neither *FLI1* nor *MLL* were identified as hits in this screen. *FLI1* was a strong candidate proto-oncogene contributing to AML, based on functional analysis of its role in erythroleukaemias, where inhibition lead to the suppression of growth and induction of cell death (Cui *et al.*, 2009). In the negative selection assay, one of the shRNAs targeted to *FLI1* was significantly selected against in both MaxP and MinP samples (Figure 56). Although this was not identified as a top hit in the assay, it may still be of interest for future studies as not all GIPZ shRNAs are validated as knock down reagents. *MLL* is a gene of interest due to its role in many leukaemias in which the *MLL* gene is involved in translocations leading to its constitutive activation through expression of fusion transcripts (Zeisig *et al.*, 2003). As shRNAs targeted to *MLL* were not selected against in this assay, this would suggest that its overexpression in 11q amplified AMLs does not contribute to leukaemogenesis.

Chapter 5 Functional Assessment of Candidate Driver Genes

5.1 Introduction

Three candidate genes of interest have been identified through the determination of minimal regions of copy number alteration followed by the use of a functional negative selection assay. One of these genes is located on chromosome 12p (*CDKN1B*), whilst the remaining 2 genes are located on chromosome 11q (*MPZL3* and *UBE4A)*. The introduction to this chapter will focus on a review of the current literature surrounding these 3 genes.

5.1.1 Cyclin-dependent Kinase Inhibitor 1B (*CDKN1B*)

Cyclin family proteins control the progression of cells through the cell cycle via the activation of cyclin-dependent kinase enzymes. There are two main classes of cyclin; G1/S phase cyclins, that are required for the transition of cells between the G_1 and S phases of the cell cycle and the G_2/M phase cyclins, which are responsible for the transition of cells between G_2 and M phases. The *CDKN1B* gene encodes a 27kDa protein that prevents the activation of two G_1/S cyclins; cyclin E, which forms a complex with CDK2; and cyclin D, which forms a complex with CDK4 or CDK6. Through the direct binding and inhibition of these complexes or their individual components, *CDKN1B* prevents phosphorylation of the Rb protein, which in turn prevents Rb from interacting with E2F-DP transcription factor complexes, halting the cell cycle (Calzone *et al.*, 2008). A schematic of this is shown in Figure 40.

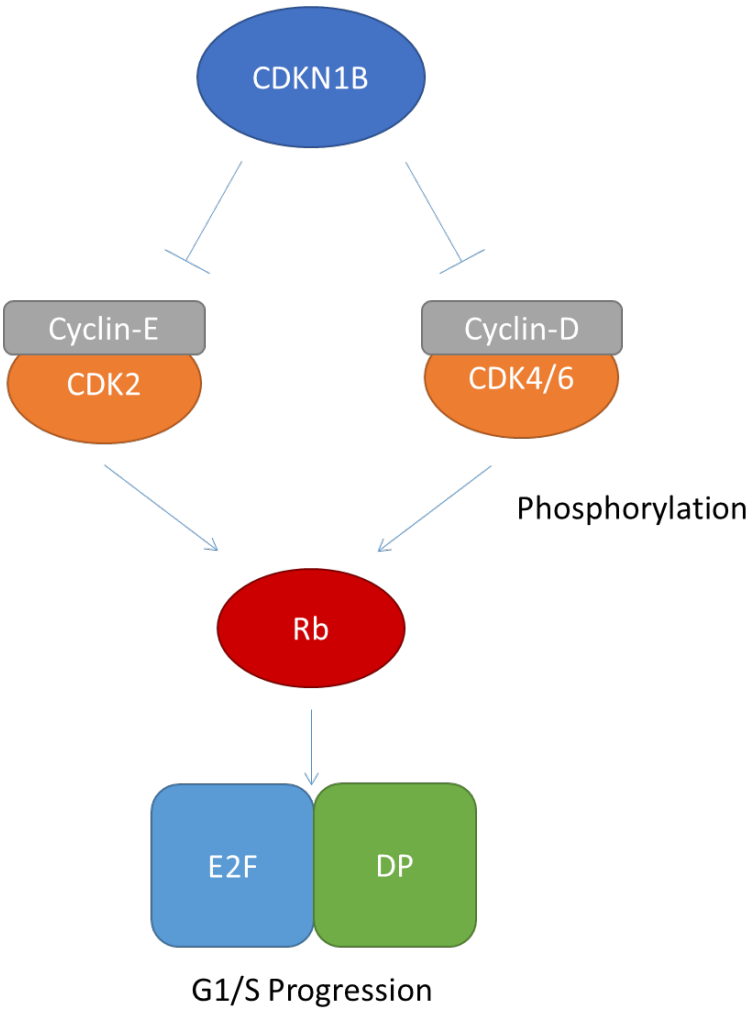


Figure 40. The primary and most well studied role of CDKN1B in G_1/S progression. Arrows represent an interaction, whilst flat headed lines represent an inhibition. CDKN1B is able to bind Cyclin-E/CDK2 complexes, Cyclin-D/CDK4/6 complexes, or their individual components to inhibit the ability of these complexes to phosphorylate Rb, halting the cell cycle at G_1/S phase.

There are multiple regulatory systems affecting the activity of *CDKN1B* and its associated gene products. The transcription of *CDKN1B* is regulated by Forkhead box class O (FOXO) transcription factors, which activate *CDKN1B* transcription in response to Akt signalling, cytokines or PML protein (Dijkers *et al.*, 2000; Medema *et al.*, 2000). The *MYC* transcription factor represses the *CDKN1B* promoter (Yang *et al.*, 2001), and induces the expression of proteins that inhibit CDKN1B such as cyclin D1, as well as the CUL1 and CKS1B proteins that proteolyse CDKN1B (O'Hagan *et al.*, 2000; Keller *et al.*, 2007).

CDKN1B translation is regulated by a number of proteins. PTBP1 binds *CDKN1B* transcripts and increases their translation (Cho *et al.*, 2005), whilst ELAVL1 and ELAVL4 bind transcripts and inhibit their translation (Kullman *et al.*, 2002). ELAV1 overexpression in ovarian and colon cancer cells has been reported, which may drive the reduction of CDKN1B levels observed in these cancer types (Lopez de Silanes *et al.*, 2003; Denkert *et al.*, 2004). Two miRNAs, miR-221

and miR-222, have been shown to bind to *CDKN1B* transcripts and subsequently inhibit their translation. High levels of these two miRNAs have been found in cancer cells, and their downregulation has been linked to an increase in CDKN1B and subsequent inhibition of tumour growth (le Sage *et al.*, 2007).

FOXO4 protein inhibits AKT to promote the nuclear localisation of CDKN1B, encouraging its cell cycle inhibitory function. Furthermore, FOXO4 acts to decrease the expression of *COP9S5*, which is translated into a protein mediator of CDKN1B degradation (Yang *et al.*, 2005). Another FOXO class member, FOXM1 also increases the stability of CDKN1B (Wang *et al.*, 2002). The proteolysis of CDKN1B is also controlled by the ubiquitination activity of the SCF complex (Bloom and Pagano, 2003), as well as a complex made up of SKP2, CUL4A and DDB1, which also marks CDKN1B for degradation by ubiquitination (Bondar *et al.*, 2006).

In cancer cells, *SRC* family members often show signs of upregulation, and this has been linked to EGFR and ERBB2 activity, both of which bind SRC to catalyse its kinase activity. The activation of SRC has been linked to a loss of CDKN1B during G1 phase progression and this is reflected in the significant correlation of *ERBB2* amplification, SRC activation, and reduced CDKN1B protein levels in breast cancers (Spataro *et al.*, 2003; Chu *et al.*, 2007). The overexpression of *EGFR* and *ERBB2* is not restricted to breast cancers only, suggesting that other epithelial cancers may downregulate CDKN1B levels in this manner to contribute to tumourigenesis (Arteaga *et al.*, 2003).

In leukaemias and lymphomas, activation of ABL, LYN, LCK and FYN have been linked with the increased proteolysis of CDKN1B (Moller *et al.*, 1999; Tsihlias *et al.*, 1999). CML cells expressing the *BCR-ABL1* fusion gene have been shown to have reduced levels of CDKN1B, which is reversed by *BCR-ABL1* inhibition using imatinib (Andreu *et al.*, 2005). This mechanism has been shown to be due to a number of interactions between *BCR-ABL1* and *CDKN1B*. *BCR-ABL1* downregulates *CDKN1B* through FOXO inhibition, as well as through *SKP2* upregegulation leading to increased CDKN1B proteolysis (Andreu *et al.*, 2005). BCR-ABL1 also directly phosphorylates CDKN1B on tyrosine-88, preventing it from functioning in CDK2 inhibition (Grimmler *et al.*, 2007).

5.1.1.1 Prognostic Indication

Decreased expression levels of *CDKN1B* have been associated with poor prognosis and disease progression in lung cancer (Esposito *et al.*, 1997; Catzavelos *et al.*, 1999; Hayashi *et al.*, 2001), head and neck cancer (Korkmaz *et al.*, 2005; Langer *et al.*, 2006) and colorectal cancer (Loda *et al.*, 1997; Manne *et al.*, 2004; Rosati *et al.*, 2004). The most extensive investigations have taken place in breast cancers, with multiple studies reporting that reduced CDKN1B levels are of

prognostic significance, with low CDKN1B levels correlating strongly with poor survival (Gillett *et al.*, 1999; Barnes *et al.*, 2003; Pohl *et al.*, 2003; Porter *et al.*, 2006).

In AML, a study of *CDKN1B* expression analysis of 286 patients found that those patients with low *CDKN1B* expression were more likely to carry certain abnormalities, including *RUNX1-RUNX1T1*, *PML-RARA* rearrangements, *MLL* rearrangements and *FLT3*-TKD mutations. The study also found that patients with low *CDKN1B* expression generally had a longer overall survival when compared to patients with intermediate or high expression levels (Haferlach *et al.*, 2012).

5.1.1.2 Cytoplasmic CDKN1B

A possible alternative role for *CDKN1B* has recently been well documented by the study of CDKN1B localised to the cytoplasm. Cytoplasmic CDKN1B has been shown to co-localise with actin fibres and stimulate cellular migration. This was demonstrated in *Cdkn1b* null mouse embryonic fibroblasts, which showed decreased cell motility compared to wild type fibroblasts (Besson *et al.*, 2004). A similar loss of motility was seen in cancer cell lines with low versus high cytoplasmic CDKN1B. Through the expression of CDKN1B protein with a defective localisation signal, excess CDKN1B accumulated in the cytoplasm. In MCF-7 breast cancer cells, this led to an increase in cell motility (Wu *et al.*, 2006). Similarly, expressing these defective proteins in melanoma increased cell motility and led to an increased incidence of metastasis *in vivo* (Denicourt *et al.*, 2007). A mouse model carrying mutant *Cdkn1b* that lacks nuclear cyclin inhibitory functions has a higher rate of spontaneous tumour development than *Cdkn1b* null mice (Serres *et al.*, 2011).

In addition to the role of *BCR-ABL1* in regulating CDKN1B activity as described above, it has been discovered that BCR-ABL1 can enhance CDKN1B cytoplasmic localisation in CML cells (Chu *et al.*, 2010). A study by Agarwal *et al.* (2014) focussed on the role of cytoplasmic CDKN1B in CML cells, as CDKN1B is primarily cytoplasmic in the majority of CML cases studied. By knocking down *CDKN1B* in CML cells that had primarily cytoplasmic CDKN1B, induction of apoptosis was observed that did not occur in those cells with largely nuclear CDKN1B. In similar work to that conducted by Serres *et al.* (2011), they also demonstrated that a murine CML model with *Cdkn1b* that lacked nuclear cyclin inhibitory functions enhanced leukaemogenesis above the levels observed in a murine CML model that was null for *Cdkn1b*. Finally, expression of two types of mutant *Cdkn1b* in a murine CML model, one with enhanced stability and another with increased nuclear retention, both showed attenuation of leukaemogenesis when compared to wild type *Cdkn1b* models (Agarwal *et al.*, 2014).

5.1.2 Myelin Protein Zero-like 3 (*MPZL3*)

The *MPZL3* gene was first identified as the mutated gene leading to the rough coat (*rc*) mutation in C57BL/6J mice. This mutation arose spontaneously in 1966 and affected mice had unkempt hair coats at weaning age, later developing into progressive hair loss (Dickie, 1966). Through the use of linkage analysis, the *rc* mutation was mapped to mouse chromosome 9, near to the *Mpi1* gene (Eicher *et al.*, 1977). Following these findings, the mutant strain was not further studied until the early 2000s. Ruvinsky *et al.* (2002) described a novel mutation in the mouse Brachyury gene that led to a similar phenotype to that of *rc*, however the authors found that the two mutations occurred in different alleles (Ruvinsky *et al.*, 2002).

Soon after this publication, the *rc* mutation was studied in detail by Hayashi *et al.* (2004), where they examined the effects of *rc* across a wide range of tissues. They demonstrated that *rc* mice showed a significant reduction in growth when compared to normal C57BL/6J littermates, with this difference progressing further with age. The hearts of *rc* mice were found to have multiple myocardial degeneration foci, with disoriented cardiac muscle fibres. They also contained reduced amounts of collagen and elastin, which was linked to the presence of distorted blood vessels with thinner walls. The bone marrow of *rc* mice showed diffused hyperplasia, and a lack of calcium deposits normally seen in the femurs of wild type mice. In the liver, *rc* mice had increased numbers of erythrocytes, lymphocytes, polymorphonuclear leukocytes and enlarged Kupffer cells in the hepatic sinusoids compared to normal mice. Finally the spleens of *rc* mice were shown to contain an increased number of macrophages containing hemosiderin granules (Hayashi *et al.*, 2004).

The *rc* mutation was successfully mapped to a gene locus through positional cloning, allowing researchers to identify a 246kb minimal region to sequence. A point mutation in a theoretical coding gene was identified and the gene was named *Mpzl3* due to its homology to the *MPZL2* gene. The mutation was found to occur in the highly conserved Immunoglobulin-like V-type domain within the coding sequence and was predicted to alter the function of MPZL3 protein through the substitution of an arginine to glutamine at position 100 of the 237 amino-acid polypeptide sequence (Cao *et al.*, 2007).

To determine the cause of reduced body weight in *rc* mice, Czyzyk *et al.* (2013) set out to investigate a possible link between metabolism and *Mpzl3*. They generated a *Mpzl3* knockout mouse model and found that these mice were hyperphagic yet had reduced body weight and fat mass, caused by a whole body increase in energy expenditure. The muscles of knockout mice were found to have increased oxidative capacity and contractile force, whilst their livers had reduced levels of triglycerides and increased Akt signalling stimulated by insulin. Overall,

this strongly suggested that in mice at least, *Mpzl3* is important in regulating metabolism and energy expenditure (Czyzyk *et al.*, 2013). Further studies of *Mpzl3* knockout mice by Leiva *et al.*, (2014) have focussed on the skin abnormalities seen in these mice. Through the transplant of *Mpzl3* -/- HSCs into immune deficient C57BL/6 Rag -/- mice, they demonstrated that *Mpzl3* -/- HSCs remained able to fully repopulate the bone marrow of the deficient mice, which did not result in replication of the abnormal skin phenotypes seen in the *Mpzl3* knockout mice. The group were able to replicate the reduced body weight seen in the Czyzyk *et al.* (2013) study and suggested that this was due to the lack of MPZL3 function leading to the suppression of hepatic lipid accumulation or increased levels of fat burning (Leiva *et al.*, 2014).

5.1.3 Ubiquitination Factor E4A (*UBE4A*)

UBE4A was first identified as the human homologue of the yeast *UFD2* gene (Koegl *et al.*, 1999). Ubiquitination is the process that results in the labelling of proteins for breakdown by the proteasome system. Substrates of the ubiquitin/proteasome system are earmarked for degradation by the conjugation of ubiquitin. Conjugation of multiple ubiquitin molecules is common (multiubiquitination) and substrates with many ubiquitin molecules are preferred by the proteasome, whilst proteins with only a single or a few ubiquitin molecules tend to be long lived or are degraded by alternative degradation pathways (Komander and Rape, 2012).

The ubiquitination system relies on multiple classes of enzymes. The ubiquitin-activating enzyme E1 activates ubiquitin molecules through the use of ATP hydrolysis. The activated ubiquitin molecules are then processed by ubiquitin-conjugating E2 enzymes, which further activate ubiquitin and either attach activated molecules to substrate proteins, or pass them on to E3 ubiquitin-protein ligases for attachment to target proteins (Komander, 2009). The E4 class of ubiquitination factors elongate ubiquitination chains in conjunction with E1, E2 and E3 enzymes to produce long ubiquitin chains and promote degradation of target proteins (Metzger and Weissman, 2010). *UFD2* and its human homologues *UBE4A* and *UBE4B* fit within this E4 class of ubiquitination enzymes (Hänzelmann *et al.*, 2010).

Both *UBE4A* and *UBE4B* lie within a commonly deleted regions in neuroblastoma, on chromosomes 11q and 1p respectively. Mutations in both of these genes identified in neuroblastoma samples are believed to be inactivating (Carén *et al.*, 2006), suggesting that in this context, both *UBE4A* and *UBE4B* function as tumour suppressor genes. A case of a *MLL-UBE4A* fusion has also been identified in an ALL. In this case, the fusion resulted in a head to head translocation of the *MLL* and *UBE4A* genes and was presumed to result in the effective loss of heterozygosity for the *UBE4A* gene (Meyer *et al.*, 2009).

The expression of *UBE4A*, as well as other ubiquitination factors (*USP2* and *USP14*) has been shown to be elevated in ovarian serous cystadenocarcinoma compared to adjacent normal tissues, suggesting that ubiquitinative activity is increased in ovarian cancer cells (Yang *et al.*, 2007). The expression of the E4 ubiquitin enzyme *UFD2* in yeast has been shown to be beneficial in high stress situations suggesting that it is involved in the degradation of stress-induced aberrant proteins (Koegl *et al.*, 1999). If the same were true of the human homologues of *UFD2*, this could explain the increased levels of expression seen in ovarian cancer cells, which are likely to experience adverse extracellular conditions.

5.2 Methods

5.2.1 Flow Cytometry

5.2.1.1 Cell Cycle Analysis

For the analysis of cell cycle distribution, the FxCycle PI/RNase staining solution (Life Technologies) was used. The Propidium Iodide (PI) in this solution is a fluorescent DNA stain which when excited, emits an emission spectrum proportional to DNA mass. The presence of RNase degrades RNA to prevent PI staining of RNAs present in the cell. A frequency histogram is plotted for signal intensity, providing information on the number of cells in G_0/G_1 phase (a single pair of chromosomes per cell), S phase (DNA synthesis stage with variable DNA content) and G_2/M phase (two sets of paired chromosomes per cell).

Cells were harvested by pelleting of cell suspensions using centrifugation at 400 x *g*, and washed in PBS. After pelleting a second time, cells were resuspended in PBS at a concentration of 1 x 10^6 cells/ml. An aliquot of 1ml was prepared and 3ml of cold ethanol chilled to -20°C was added dropwise whilst the cell suspension was mixed by pulse vortexing. The cells were then incubated for at least 1 hour at 4°C to complete the fixation process.

Fixative was removed through pelleting of the cell suspension by centrifugation at 1200 x *g*. This wash step was repeated to ensure that no fixative remained, and the cell pellet was resuspended in 500µl of FxCycle PI/RNase staining solution. The samples were then incubated in the dark for 30 minutes at room temperature before cytometric analysis using the FACS Calibur (BD Biosciences). Samples were excited at 488nm and emission was detected at 585nm on the FL2 channel.

5.2.2 Reverse Transcription Polymerase Chain Reaction

5.2.2.1 cDNA Generation

cDNA was reverse transcribed from RNA using the High Capacity RNA-to-cDNA kit (Life Technologies). 1µg of total RNA was used per reaction and reactions were prepared as described in Table 27.

Reagent	**Volume**
Reverse Transcription Buffer	5µl
Reverse Transcription Enzyme Mix	0.5µl
RNA	1µg in up to 4.5µl
H_2O	Quantity sufficient to 10µl total volume

Table 27. RNA to cDNA reverse transcription reaction mixture.

Samples were loaded into a thermal cycler (Applied Biosystems) and the program was run as described in Table 28.

Step	Temperature	Time (Minutes)
Reverse Transcription	37°C	60:00
Denaturation	95°C	05:00
	4°C	*Hold*

Table 28. RNA to cDNA reverse transcription thermocycler settings.

5.2.2.2 Quantitative Polymerase Chain Reaction

Quantitative Polymerase Chain Reaction (qPCR) was used to assess relative levels of RNA transcripts present in samples. SYBR Green PCR Master Mix (Life Technologies) was used for qPCR reactions. The SYBR Green dye preferentially binds double stranded DNA, producing a fluorescent DNA-dye complex which, when excited, emits an emission spectrum with a peak of 520nm. This allows SYBR Green to be used to quantify the amount of double stranded DNA when performing qPCR, and this signal can be continuously monitored in real time, as thermal cycling and the DNA amplification process proceeds.

A master mix was prepared using the reagents and volumes per reaction listed in Table 29. The master mix was added to 4ng RNA equivalent of sample cDNA in a 384 well plate. A standard curve was prepared for each primer pair present on the plate using qPCR Human Reference cDNA, Random-primed (Clontech), containing the following range of cDNA amounts; 10ng, 5ng, 2.5ng, 1ng, 0.5ng and 0.25ng. qPCR was performed using a ViiA 7 Real-Time PCR System (Life Technologies) on the standard Real Time PCR setting. All experiments were performed in triplicate, and obtained mean values were plotted onto a standard curve obtained for each primer pair using the reference samples. This allowed relative expression between experimental and reference samples to be calculated.

Reagent	Volume
SYBR Green PCR Master Mix	5.1µl
Forward Primer (10µM)	0.2µl
Reverse Primer (10µM)	0.2µl
H_2O	0.5µl

Table 29. qPCR master mix for determining relative transcript abundance.

5.2.2.3 Primer Optimisation

Before performing qPCR on samples, primer pairs were used to amplify cDNA from control Human Reference cDNA (Clontech), alongside a panel of primers targeted to housekeeping genes (*GAPDH, ALAS, GUSB* and *TBP*). A housekeeping gene was then selected that closely

matched the amplification efficiency of the target gene and was used in subsequent experiments to normalise target cDNA quantity between samples.

5.2.3 Protein Extraction

Protein was extracted from cultured cells using urea buffer (90% 8.8M Urea, 2% 5M NaH_2PO_4, 8% 1M Tris pH 8.0). Urea acts as a protein denaturant, breaking disulphide bonds and increasing protein solubility, allowing protein concentration to be quantified and subsequently used for downstream applications requiring accurate amounts of protein.

Up to 1 x 10^7 cells were pelleted by centrifugation at 400 x *g* for 5 minutes, and resuspended in 50µl ice-cold urea buffer. The sample was then pulse vortexed for 15 seconds, and incubated at 4°C for 30 minutes before centrifugation at 15000 x *g* in a microcentrifuge chilled to 4°C for a further 30 minutes. The supernatant contained solubilised proteins and was stored long term in a freezer at -80°C.

5.2.4 Pierce BCA Assay

To quantify the amounts of solubilised proteins, the bicinchoninic acid (BCA) Protein Assay kit was used (Thermo Scientific). This kit is a colourimetric assay used to estimate protein concentration through comparison to a diluted bovine serum albumin (BSA) standard. The formation of Cu^2+-protein complexes under alkaline conditions leads to the production of Cu^1+ ions, through the reduction of Cu^2+. BCA forms a purple coloured complex with Cu^1+ under the alkaline conditions present, and this colorimetric change can be detected through the measurement of absorbance at 562nm.

A series of standards were prepared for protein quantification. A stock solution of BSA provided with the kit (2000µg/ml) was serially diluted in PBS as detailed in Table 30.

Dilution	Volume of PBS (µl)	Volume of BSA (µl)	Source of BSA	Final Concentration (µg/ml)
1	0	30	Stock	2000
2	12.5	37.5	Stock	1500
3	32.5	32.5	Stock	100
4	17.5	17.5	Dilution 2	750
5	32.5	32.5	Dilution 3	500
6	32.5	32.5	Dilution 5	250
7	32.5	32.5	Dilution 6	125
8	40	10	Dilution 7	25
9	40	0	N/A	0

Table 30. Standards produced for protein quantification using the Pierce BCA assay.

A working reagent was prepared by mixing 200µl BCA Reagent A with 4µl BCA Reagent B for each sample to be measured. 25µl of each standard and 25µl of protein sample (diluted 1:5 with PBS) was loaded into individual wells of a clear 96 well plate. 200µl of working reagent was added to each well and the plate was shaken for 30 seconds using a plate shaker. The plate was sealed with Parafilm and incubated at 37°C for 30 minutes. After incubation, the absorbance of each well was measured at 562nm using a Fluostar Omega plate reader (BMG Labtech).

5.2.5 Western Immunoblotting

In Western immunoblotting, proteins are first separated by molecular weight and 3-D structure by polyacrylamide gel electrophoresis. The proteins are then transferred to a membrane where they can be probed using antibodies specific to the protein of interest. The antibodies used are conjugated to enzymes or fluorophores that can be detected to show the physical location and intensity of a protein band present on the membrane.

Buffer	Components
10x Running Buffer	30.3g Tris(hydroxymethyl)aminomethane (Tris)
	144g Glycine
	10g Sodium Dodecyl Sulphate (SDS)
10x Transfer Buffer	144g Glycine
	30.3g Tris
10x Tris Buffered Saline-Tween-20 (TBS-T)	12.11g Tris
	29.22g Sodium Chloride
	10ml Tween-20
All buffers were made up to 1L with distilled water for storage and diluted 1:10 with distilled water when required for use.	

Table 31. List of buffers used in Western immunoblotting.

5.2.5.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

In order to separate whole protein lysates by size, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used. 25µg of protein was made up to 30µl with fresh urea buffer. 5µl of 6x loading buffer (Cell Signalling Technologies) was added to each sample and briefly mixed by pulse vortexing. Samples were then heated to 96°C for 5 minutes in a thermocycler and subsequently kept on ice until ready to load. An Amersham ECL Gel (GE Life Sciences) was rinsed in dH_2O and placed into an Amersham ECL Gel Box (GE Life Sciences), filled with 180ml 1x running buffer. The gel was pre-run for 12 minutes at 160V and the well comb removed. The wells were then filled with 6ml of 1x running buffer and 35µg of protein samples loaded into each well. Into one empty well of each gel, 10µl of Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific) was loaded. Remaining empty wells were filled with 30µl of urea buffer mixed with 5µl of loading buffer and when loading was complete, the gel was run at 160V for 60 minutes.

5.2.5.2 Protein Transfer

A square of PVDF membrane (Merck Millipore) approximately equal to the gel size was cut and activated by placing into 20ml of 100% methanol. The membrane was then transferred to 10ml of 1x transfer buffer to equilibrate for at least five minutes.

After the gel had completed running, it was removed from the cassette and the stacking gel was cut away. The gel was placed onto a nitrocellulose blotting paper that had been pre-soaked in 1x transfer buffer. The activated PVDF membrane was placed on top of the gel, and to this another pre-soaked piece of nitrocellulose blotting paper was added. This assembly was placed into a transfer cassette and inserted into a protein transfer tank filled with 1L of 1x

transfer buffer. An ice pack was added to the tank to maintain a low temperature during the transfer process, and proteins were transferred by running the equipment at 100V for 60 minutes.

5.2.5.3 Immunoblotting

Successful protein transfer was confirmed by the transfer of coloured bands present in the protein ladder onto the PVDF membrane. The membrane was transferred to a solution of 5% skimmed milk powder in 50ml TBS-T and incubated at room temperature on a rocker (Bibby Scientific) for at least four hours to block the membrane. Antibodies were prepared in dilutions according to the manufacturer's guidelines. The blocked membrane was placed into a 50ml Falcon tube with 5ml of antibody solution and incubated at room temperature on a roller for 1 hour. Excess antibody was washed off by pouring off excess solution followed by the addition of 10ml TBS-T. The membrane was rocked at room temperature for 5 minutes, and the washes repeated a further two times. The corresponding secondary antibody was then incubated with the membrane following the same procedure, and the same wash process repeated. Antibodies and dilutions used are listed in Table 32.

Primary Antibody	**Manufacturer**	**Dilution**	**Secondary Antibody**	**Conjugate**	**Manufacturer**	**Dilution**
Rabbit Anti-CDKN1B	Santa Cruz Biotechnology	1:100	Goat Anti-Rabbit	Horseradish Peroxidase (HRP)	Cell Signalling Technology	1:2000
Mouse Anti-α-tubulin	Cell Signalling Technology	1:1000	Rabbit Anti-Mouse	HRP	Cell Signalling Technology	1:1000

Table 32. Antibodies used in Western Immunoblotting.

After washing, a signal was generated using Amersham Enhanced Chemiluminescence (ECL) Prime Western Blotting Detection Reagent (GE Healthcare). 500µl each of reagents A and B were mixed and subsequently added to the PVDF membrane. After incubation at room temperature for one minute, excess reagent was removed by touching the edge of the membrane to tissue paper. The membrane was placed into a G:Box (SynGene) for the detection of chemiluminescent signals produced by the reaction of HRP with ECL reagent.

5.2.5.4 Image Densitometry Analysis

Images produced by the G:Box were analysed using the ImageJ software (Version 1.48, http://rsb.info.nih.gov/ij/, National Institutes of Health). The software was used to relate

target band density to those produced by control antibodies, allowing for the normalisation of relative protein abundance between experiments and across gels.

5.2.6 Cell Proliferation Assay (MTS)

CellTiter 96 AQueous One Solution Cell Proliferation Assay (Promega) was used for measuring cell proliferation rates in a 96 well format. This colourimetric assay relies on the cleavage of tetrazolium dye by nicotinamide adenine dinucleotide phosphate (NADPH)-dependent cellular oxireductase enzymes, reflecting the number of viable cells present. Cells were seeded into each well of a 96-well plate, at a concentration of 5 x 10^5 cells/ml in a volume of 90µl. The top and bottom rows of each test plate contained 90µl of media only to allow for blanking and control wells to be established. For drug dosing experiments, 10µl of media containing serially-diluted drug or vehicle only control was added to this suspension. After 72 hours of incubation, 10µl of AQueous One reagent was added to each well and the plate incubated at 37°C, 5% CO_2 for two hours. Absorbance was then measured at 490nm using a Fluostar Omega plate reader, allowing relative proliferation rates to be calculated after subtracting blanking values for media only wells.

For the statistical analysis of correlation between transcript or protein and 50% growth inhibition (GI50) values, a Shapiro-Wilk test of normality was applied. As all tests performed found data that were not normally distributed, a non-parametric test of correlation was applied. The Spearman's Rank Correlation Coefficient was used to calculate rho and an associated p-value. Shapiro-Wilk test of normality and Spearman's Rank Correlation Coefficient tests were performed in STATA 11 (StataCorp).

5.3 Results

5.3.1 Confirmation of *CDKN1B* Overexpression in NKM-1 Cell Lines

A population of NKM-1 cells were transduced with SIN-SIEW-*CDKN1B* lentiviral particles and the expression levels of *CDKN1B* were measured using quantitative PCR. The qPCR showed an almost 30-fold increase in the expression levels of *CDKN1B* in the transduced population when compared to a non-transduced control population, as demonstrated in Figure 41.

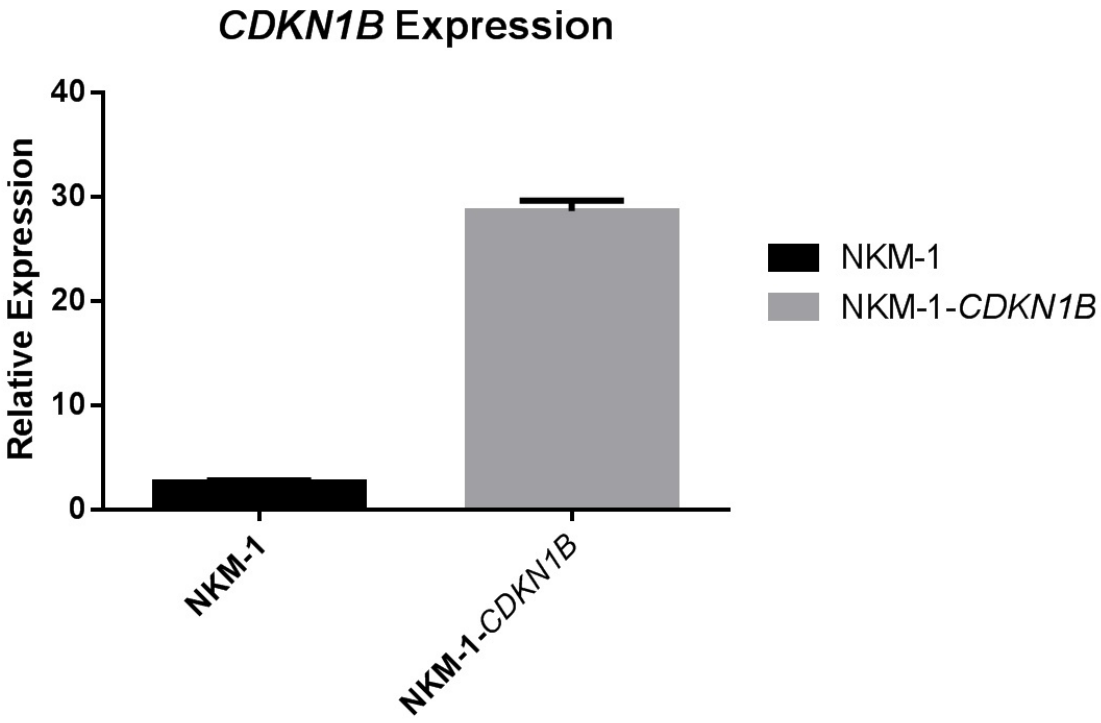


Figure 41. Mean transcript levels of CDKN1B measured by quantitative PCR in a control population of NKM-1 cells compared to NKM-1 cells transduced with SIN-SIEW-CDKN1B, 72 hours after transduction. Error bars show standard deviation of triplicate qPCR results.

To confirm that *CDKN1B* transcripts produced from integrated lentiviral vectors were translated to protein, Western immunoblotting was performed to assess the levels of CDKN1B protein present. In the transduced population a clear increase in CDKN1B levels was observed, as shown in Figure 42. Image densitometry analysis of the blotting image estimated a 20 fold increase in the amount of CDKN1B protein present.

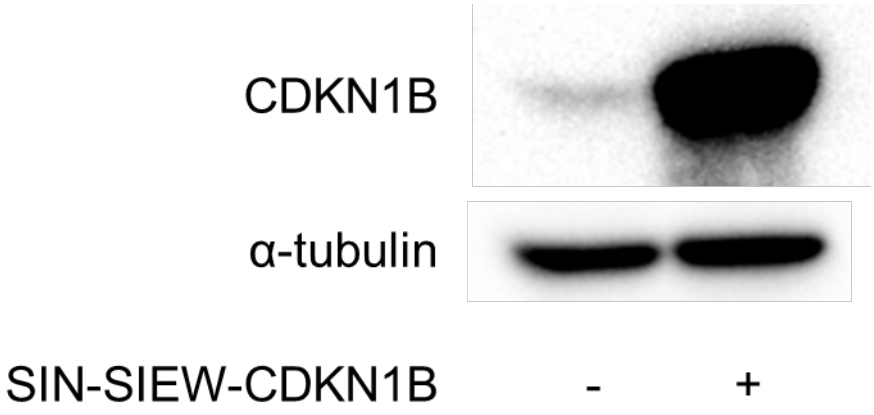


Figure 42. Western immunoblot of CDKN1B *in a control population of NKM-1 cells compared to NKM-1 cells transduced with SIN-SIEW-CDKN1B, 72 hours after transduction.*

To validate the negative relationship between overexpression of *CDKN1B* and cell proliferation, an MTS ((3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)) assay was used to assess the rate of proliferation in cell populations transduced with SIN-SIEW-*CDKN1B* as compared to control cells. As shown in Figure 43, transduction with SIN-SIEW-*CDKN1B* reduced the number of viable cells present 72 hours after transduction to less than half of those present in the non-transduced population. A t-test confirmed a significant difference between the two groups (p=0.012).

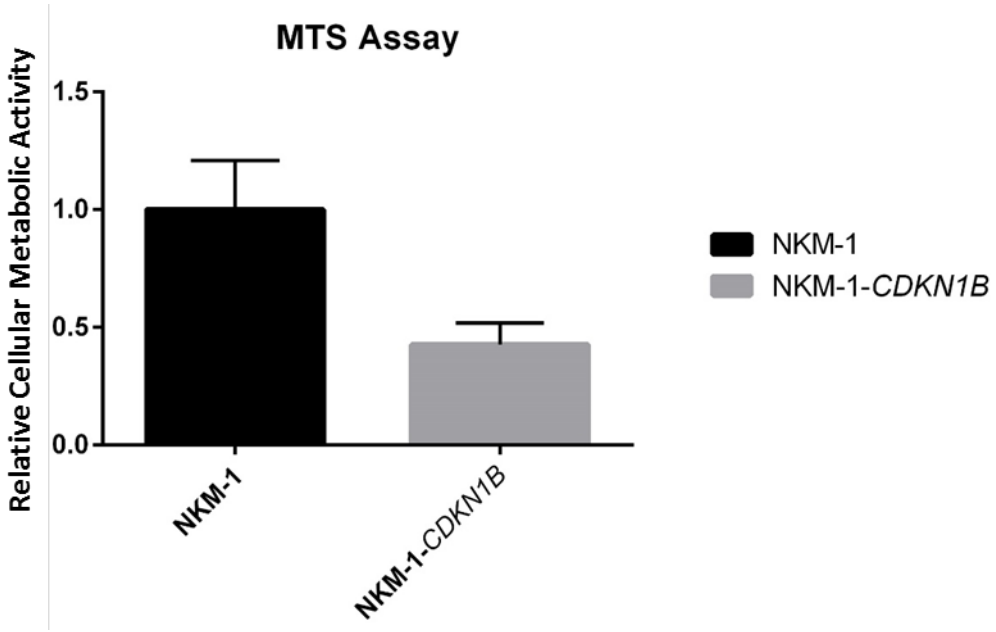


Figure 43. NKM-1 cell proliferation in non-transduced cells compared to cells transduced with SIN-SIEW-CDKN1B. Proliferation assay was performed using AQueousOne reagent in a colourimetric MTS assay, 72 hours after transduction and seeding in equal numbers into a 96 well microplate. Error bars represent standard deviation of triplicate repeats.

Cell cycle analyses using PI were performed on the same populations, the results of which are shown in Figure 44. The data showed a small increase in the number of cells in G_2/M phase in the NKM-1-*CDKN1B* population, with slight reductions in the proportion of cells in G_0/G_1 and S phase. However, as shown in Figure 44A, there was an accumulation of cells at the end of S phase that appeared to be unable to transition into the G_2/M phase.

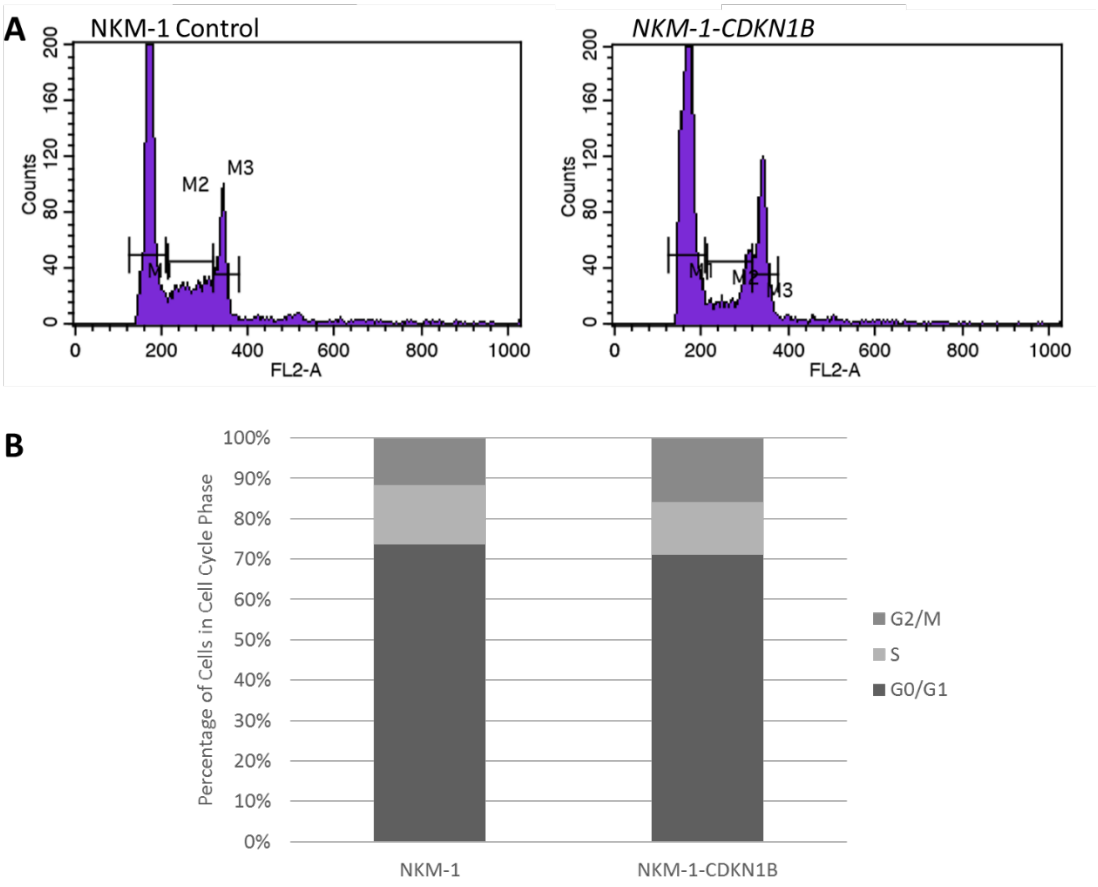


Figure 44. Cell cycle analyses performed on NKM-1 control and SIN-SIEW-CDKN1B transduced populations using PI staining to determine cellular DNA content 72 hours after transduction. Samples were gated after flow cytometry analysis to assess the numbers of cells in each cell cycle phase, though the use of three markers (M1-3) (A). Obtained values were subsequently plotted onto stacked column graphs (B).

5.3.2 *CDKN1B* Expression in AML Cell Lines

Expression levels of *CDKN1B* were measured in a panel of 13 AML cell lines using qPCR, as shown in Figure 45. Expression levels were calculated by relating observed *CDKN1B* levels to the expression levels of the housekeeping gene *GAPDH* (as selected by qPCR optimisation described in 5.2.2.3). A wide range of *CDKN1B* expression levels were observed across cell lines, with a particularly low level of expression in the GDM-1 cell line.

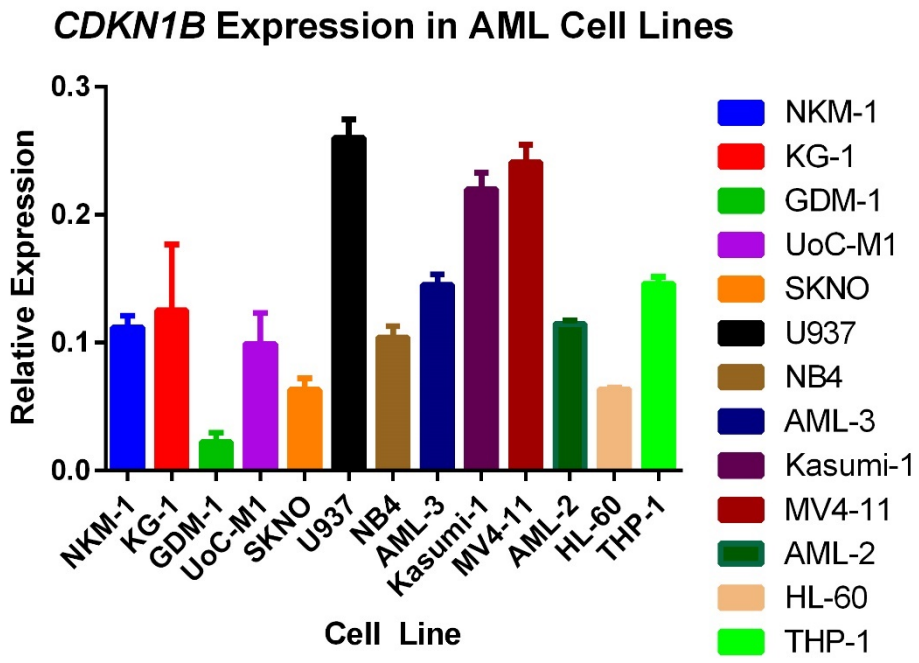


Figure 45. Expression of CDKN1B measured in a panel of AML cell lines using quantitative Real Time PCR. Data shown are representative of CDKN1B expression normalised to a housekeeping gene (GAPDH). Error bars represent standard deviation calculated from triplicate qPCR repeats.

For statistical analysis, cell lines were grouped into 12p deleted (NKM-1, KG-1 and GDM-1) and non 12p deleted lines (all other lines in the panel) as shown in Figure 46. An unpaired t-test showed no significant difference between the two groups (p=0.219).

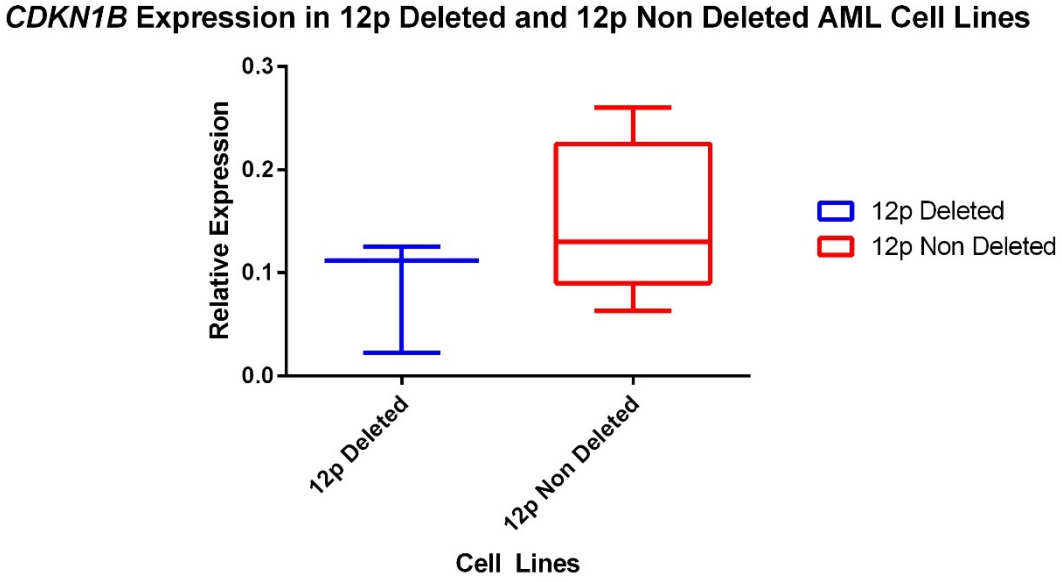


Figure 46. Expression of CDKN1B measured in a panel of AML cell lines split into two groups depending on 12p deletion status. CDKN1B expression levels were measured using quantitative Real Time PCR and normalised using the GAPDH housekeeping gene. The central bar shows the median, whilst the upper and lower bars of the inner box

represent the interquartile range (Non 12p Deleted only). The furthest bars from the central point of each box show the range.

Protein was extracted from the same panel of AML cell lines and assessed using Western blotting. Band density was analysed and CDKN1B protein levels were normalised to α-tubulin controls. The 12p deleted samples (NKM-1, KG-1 and GDM-1) were run on all gels and used to normalise band density between experiments, as shown in Figure 47 (original gel pictures are displayed in Supplementary Figure 57).

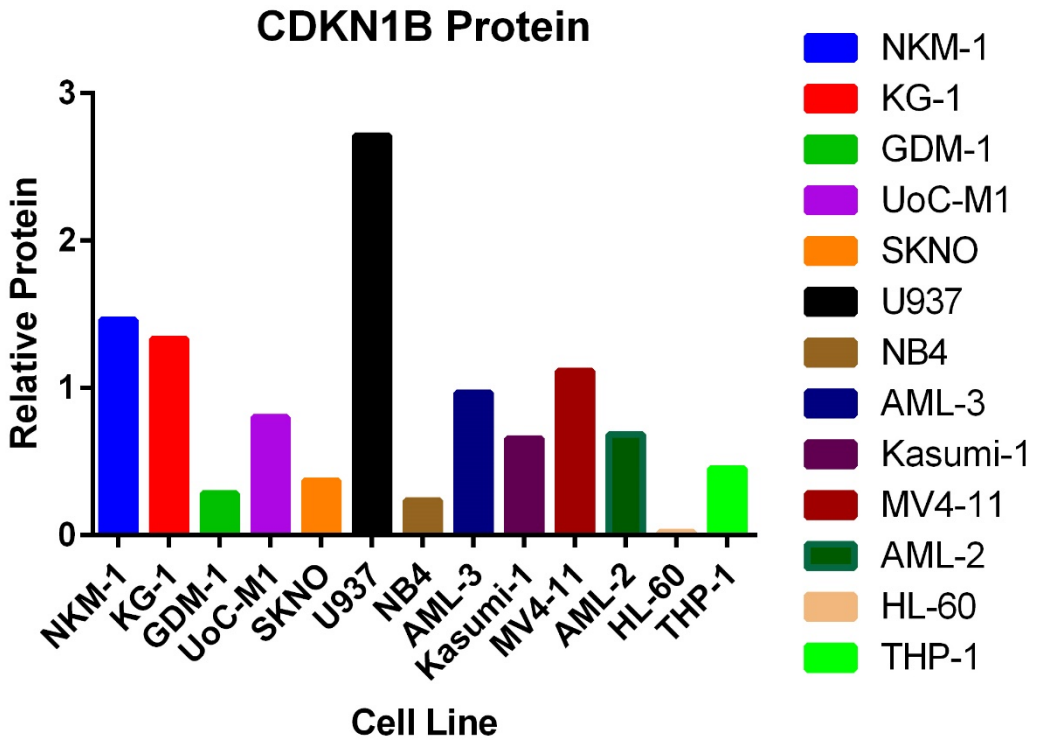


Figure 47. CDKN1B protein expression levels measured in a panel of AML cell lines. Bands imaged from Western blotting were quantified using band densitometry and normalised to α-tubulin controls. NKM-1, KG-1 and GDM-1 samples were run on all gels and used to normalise samples run on different gels.

For statistical analysis, cell lines were grouped into 12p deleted (NKM-1, KG-1 and GDM-1) and non 12p deleted lines (all other lines in the panel) as shown in Figure 48. An unpaired t-test between the two groups found no significant difference in the levels of CDKN1B protein detected (p=0.649).

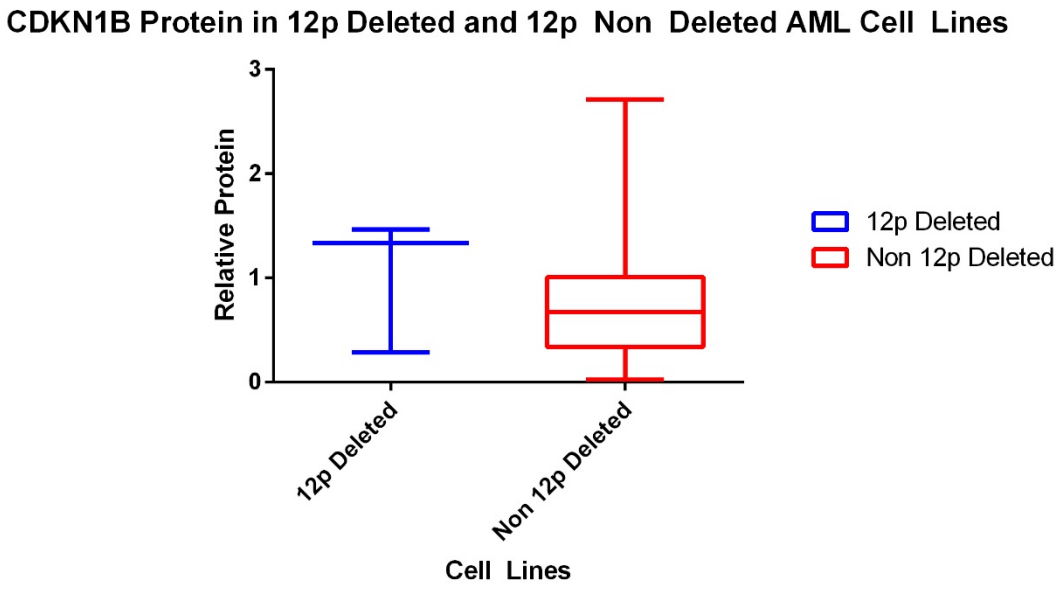


Figure 48. Expression of CDKN1B protein measured in a panel of AML cell lines split into two groups depending on 12p deletion status. The central bar shows the median, whilst the upper and lower bars of the inner box represent the interquartile range (Non 12p Deleted only). The furthest bars from the central point of each box show the range.

5.3.3 *CDKN1B* Expression in Patient Samples

The expression levels of *CDKN1B* were assessed in 76 AML patient cDNA samples (kindly donated by Prof. Claudia Haferlach, Munich Leukemia Laboratory). Of these, 45 had deletions of 12p, whilst 31 samples were provided for use as controls. As shown in Figure 49, the levels of *CDKN1B* expression between the two groups were very similar in both median and range, although the median was slightly lower for the 12p deleted group. An unpaired t-test confirmed that there was no significant difference between the two groups (p=0.606).

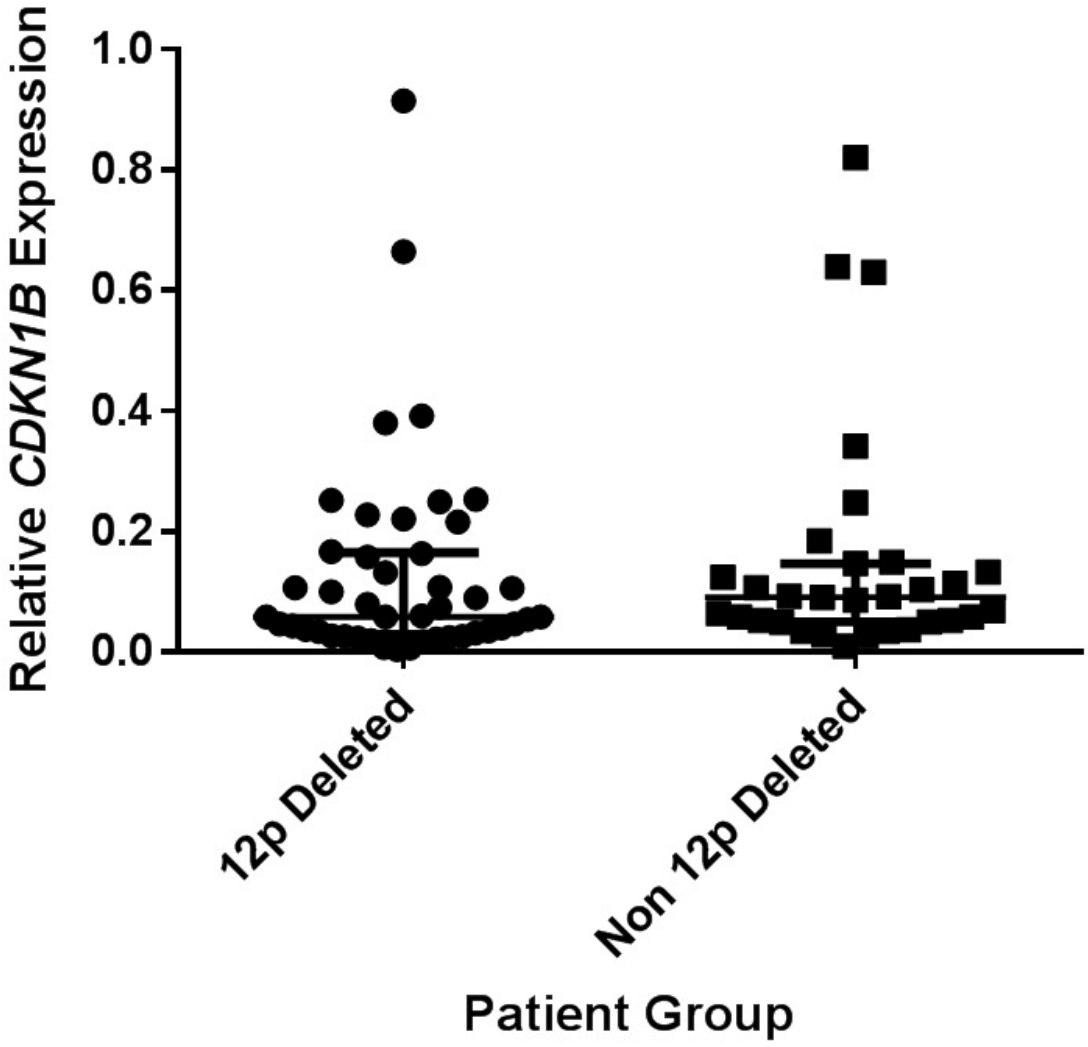


Figure 49. CDKN1B expression levels measured by quantitative Real Time PCR in a panel of 76 AML patients. In the 12p deleted group were 45 patients, whilst there were 31 patient samples analysed in the control non 12p deleted group. CDKN1B expression levels were calculated relative to the GAPDH housekeeping gene. The central bar represents the median of the group, whilst the bars above and below the mean represent the interquartile range.

5.3.4 *CDKN1B* Expression in Normal Myeloid Cells

To gain a greater understanding of the role of *CDKN1B* in the samples analysed, expression microarray data was analysed to ascertain the *CDKN1B* expression levels relative to *GAPDH* in normal haematopoietic cells. Using the R2: Genomics Analysis and Visualization Platform (Academic Medical Center, the Netherlands), an expression microarray performed on flow sorted haematopoietic cell populations was analysed (Dataset GSE19599, Fioretos *et al.*, 2012, Affymetrix Human Genome U133 Plus 2.0 Array). As shown in Figure 50, *CDKN1B* expression relative to *GAPDH* ranged from 0.05 to 0.10 across a range of myeloid and myeloid progenitor cells. Although analysed on a different platform, these are within a similar range to the majority of cells lines and patient samples using qPCR in this study.

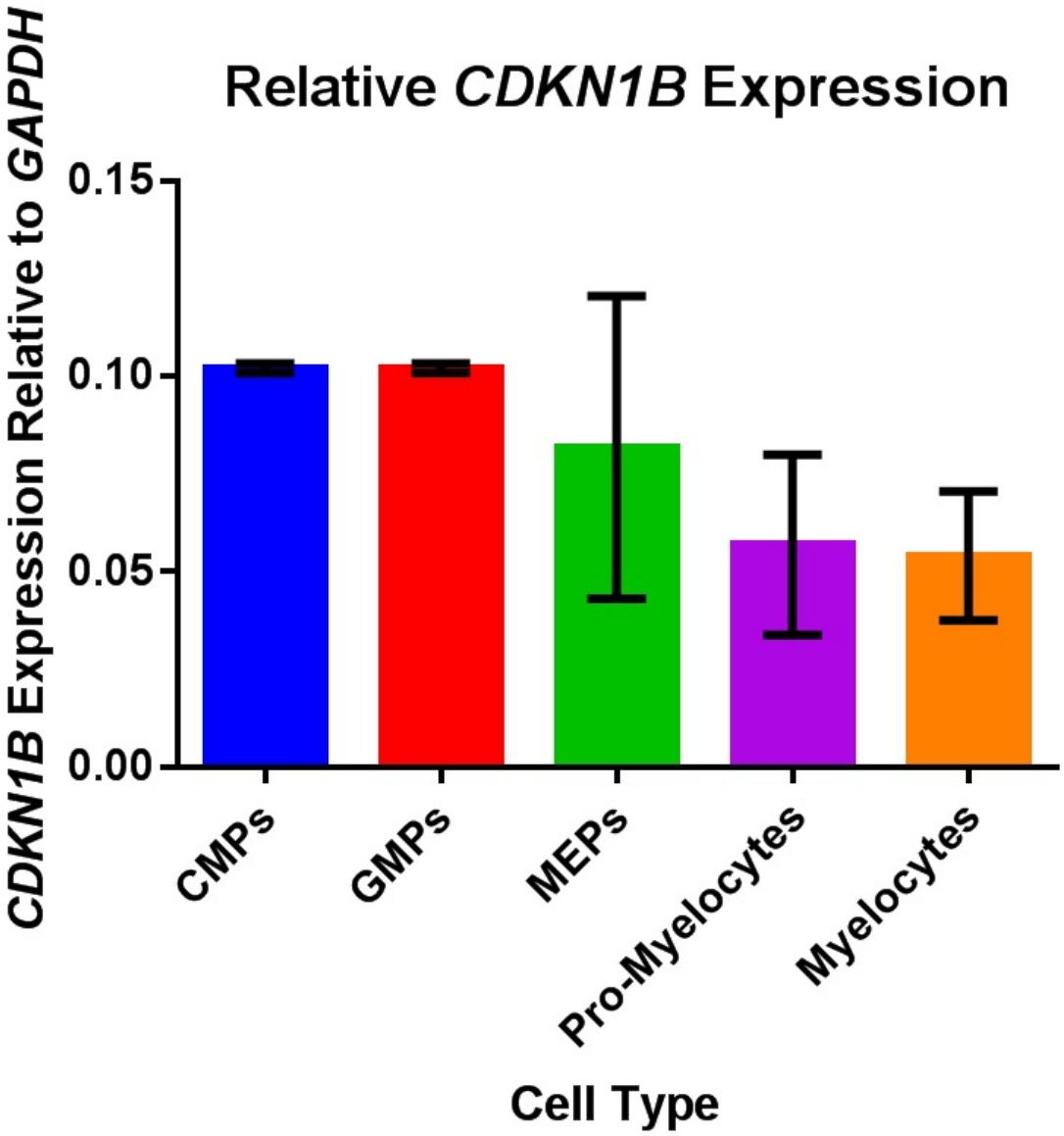


Figure 50. CDKN1B expression relative to GAPDH in normal haematopoietic cells, as measured by expression microarray on the Affymetrix Human Genome U133 Plus 2.0 Array platform. Array analysis was performed on flow sorted haematopoietic cells by Fioretos et al. and publicly released to the Gene Expression Omnibus (Accession number GSE19599). Expression data was then accessed using the R2: Genomics Analysis and Visualization Platform (Academic Medical Center, the Netherlands).

5.3.5 Targeting *CDKN1B* with shRNAs in U937 Cells

To analyse the role of *CDKN1B* in cell cycle regulation, the cell line with the highest level of *CDKN1B* expression (Figure 45 and Figure 47) was selected (U937). Cells were transduced with a pool of GIPZ lentiviral vectors driving the expression of shRNAs targeted to *CDKN1B* to reduce its expression. Puromycin selection was used to isolate a population of transduced cells and successful selection confirmed through the analysis of GFP expression in these cells (Figure 51A). Cell cycle analysis using PI was performed on these populations, showing a reduction in

the number of cells in G_0/G_1 and S phases in the U937 cells transduced with the *CDKN1B* shRNAs.

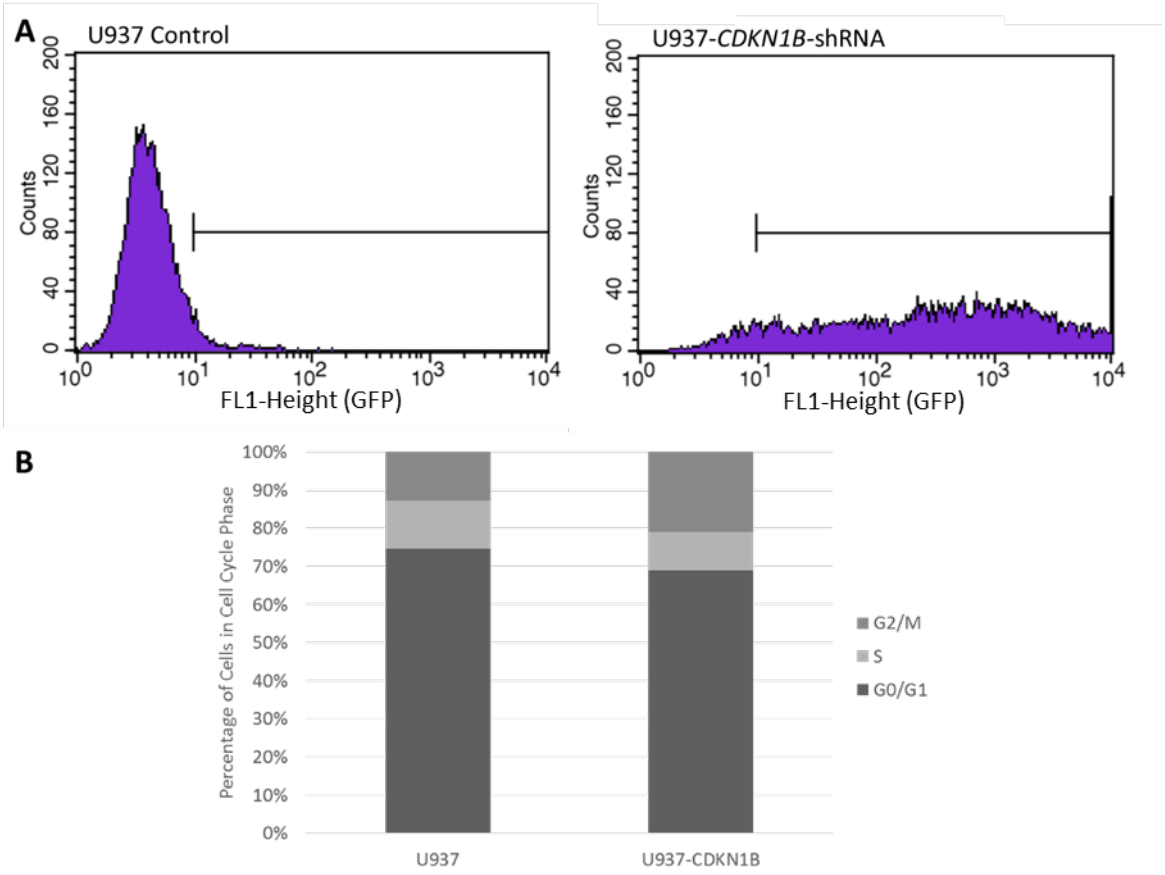


Figure 51. U937 cells transduced with a pool of GIPZ lentiviral vectors to drive the expression of shRNAs targeted to CDKN1B. Puromycin selection was used to isolate a population of successfully transduced cells which was confirmed by flow cytometry measuring GFP expression levels as shown in A. Cell cycle analysis using PI staining was performed using flow cytometry. The x-axis represents GFP expression levels whilst the y-axis shows the number of counts for each section of the histogram. B. The proportions of cells in different cell cycle phases were then calculated using propidium iodide staining of fixed cells.

A cell viability assay was performed on equal numbers of U937 control and transduced cells seeded into individual wells of a 96 well microplate, after 72 hours of incubation. Whilst there was a slight increase in the number of viable transduced cells as shown in Figure 52 (1.08 vs. 1), this was not significant (t-test p=0.101).

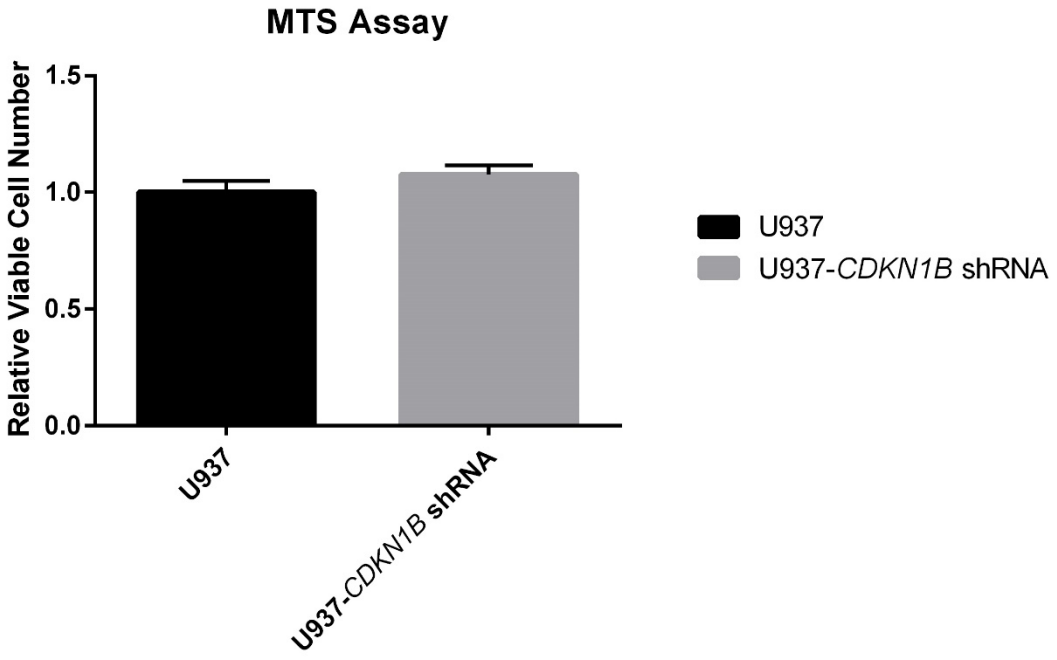


Figure 52. Cell viability assay (MTS) performed on U937 control and CDKN1B GIPZ shRNA transduced cells cultured for 72 hours after seeding in equal numbers in triplicate wells of a 96 well microplate.

5.3.6 Exposure of AML Cell Lines to Flavopiridol

As discussed in 5.1.1, CDKN1B has been demonstrated to inhibit cell cycle progression through the binding of CDK2 and CDK4 complexes and their subsequent inhibition. If CDK2 and CDK4 complexes were more readily activated in the absence of CDKN1B inhibition, in those AML cell lines with a reduced level of *CDKN1B* expression, it could be hypothesised that treatment with a CDK2/CDK4 inhibitor may be an effective replacement for CDKN1B function. The panel of AML cell lines described in previous experiments was treated with a range of flavopiridol concentrations for 72 hours to assess whether this link could be established through the calculation of 50% Growth Inhibition (GI50) values. An MTS assay was performed to assess cell viability and GI50 values were calculated and compared to *CDKN1B* transcript and CDKN1B protein levels, as shown in Figure 53.

GI50 values were tested for normality using the Shapiro-Wilk test. The data were not found to be normally distributed, so a non-parametric test (Spearman's rank correlation coefficient) was used to calculate correlation between GI50 and *CDKN1B* transcript or CDKN1B protein values. There was a moderate correlation between flavopiridol GI50 and *CDKN1B* transcript levels (rho=0.405), however this was not significant (p=0.320). There was a weak correlation between flavopiridol GI50 and CDKN1B protein levels (rho=0.262), but this was also not significant (p=0.531).

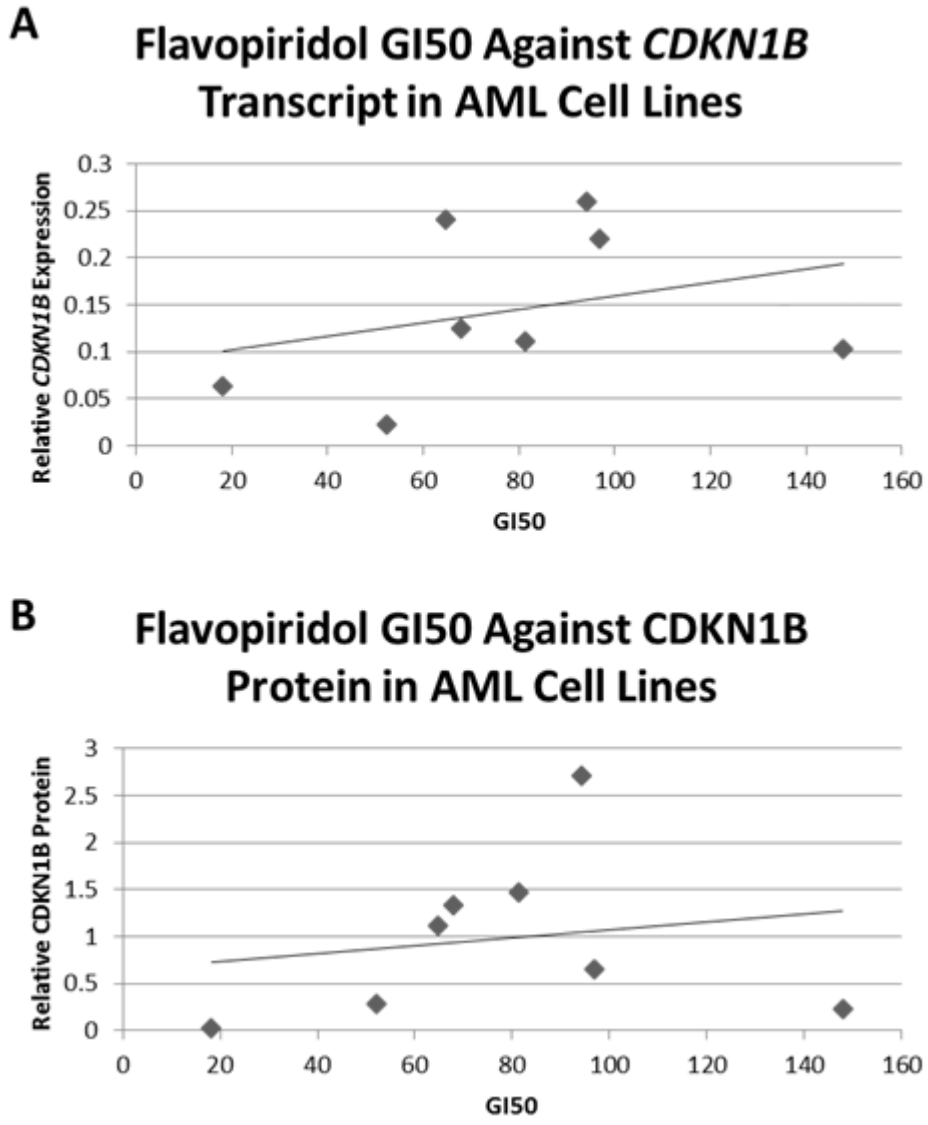


Figure 53. A panel of AML cell lines were seeded into individual wells of a 96 well microplate and treated with flavopiridol for 72 hours to calculate GI50 values using an MTS assay. GI50 values were then correlated with CDKN1B transcript (A) and CDKN1B protein (B) levels, as calculated in Figure 45 and Figure 47.

5.3.7 Validation of *MPZL3* and *UBE4A* Knockdown in UoC-M1 Cells

To confirm successful knockdown of the 11q screen hits *MPZL3* and *UBE4A*, UoC-M1 cells were transduced on an individual basis with each lentiviral vector and puromycin used to select positive cell populations. Transcript levels of *MPZL3* and *UBE4A* were then assessed using qPCR as shown in Figure 54. For all shRNAs targeted to both *MPZL3* and *UBE4A*, transcript knockdown was found to be greater than 50% knockdown compared to NSC (Non Silencing Control) samples.

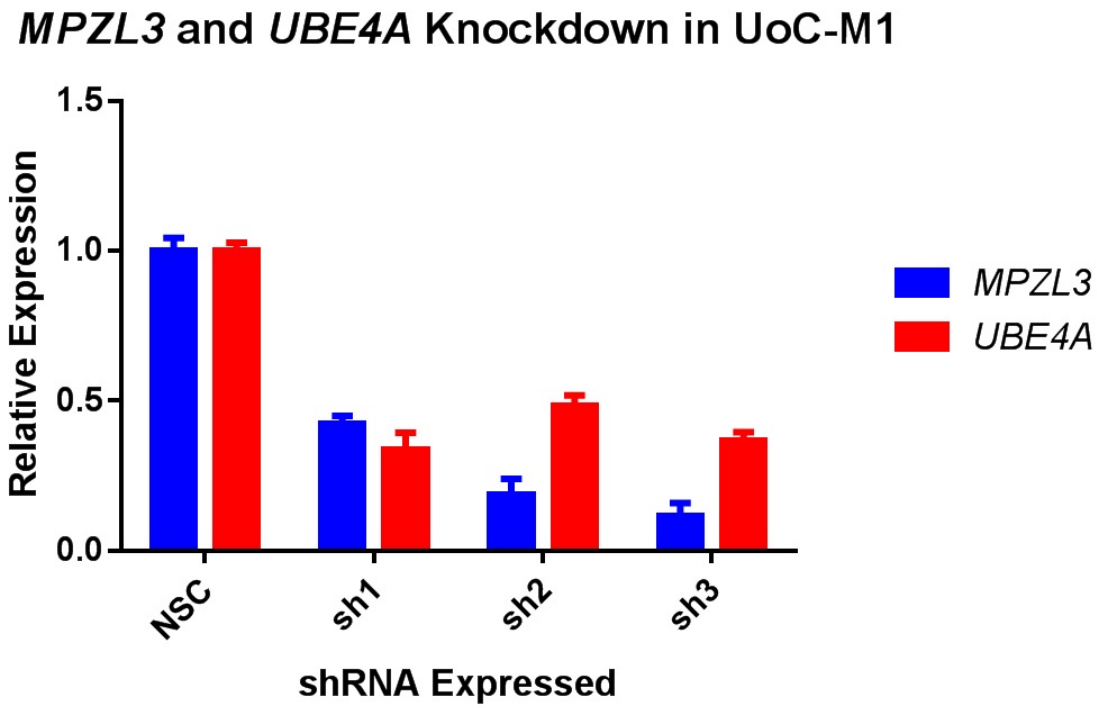


Figure 54. MPZL3 and UBE4A expression in UoC-M1 cells transduced with pGIPZ non silencing control (NSC), or one of three shRNAs targeted to either MPZL3 or UBE4A. Puromycin selection was then applied to obtain a population of GFP positive cells indicating positive shRNA selection. mRNA was extracted and qPCR performed using primers specific to MPZL3 or UBE4A to calculate transcript abundance relative to GAPDH. Data shown is adjusted to relative transcript abundance compared to NSC samples.

5.4 Discussion

5.4.1 Candidate Tumour Suppressor Gene *CDKN1B*

To validate the effects of *CDKN1B* overexpression observed in Chapter 4, NKM-1 cells were transduced with the SIN-SIEW-*CDKN1B* lentiviral construct alone. Analysis of *CDKN1B* mRNA and protein levels were assessed in the transduced population and found to be greatly elevated compared to the non-transduced NKM-1 cells. A second assay measuring the number of viable cells 72 hours after seeding further confirmed the reduced rate of proliferation of these cells.

As *CDKN1B* has a primary role in cell cycle checkpoint regulation, the cell cycle distribution of transduced cells was assessed and compared to control cells. The cell cycle distribution was similar between the two populations, but with an increased proportion of transduced cells in the G_2/M phase. This was somewhat unexpected, as the primary and most well studied role of CDKN1B is in the inhibition of CDK2 and CDK4/6 complexes, which are regulators of the G_1/S phase. Therefore, one would expect that overexpression of *CDKN1B* would lead to a block at this stage, increasing the proportion of cells in G_1.

The transcript and protein levels of *CDKN1B* were then measured in a panel of AML cell lines. The three cell lines carrying deletions of 12p (NKM-1, GDM-1 and KG-1) and UoC-M1 (used in the 11q project) have been used in previous experiments. An additional nine AML cell lines were acquired and *CDKN1B* expression assessed. A wide range of expression levels were observed across the cell lines, with no significant difference between cell lines carrying deletions of 12p and those without deletions of 12p. This could suggest that if *CDKN1B* is an important haploinsufficient tumour suppressor gene in AML, it is dysregulated through other mechanisms in different cell lines. Supporting this hypothesis is the research surrounding other mechanisms that exist to regulate *CDKN1B*, through its transcription, translation and proteolysis, as described in 5.1.1. Additionally, this would explain the lack of correlation between transcript and protein levels observed in some cell line samples. Targeting these mechanisms for treatment may be beneficial in returning *CDKN1B* expression levels to normal.

The cell line with the lowest expression levels of *CDKN1B* transcript was the GDM-1 cell line used in the repeat negative selection assay performed in Chapter 4. This very low level of *CDKN1B* transcription, in conjunction with the lower level of expression from integrated lentiviral vectors in this cell line, could explain the smaller changes in relative construct copy number observed during leukaemic expansion as its expression from the integrated SIN-SIEW lentiviral vector may have been better tolerated.

The expression levels of *CDKN1B* were also measured in AML patient samples, both with and without deletion of 12p. Again, a range of *CDKN1B* expression levels were observed, and there was no significant difference between patients carrying a deletion of 12p and those without deletions of 12p (p=0.606). This further supports the hypothesis that *CDKN1B* may be down regulated by different mechanisms in samples that do not have a deletion of 12p. Additionally, the presence of cell lines and patients with higher levels of *CDKN1B* expression raises the possibility that *CDKN1B* may have an oncogenic role in some AML cases, which in turn could be explained by the alternative role of cytoplasmic CDKN1B that was described above. For this reason, it would be highly informative to quantify RNA and protein levels in nuclear and cytoplasmic fractions separated from AML cells.

To assess whether the expression levels of *CDKN1B* found in the cell line and patient samples used in this study was similar to those found in normal haematopoietic cells, the R2: Genomics Analysis and Visualization Platform was utilised to extract data from a study performed by Fioretos *et al.* (2012) on flow sorted normal haematopoietic cells. *CDKN1B* expression was analysed relative to *GAPDH* expression, as had been performed in this study using qPCR. In the Fioretos *et al.* (2012) study, *CDKN1B* expression levels appeared to be between 5% and 10% of *GAPDH* expression. This is in line with the qPCR data obtained for both cell lines and patient samples, although there were exceptions from both datasets. Three cell lines (U937, Kasumi-1 and MV4-11) demonstrated expression of *CDKN1B* at above 0.2 relative to *GAPDH*, whilst two 12p deleted and three non-12p deleted patient samples had *CDKN1B* expression levels greater than 0.6 relative to *GAPDH*. This would suggest that there are some cases in which *CDKN1B* expression is elevated above normal myeloid cell expression levels, although due to the use of two different platforms (microarray and qPCR) it is not possible to draw any solid conclusions, but further investigation of this in the future may be warranted.

To investigate the role of *CDKN1B* when highly expressed in AML, U937 cells were selected to be targeted with a pool of GIPZ shRNAs to knockdown *CDKN1B*. Puromycin was used to select for a population made up of transduced cells expressing these shRNAs. Cell cycle analysis showed a decreased proportion of cells in G_0/G_1 phase, suggesting that a greater proportion of cells were actively dividing. However, the difference was not significant when assessed using a cell proliferation and viability assay (p=0.101). Due to the lower level of expression found with GIPZ lentiviral constructs throughout this project, it would be beneficial to repeat this experiment with longer culture times to ascertain the true nature of *CDKN1B* in U937 cells.

Finally, the effects of flavopiridol on AML cell line growth were assessed to investigate a possible link between the loss of CDK inhibition through low *CDKN1B* expression and the ability to re-establish it through treatment using a CDK inhibitor. Flavopiridol was selected as it has

been demonstrated to inhibit CDK2 and CDK4, both targets of CDKN1B, as well as having been used in clinical trials for a number of cancer types (primarily CLL) (Senderowicz, 1999; Stephens *et al.*, 2013; Pierceall *et al.*, 2014) and it is currently in phase II trials for the treatment of intermediate and high-risk AML (Zeidner *et al.*, 2014). This study found a moderate correlation between *CDKN1B* transcript levels and GI50 values calculated for sensitivity to flavopiridol, although in a test of dependency these values were found to be independent (p=0.320). It would be desirable to perform further investigation of this link in additional cell lines, as well as combining the data with information on cytoplasmic and nuclear *CDKN1B* transcript and protein abundance in the cell lines tested.

Although the data presented here are not entirely clear in implicating *CDKN1B* directly in cell cycle regulation in 12p deleted AML cases, it appears that *CDKN1B* does have a role in leukaemogenesis. These data suggest that *CDKN1B* can be dysregulated by alternative routes in non 12p deleted AML, although the existence of highly expressing cell lines and patient samples suggest that *CDKN1B* may also have an oncogenic role. The dual roles of *CDKN1B* could explain why neither homozygous deletions of 12p or mutations affecting the remaining allele in cases of heterozygous deletion have been reported. It is clear that more detailed investigations into the role of *CDKN1B* in AML are required to elucidate the contribution of *CDKN1B* in leukaemogenesis.

5.4.2 Candidate Oncogenes *MPZL3* and *UBE4A*

To confirm transcript knockdown using the GIPZ lentiviral shRNA-containing vectors, UoC-M1 cells were transduced with individual GIPZ vectors targeted to either *MPZL3* or *UBE4A*. Puromycin was then used to select for a pure population of transduced cells and transcript levels were measured relative to *GAPDH*. A knockdown of over 50% was achieved with all shRNAs.

For shRNAs targeted to *MPZL3*, the largest knockdown was achieved with shRNA 3, which showed the second greatest effect in the negative selection assay. Somewhat unexpectedly, shRNA 1, which showed the greatest effect in the negative selection assay, demonstrated the lowest level knockdown of the three shRNAs targeted to *MPZL3*. Of the shRNAs targeted to *UBE4A*, the greatest knockdown was achieved with shRNA 1. This is concordant with the data obtained in the negative selection assay, where shRNA 1 was selected against most strongly in both the MinP and MaxP samples. Similar levels of knockdown were achieved with shRNAs 2 and 3. Whilst expression of shRNA 3 was selected against in the screening assay, expression of shRNA 2 was not.

It is somewhat conflicting that whilst knockdown of the target genes was achieved with all shRNAs, this was not always associated with negative selection in the assays performed in Chapter 4. Without further functional investigation, it is unclear for the reasons behind this, but possible reasons could be speculated. Due to the large number of lentiviral vectors present in the pools, it may be that relatively few starting cells were transduced with each GIPZ vector. This could lead to a situation where the majority of cells containing a particular vector express the contained shRNA at a low level, due to integration into a genomic site of low expression. Without the pressures of puromycin selection, these cells are not selected against and may therefore remain in the cell pool without any adverse effects due to shRNA expression. Alternatively, it may be the case that particular shRNAs are only capable of reducing the level of specific transcript splice variants, the loss of which may not lead to a reduction in functional protein levels. To assess this, it would be ideal to quantify the amounts of *MPZL3* and *UBE4A* protein present in cells selected using puromycin.

Chapter 6 General Discussion

6.1 Summary of Aims

6.1.1 Defining the Extent of Chromosome 12p Deletions and Chromosome 11q Amplifications

As there was no clear consensus on the extent of 12p and 11q copy number alterations in AML, the main objective of this body of work was to establish minimal regions of copy number alteration for both of these chromosomal regions.

6.1.1.1 Deletions of Chromosome 12p

Deletions and translocations of the short arm of chromosome 12 have previously been reported in a range of haematological malignancies, including AML, ALL, MDS and non-Hodgkin lymphoma. In AML, deletions have been reported in between 1% and 5% of AML patients (Grimwade *et al.*, 2010; Harrison *et al.*, 2010), and have been frequently associated with other abnormalities within a complex karyotype, including the unfavourable -7/-7q and -5/-5q (Wall *et al.*, 2012). Further studies have suggested that up to 10% of normal karyotype AML patients may have deletions of 12p that are not detectable by cytogenetic analysis (Andreasson *et al.*, 1997).

Although there are few studies that have investigated the prognostic impact of 12p deletions, two studies have demonstrated that in paediatric AML patients, deletions of 12p have a poor prognosis, with an event free survival of 35% compared to 59% in patients without deletions of 12p (Harrison *et al.*, 2010, von Neuhoff *et al.,* 2010). There are no studies that have looked specifically at the prognostic relevance of these deletions in adult AML as they occur most frequently within a complex karyotype, however analysis of annotated microarray data (Parkin *et al.*, 2010) revealed a poor prognosis in one study, albeit with relatively few patients (Section 0, Figure 4).

To define a miminally deleted region for chromosome 12p, eight studies that had made use of copy number microarrays, including one focussing 12p, were selected for reanalysis. A total of 866 AML patient samples were assessed, and 5.2% of samples from studies that had not specifically been selected for deletions of 12p were identified as having a deletion encompassing this region. By overlaying the deletions in this region, a minimally deleted region was defined containing 10 genes: *BCL2L14*, *LRP6*, *MANSC1*, *LOH12CR1*, *DUSP16*, *CREBL2*, *GPR19*, *CDKN1B*, *APOLD1* and *DDX47*.

To determine the incidence of submicroscopic 12p deletions in normal karyotype AML patients, a MLPA probeset was custom designed to interrogate the copy number status of a

selection of genes from within and around the MDR. Nine probes were designed and optimised before testing on cell line and patient material with known deletions of 12p. The probeset was then used to screen 75 normal karyotype AML patients, five of which were found to have deletions of 12p (6.7%). The deletions in four of these five patients were confirmed using FISH. The remaining patient did not have the required material available for this verification. All five patients identified using MLPA were found to have deletions larger than the established MDR, and were therefore not informative in reducing the number of genes.

At this stage, three AML cell lines with deletions of 12p encompassing the MDR, as well as relatively few other genetic aberrations that may have affected results obtained in functional anaylsis, were identified using the Sanger COSMIC Cell Line database – GDM-1, KG-1 and NKM-1.

The putative tumour suppressor gene *ETV6* was absent from the 12p MDR from the cell lines chosen. However, *ETV6* was included in the functional analysis for a number of reasons. The MDR was determined using copy number SNP arrays, thus it was unknown whether other chromosomal rearrangements may have been present that could alter the expression of *ETV6* without affecting copy number. This scenario is not unlikely as the 12p13 chromosomal region is known to be unstable, with multiple complex rearrangements previously reported (Sato *et al.*, 2001). Translocations of *ETV6* in AML have been shown to form fusion genes that do not result in a functional transcript, and these have been suggested to contribute to AML through lower expression of *ETV6* resulting from haploinsufficiency (Silva *et al.*, 2008). Although relatively rare in AML, some mutations of *ETV6* have been reported to act in a dominant negative manner, disrupting *ETV6* function in the absence of a second hit (van Doorn-Khosrovani *et al.*, 2005).

6.1.1.2 Amplifications of Chromosome 11q

Amplifications of chromosome 11q are relatively rare, occurring in 1 to 2% of AML patients. Gain of material from this region is among the most common aberrations observed in 15-40% of complex karyotype AML (Schoch *et al.*, 2002; Rücker *et al.*, 2006). It is likely that if there is a gene or genes with an oncogenic dosage effect present in this region, it would be present within common regions of gain and amplification observed across multiple patients. Due to the strong association with complex karyotype AML as well as an average older age of disease manifestation, amplifications of 11q carry a very poor prognosis.

There have been two recent studies that have focussed on establishing a minimally amplified region of chromosome 11q in AML. In 2006, Rücker *et al.* identified three separate common regions of gain on chromosome 11q through the analysis of complex karyotype AML patients

using microarray technology. A more recent study by Zatkova *et al.* (2009) found nine recurrent regions of amplification on 11q, three of which were observed in all cases studied. These regions had considerable overlap with those found in the 2006 study and contained the suspected proto-oncogenes *MLL*, *ETS1* and *FLI1*.

To gain a greater understanding of 11q abnormalities in AML patients, gains and amplifications were mapped in the datasets also used in the 12p study. Across the entire cohort (n=866), 25 patients were identified with gains of 11q, and five patients with amplifications of 11q. Through alignment of these data with the previously mentioned studies, three commonly amplified regions containing a total of 32 genes were identified and selected for functional analysis.

The UoC-M1 cell line was selected for use in functional analysis work. Although this cell line has only four copies of the genes of interest, it had previously been demonstrated that expression of genes from the 11q arm were significantly elevated when compared to control samples (Allen *et al.*, 1998).

6.1.2 Identification of Genes Contributing to Disease Development Through the use of a Functional Negative Selection Assay

The candidate gene lists established through analysis of array and MLPA data were taken forward for functional assessment using a negative screening assay. In this assay, cDNA or shRNA constructs targeting candidate tumour suppressor or oncogenes respectively were introduced into AML cell lines containing the genetic aberrations of interest using lentiviral transduction. Transduced cells were then expanded *in vitro* and *in vivo* through intrafemoral injection of cells into immunocompromised mice (12p project only). Detrimental effects on growth due to expression of cDNAs or shRNAs from integrated lentiviral vectors were measured through the comparison of relative construct proportion before and after leukaemic cell expansion using high throughput NGS. Optimised methodology for production of lentiviral particles as well as transfection and transduction conditions, were developed as part of this study.

The engraftment of cell lines when intrafemorally injected into immunocompromised mice was assessed. Whole body bioluminescent imaging of mice injected intrafemorally with cell lines transduced with lentivirus expressing luciferase resulted in signals consistent with the development of solid tumours around the site of injection. However upon sacrifice of the mice, 50% showed an increased spleen weight consistent with leukaemic infiltration. Subsequent analysis of preserved skulls (performed by Dr. Christina Halsey, University of Glasgow) showed

further leukaemic infiltration of the calvaria and leptomeningies in some samples, suggesting more extensive engraftment than first thought.

6.1.2.1 Chromosome 12p Candidate Gene List

An initial pilot study was performed using the NKM-1 cell line to assess the candidate genes from the 12p MDR. The pilot *in vitro* study showed strong and rapid selection against the *CDKN1B* expression construct, as well as weaker selection against constructs driving the expression of *CREBL2* and *ETV6*. The pilot study *in vivo* results were less clear, with highly variable construct representation between samples, both between mice as well as between organs within an individual mouse. In some samples, a single expression construct was highly overrepresented. Whilst no solid conclusions could be drawn from these data, the expression clone for *ETV6* was never overrepresented, the expression clone for *LOH12CR1* was only overrepresented once, and the expression clone for *CDKN1B* was overrepresented twice. The remaining expression clones were overrepresented in 3 or more of the mouse samples analysed, suggesting that expression of these genes did not apply significant negative selection pressures.

Following on from the pilot study, repeat studies were performed in triplicate using the NKM-1 cell line. Expression of *CDKN1B* was strongly and rapidly selected against, and the selection against *APOLD1* and *CREBL2* was not recurrent, suggesting that they were spurious results from the pilot study that were eliminated through repetition. A weak selection effect against *ETV6* remained and was observed across all three repeats. The negative selection assay was then performed in an additional cell line carrying deletion of 12p, GDM-1. Whilst the observed effects of negative selection were lesser in this cell line, there was still a highly significant selection against *CDKN1B*. However, the weak selection effect against *ETV6* was not seen in this line and is likely due to a lower level of construct expression, previously demonstrated in FACS experiments assessing GFP levels in transduced cells.

6.1.2.2 Chromosome 11q Candidate Gene List

The negative selection assay performed on the genes present in the amplified 11q region targeted each gene with three shRNAs. Two methods of pooled lentiviral particle preparation were undertaken for this section of work. In the first, each shRNA-containing transfer plasmid was prepared on an individual basis in Plasmid Miniprep size cultures (MinP). In the second method, bacterial stocks were pooled in groups of 10 shRNAs each and extracted in nine separate Maxiprep size cultures (MaxP). Lentiviral particles were then produced in pools of 10 shRNAs each for both methods, and used to transduce UoC-M1 cells. These two methods were carried out to assess whether pooling of lentiviral vectors at the bacterial stage (as in MaxP) might have led to highly uneven representation or even loss of some shRNAs, although this

needed to be balanced with the greatly improved vector yield obtained from MaxiPrep cultures when compared to MiniPrep cultures.

Although there were no cases where all shRNAs targeted to one gene were found to be hits, shRNAs targeted to two genes were repeatedly selected against. The knockdown of *MPZL3* was significantly selected against by two shRNAs in both the MaxP and MinP samples, and the knockdown of *UBE4A* was also selected against by one MaxP and two MinP shRNAs. The relative changes in construct proportion observed across all experiments using GIPZ shRNAs was far lower than the changes observed using the SIN-SIEW lentiviral vectors for the 12p project, suggesting either a lower level of expression or silencing of lentiviral vectors over time.

6.1.3 Investigating the Relevance of Targets Identified by the Screening Assay

The functional follow up assessment of the targets identified by the negative screening assay mainly focussed on the strongest hit identified, which was *CDKN1B*. *CDKN1B* overexpression was confirmed in NKM-1 cells transduced with SIN-SIEW-*CDKN1B* through the use of qPCR, and subsequent translation of *CDKN1B* mRNA into protein was confirmed using Western immunoblotting. Culturing control and transduced cells for 72 hours resulted in a significantly reduced number of viable cells present in populations transduced with SIN-SIEW-*CDKN1B*, when assessed using an MTS assay. Cell cycle analysis of the same cells demonstrated a slight reduction in the proportion of cells in G_0/G_1 and S phase, and an increased number of cells in G_2/M phase. The most well studied role of CDKN1B is in the inhibition of CDK complexes that regulate the G_1/S phase, so this result was a little unexpected as it could be hypothesised that overexpression of *CDKN1B* would block this stage of the cell cycle and increase the proportion of cells in the G_1 phase.

The expression levels of *CDKN1B* were then measured in a panel of AML cell lines. The cell lines demonstrated a range of expression levels, with a particularly low level of *CDKN1B* expression found in the GDM-1 cell line. Grouping cell lines into 12p deleted and non-12p deleted categories did not reveal a significant difference between the two groups. The levels of CDKN1B protein in the same panel of cell lines was also assessed, again with a wide range of expression levels observed. Grouping the cell lines into the two categories mentioned above did not show a significant difference in the protein levels observed. These data therefore demonstrated that deletion of the 12p region does not correlate with a reduction in *CDKN1B* transcript or protein levels relative to other AML cell lines.

Continuing on this theme, the expression levels of *CDKN1B* were measured in a panel of 76 AML patient samples (kindly donated by Prof. Claudia Haferlach, Munich Leukemia Laboratory). Of these, 45 had deletions of 12p, whilst 31 samples were included as controls.

There was no significant difference in expression levels between the two groups when assessed using an unpaired t-test (p=0.606), although the median was slightly lower in the 12p deleted group when compared to the non-12p deleted group. This indicated that in AML patient samples, deletion of 12p did not lead to a significant relative reduction in *CDKN1B* transcript levels compared to non 12p deleted samples. It is possible that *CDKN1B* expression is reduced in most AML cell lines and patient samples that do not have a deletion of 12p through an alternative mechanism, which would result in no observed relative difference between the two groups.

CDKN1B expression was then analysed in a panel of normal myeloid and myeloid progenitor cells, using the R2: Genomics Analysis and Visualization Platform and the GSE19599 dataset provided by Fioretos *et al.* (2012). Although this involved comparing data from two different platforms (expression microarray and qPCR), it was found that the levels of *CDKN1B* expression measured in this project were comparable to the *CDKN1B* expression levels found in the Fioretos *et al.* (2012) dataset, with a few exceptions in both cell lines and patient samples, which demonstrated elevated levels of *CDKN1B* expression relative to *GAPDH*. Although further investigation is required, if this was true, it would suggest that deletion of 12p does not result in a significantly altered expression of *CDKN1B.*

Following this, the cell line with the highest level of *CDKN1B* expression, U937, was selected for further functional analysis. U937 cells transduced with GIPZ lentiviral targeted to *CDKN1B* showed a reduction in the number of cells in G_0/G_1 and S phases, but demonstrated no significant reduction in growth 72 hours after transduction.

Finally, a panel of AML cell lines were exposed to the CDK2 and CDK4 inhibitor flavopiridol, with the hypothesis that inhibition of CDKN1B through reduced expression would lead to an increased susceptibility to treatment with a CDK2/CDK4 inhibitor. GI50 values were calculated and analysis with Spearman's rank correlation coefficient testing showed a moderate correlation between flavopiridol GI50 and *CDKN1B* transcript levels (rho=0.405), although this was not significant (p=0.320). There was a moderate correlation between flavopiridol GI50 and CDKN1B protein levels (rho=0.262), but again this was not significant (p=0.531). The presence of a trend suggests that further work is required to confirm or refute this hypothesis, and this is discussed further in Section 6.3.3.

The follow up work performed on the candidate genes identified by the 11q negative selection assay focussed on knockdown validation of shRNAs targeted to *MPZL3* and *UBE4A*. UoC-M1 cells were transduced with individual shRNAs targeted to each candidate gene and positively transduced cells were selected using puromycin. Transcript knockdown of over 50% was

confirmed for each lentiviral vector, however the levels of knockdown did not fully correlate with the level of negative selection observed in Chapter 4. The greatest levels of *MPZL3* knockdown was achieved with shRNA 3, followed by shRNA 2. However, in the negative selection assay, the greatest selection effect was seen with shRNA 1, followed by shRNA 3, whilst shRNA 2 showed no significant negative selection effect. For *UBE4A* knockdown, the greatest knockdown was achieved with shRNA 1, which also showed the largest effect in the negative selection assay. Similar levels of knockdown were seen with both shRNA 2 and shRNA 3, but in the negative selection assay shRNA 3 was selected against, whilst shRNA 2 was not. The discordance between levels of knockdown and the results of the negative selection assay suggest that further validation of the negative selection assay results would be required to ensure that the results obtained were not due to random fluctuation of relative construct proportion.

6.2 Study Strengths and Limitations

6.2.1 Chapter 3

One of the major strengths of this chapter stemmed from the large number of researchers who have published their microarray data on databases such as Gene Expression Omnibus and ArrayExpress, without which this body of work would not have been possible. It allowed 866 AML patient data files to be reanalysed with a focus on 12p deletions and 11q amplifications, many of which had matched remission or relapse data, allowing germline and relapse specific copy number alterations to be assessed. The high numbers involved made this study larger than any other published studies that have focussed specifically on deletions of 12p (Sato *et al.*, 1995; Wlodarska *et al.*, 1996; Andreasson *et al.*, 1997; Andreasson *et al.*, 1998; Baens *et al.*, 1999; Silva *et al.*, 2008; Haferlach *et al.*, 2009) providing greater power to identify a minimally deleted region and candidate gene list for deletions of 12p. Due to the low frequency of 11q amplifications in the data sets, this project found fewer patients with amplifications of 11q than other published studies (Rücker *et al.*, 2006; Zatkova *et al.*, 2009); although it was still possible to combine data from all three sources define a minimal region of gain and amplification on 11q. One limitation of this section of work is that only Affymetrix microarray data was used, although many other platforms could have been utilised, as well as other techniques, which will be discussed in 6.3.1 below. However, it could be argued as a strength, as the use of the same analysis protocol ensures that all samples are analysed in a standardised fashion, avoiding the introduction of technical bias.

Another strength of the study was the custom designed MLPA probeset to target genes on 12p. It provided rapid screening of large numbers of patients for deletions of 12p which was not previously available. Unfortunately, due to the small number of patient samples available, it was not possible to screen sufficient samples to build on conclusions drawn from the analysis of the microarray data.

The Sanger database of cancer cell lines was invaluable to identify suitable AML cell lines for the negative selection assay. Three cell lines with deletions of 12p and relatively few other genetic aberrations were found (GDM-1, Ben-Bassat *et al.,* 1982; KG-1, Furley *et al.*, 1986; NKM-1, Kataoka *et al.*, 1990). Due to the rarity of 11q amplifications in AML, cell lines were scarce. A literature review identified three candidate cell lines, however it was only possible to obtain one with four copies of the 11q region of interest. Although the cell line demonstrated elevated expression of genes present on 11q, a model with a higher level of amplification would have been more representative.

6.2.2 Chapter 4

The main strength of the negative selection described in Chapter 4 was the novel nature of the assay. The use of an assay of this type for the functional interrogation of candidate genes from within regions of deletion or amplification in cancer has not previously been reported. It enabled a functional approach to identify those copy number aberrations driving leukaemic progression in AML with deletions of 12p and amplifications of 11q.

One of the main drawbacks to this body of work was that only AML cell lines were assessed, while the use of primary patient material would have been preferred. As reviewed by Gillet *et al.* (2013), whilst some cancer cell lines have been valuable as models in a preclinical setting, others have failed. They suggested a number of limitations of using cell lines, such as a low relevance to actual disease states. This could produce misleading results if a gene assessed using the negative selection assay was important in leukaemic development, but its knockdown or overexpression was tolerated in a cell line setting.

Another limitation concerns the negative selection assay performed for the 11q candidate genes. The two main hits identified were only identified as hits for three or four of the six shRNAs targeted to the genes of interest (three separate shRNAs total, with two methods of lentiviral vector replication). As demonstrated in the 12p part of the project, data replication is important to exclude spurious results and the data would have been stronger with more consistent findings. The source of this weakness is unclear, but it could be speculated that low levels of shRNA expression from integrated GIPZ vectors, possibly in combination with silencing of the CMV promoter, may have led to weak selection pressures when the shRNAs were expressed. The use of the UoC-M1 cell line, which only contains four copies of the region of interest, may demonstrate a lower expression of candidate genes when compared to AML samples with higher levels of 11q expression. This may mean that the cell line is less susceptible to shRNA targeting of these genes. Careful validation using alternative models such as patient material or mouse models would be required to confirm these findings.

The *in vivo* body of this project was useful in furthering understanding of AML development *in vivo*, but failed to fully replicate the *in vitro* negative selection assay results, likely due to the pattern of leukaemic engraftment in the mice, as it appeared that a small population of cells had engrafted and expanded at each site, leading to overrepresentation of a small number of lentiviral vectors. There was little consistency in which lentiviral vectors were over or underrepresented, with major variation between both mice and between sites of engraftment isolated from within mice. Subsequent discussion and research has implied that tail vein

injection may be a more suitable route of leukaemic cell delivery, as demonstrated by Wang *et al.* (2012) and Miller *et al.* (2013).

6.2.3 Chapter 5

A strength of the 12p portion of this chapter was the successful confirmation of *CDKN1B* overexpression and investigation into its effects. However, this functional assessment ran into problems, where not all experiments corroborated the evidence presented in Chapter 4, suggesting that *CDKN1B* acts as a tumour suppressor through its mechanisms of action upon the cell cycle. This could be a highly significant finding if it was found that *CDKN1B* may have some oncogenic role as it does in CML (discussed in Section 5.1), which may explain why biallelic *CDKN1B* deletions are not reported in AML, if one copy is required to fulfil these oncogenic functions. It appeared that *CDKN1B* expression in both AML cell lines and patient samples was not always linked to 12p copy number status, an unexpected finding, but one that may be of interest in understanding the role of *CDKN1B* in AML. This chapter demonstrated that much work remains to be done before clear conclusions can be drawn (discussed further below in Section 6.3.3), but it has provided a solid foundation for future studies.

Due to time constraints, it was not possible to perform more detailed investigations into the 11q candidate oncogenes identified in the negative selection assay, which is a limitation of this portion of the project. However, successful knockdown of target genes was established, providing a useful starting point for future functional investigations.

6.3 Future Work

6.3.1 Further Investigation of Copy Number Alterations in AML

The work performed in Chapter 3 that investigated the extent of chromosome 12p deletions and chromosome 11q amplifications in AML could be further expanded. Since the investigation was performed, there are new platforms available for investigating copy number alterations (for example the Affymetrix CytoScan HD) and more data sets are now available on the GEO database. Expanding the database with additional data could identify more AML patients with the chromosomal aberrations of interest, providing further information about the extent of minimally altered regions. This would be particularly beneficial to the 11q section of this project as fewer individuals were found with this chromosomal aberration. As this project only made use of Affymetrix SNP arrays, further expansion could include a wider range of SNP array manufacturers, for example the Illumina range of Human Genotyping Arrays, to provide additional data. The inclusion of NGS data would also provide further information about the copy number status of genes within the regions of interest.

As considerable effort was put into the development of a custom 12p MLPA probeset, it would be ideal to screen additional patient samples. Through screening additional karyotypically normal samples, more patients could be identified with small deletions of 12p, potentially refining the minimally deleted region and providing additional information on the occurrence of 12p deletions in these patients. Performing FISH on these samples may also identify additional mechanisms of dysregulation of the genes within the MDR, such as translocations of 12p as previously shown (Harrison *et al.,* 2010). Should appropriate material be available, it would be of interest to study the clonality of these deletions in leukaemic populations and assess the proportion of cells in each leukaemia that carry the deletion of 12p. This may inform the relative importance of the aberration in leukaemic development.

Although publicly available sequencing data were screened in an attempt to identify inactivating mutations in *CDKN1B*, no such mutations had been reported in AML. A targeted sequencing approach, using PCR enrichment, could be undertaken to screen for *CDKN1B* mutations in a large cohort of AML patients, both with and without deletions of 12p. Screening for mutations affecting the other genes in the MDR, with a particular interest in *ETV6*, which showed a weak effect in the negative selection assay, would provide significant support to their role as tumour suppressor genes in AML.

6.3.2 Expansion of Negative Selection Assay

In vivo work undertaken in this project did not successfully replicate the *in vitro* negative selection assay within the mouse model. As discussed in Section 4.5.1, other researchers have

successfully used intravenous tail vein injection of mice following whole body irradiation to engraft a wide range of AML material, including difficult to engraft patient samples (Wang *et al.*, 2012; Miller *et al.*, 2013). These systems have shown considerable engraftment in both the bone marrow and organs of injected mice, and successful shRNA screens have been performed which have identified hits using similar methods. Repeating the *in vivo* negative selection assay with tail vein injection may yield more informative data.

This work could then be further expanded by performing the negative selection assay in AML material derived from patients with deletions of 12p. As cell lines have often been cultured for many years, during which time they have often gained additional chromosomal aberrations, they are not viewed as the gold standard for performing this type of study (Gillet *et al.*, 2013). Patient's leukaemic cells would then be engrafted and subsequently expanded in mice. This xenograft material could then be transduced and the negative selection assay performed as before. This model of AML would be more representative of human, providing a more accurate representation of the role of each gene studied in the selection assay. For the 11q project, an additional benefit would be the use of a model system from a patient with true amplification of 11q, rather than the UoC-M1 cell line which only has four copies of the 11q region of interest.

Part of the original plan of the project was to perform a positive selection assay in cells derived from INK4a/ARF mutant mice, however due to time constraints this work could not be completed. Foetal liver cells were collected from these mice, which provide a source of HSCs that can be lentivirally transduced and engrafted into NSG mice. In the positive selection assay, putative tumour suppressor genes would be knocked down, and putative proto-oncogenes would be overexpressed. INK4a/ARF null foetal liver cells lead to the development of leukaemias when transplanted into NSG mice, thus the positive selection assay would identify which genes were able to potentiate this process as their corresponding shRNAs or cDNAs would have an elevated relative copy number after expansion compared to their starting copy number.

6.3.3 Further Functional Analysis of Candidate Genes

As discussed in Section 5.4.1, it would be important to gain a greater understanding of the expression levels of *CDKN1B* in normal haematopoietic cells compared to the AML cell lines and patient samples analysed in this project. Although it was possible to use an online resource to extract data from an expression microarray, it is not ideal to compare expression data between two different platforms. For this reason, future work might encompass extracting haematopoietic cells from normal blood and bone marrow samples and flow sorting

them into a range of haematopoietic cell compartments. The expression levels of *CDKN1B* could be directly compared to the expression levels of *CDKN1B* found in the AML samples. This comparison would shed further light on how expression levels of *CDKN1B* change in leukaemogenesis, in both 12p deleted and 12p non-deleted cases as well as in AML cell lines.

As gene editing techniques such as the CRISPR/Cas system have become more popular, the techniques have been further refined and are now useful tools for a wider range of researchers. These techniques could be used to delete specific genes from target cells, which could be useful to specifically delete *CDKN1B* to assess the effects of its loss in haematopoietic stem cells and its role in leukaemogenesis. Additionally, knockout mouse models of *CDKN1B* could be beneficial to assess the role of *CDKN1B* in haematopoiesis and leukaemogenesis. A conditional knockout of *CDKN1B* in the haematopoietic compartment would be ideal, and the range of haematopoietic cells produced by the model could be studied to identify any bias towards the production of a specific cell type. The model could also be exposed to mutagens and the incidence of leukaemic development could be studied in relation to mouse models that are not missing a copy of *CDKN1B*.

It would be of interest to co-express both *CDKN1B* and *ETV6* in the same cell to assess whether the negative selection effect seen in the assay was increased by the expression of both genes together. If so, it would suggest cooperation between these genes in deleted 12p cases is able to drive leukaemogenesis more than the loss of a single one of them. It would also explain why the majority of patients with deletions of 12p have loss of both genes.

The work performed exposing AML cell lines to flavopiridol and relating this to *CDKN1B* expression levels showed a trend towards a correlation between the IC50 values and *CDKN1B* expression. It would be of interest to assess more cell lines to confirm or refute the link between these two factors. Although flavopiridol affects CDKs 2, 4 and 6, it also inhibits other CDK family members, which may lead to its effects becoming less pronounced. For this reason, the test could be improved by utilising a more specific CDK inhibitor to investigate this link.

As there was little time to perform any follow up work on the hits identified in the 11q negative selection assay, it would be of interest to assess the potential of the candidate genes as drivers of leukaemogenesis. Further insight is required into why specific shRNAs were not selected against in the functional assay, despite demonstrating knockdown in a puromycin selected population transduced with individual shRNA containing lentiviral vectors. The first step would be to validate the 11q negative selection assay using further repeats in a smaller pool of shRNAs. Next, the level of protein knockdown achieved using the GIPZ shRNAs should be assessed, as transcript knockdown may not lead to reduction in the levels of protein

present. As there is little known about the function and role of the identified candidate genes in driving oncogenesis, further work should focus on establishing how these genes may contribute to cancer development, before subsequently assessing how these mechanisms might be harnessed for targeted treatments in the future.

6.3.4 Additional Project Applications

Finally, the negative selection assay developed here could easily be applied to other chromosomal copy number aberrations as well as other candidate gene lists. For example, large scale expression studies could be performed to identify a list of genes most commonly dysregulated in AML leukaemogenesis, and this could be used as the basis for a larger scale study to determine which genes are functionally relevant and amenable to targeted treatment. The assay could also be applied to a range of cancer types, either with similar regions of copy number aberrations found across cancer types or by studying individual cancer subtypes with their own specific copy number aberrations.

6.4 Final Conclusions

Utilising a number of methods and novel approaches, a greater understanding of the effects of deletions of the short arm of chromosome 12 and amplification of the long arm of chromosome 11 have in driving leukaemic development has been gained. This body of work has identified candidate genes that will contribute to the understanding of AML and may represent the first steps into the development of novel therapies.

Chapter 7 Supplementary Data

7.1 Supplementary Tables

7.1.1 Table of Primers

Primer	**Primer Sequence**
CS1-SFFV F	ACACTGACGACATGGTTCTACAAGCTGACCTGCAAGAAGAGG
ETV6 R	TACGGTAGCAGAGACTTGGTCTTGGGCTCTCTGGAGGTGTAT
BCL2L14 R	TACGGTAGCAGAGACTTGGTCTGGCCGAGGATTTTGAATTCTATGG
DUSP16 R	TACGGTAGCAGAGACTTGGTCTGTTCCACTTTCCAGCAGAGC
CREBL2 R	TACGGTAGCAGAGACTTGGTCTCACGTTTACCGGGCTTCTTT
GPR19 R	TACGGTAGCAGAGACTTGGTCTGGGCACCAGAAGTGTAGGAA
CDKN1B R	TACGGTAGCAGAGACTTGGTCTCATCCGCTCCAGGCTAGG
APOLD1 R	TACGGTAGCAGAGACTTGGTCTCTTCCGAGTACCCCTGGC
DDX47 R	TACGGTAGCAGAGACTTGGTCTTCCTCCTCTTCCACAATCGG
LOH12CR1 R	TACGGTAGCAGAGACTTGGTCTTGGCTCTATGCTTGGCTG
MANSC1 R	TACGGTAGCAGAGACTTGGTCTAGCAGACAGCCTTAGTGTCA
LRP6 R	TACGGTAGCAGAGACTTGGTCTATAAAGCAACAAAGGGGCCG
pGIPZ Hairpin F	ACACTGACGACATGGTTCTACACAACAGAAGGCTCGAGAAGG
pGIPZ Hairpin R	TACGGTAGCAGAGACTTGGTCTCCTTGAATTCCGAGGCAGTA
GAPDH F (qPCR)	CGACCACTTTGTCAAGCTCA
GAPDH R (qPCR)	GACTGAGTGTGGCAGGGACT
CDKN1B F (qPCR)	TAATTGGGGCTCCGGCTAAC
CDKN1B R (qPCR)	GAAGAATCGTCGGTTGCAGGT

Table 33. Table of primers and associated sequences.

7.1.2 Table of 12p Deletions Detected by SNP Microarray

Series Author	Platform	Sample	Chromosomal Location	Start	End
Radtke	500K	393774	12pter - p12.1	0	22493864
Radtke	500K	393776	12p	0	34855126
Bullinger	500K	472597	12p13.2 - p13.1	11653772	13068046
Bullinger	500K	472618	12p13.2 - p12.3	11388025	19047806
Bullinger	500K	472670	12p13.2 - p13.1	11720446	13014076
Bullinger	500K	472691	12p13.2 - p12.3	11276681	19039754
Bullinger	100K	473086	12p13.2	11354241	12606882
Bullinger	100K	473161	12p13.2 - p12.3	10869580	19279573
Parkin	SNP6.0	631015	12p13.2 - p12.2	10894814	20417967
Parkin	SNP6.0	631021	12p13.31 - p11.21	9178874	31119940
Parkin	SNP6.0	631027	12p13.31 - p11.22	9390271	28481984
Parkin	SNP6.0	631041	12p13.31 - p12.3	6822135	17766576
Parkin	SNP6.0	631075	12p13.31 - p13.2	9997003	14229914
Parkin	SNP6.0	631087	12p13.2 - p11.22	11526805	29335762
Parkin	SNP6.0	631099	12p13.2 - p13.1	11419362	14253967
Parkin	SNP6.0	631115	12p	0	34855126
Parkin	SNP6.0	631145	12p13.31 - p12.1	6543326	22039542
Parkin	SNP6.0	631167	12p13.31 - p12.1	9794659	22372509
Parkin	SNP6.0	631171	12pter - p12.2	0	21195776
Haferlach	SNP6.0	687172	12p13.2 - p13.1	11723293	12934854
Haferlach	SNP6.0	687173	12p13.2 - p12.1	11943146	22510427
Haferlach	SNP6.0	687174	12p13.2 - p13.1	12113639	13763702
Haferlach	SNP6.0	687175	12p13.2 - p13.1	11694841	12990759
Haferlach	SNP6.0	687176	12p13.31 - p11.21	9524658	32829294
Haferlach	SNP6.0	687177	12p13.2 - p13.1	11593665	13008381
Haferlach	SNP6.0	687178	12p13.2 - p12.2	11848875	20626553
Kühn	SNP6.0	802992	12pter - p12.1	0	22450199
Kühn	SNP6.0	802993	12p	0	34855126
Kühn	SNP6.0	808025	12p	0	34855126
Rücker	SNP6.0	850733	12pter - p13.12	0	49225974
Rücker	SNP6.0	850736	12p13.2 - p11.22	11712106	28075420
Rücker	SNP6.0	850746	12p13.31 - p13.1	9402956	14396622
Rücker	SNP6.0	850753	12p13.31 - p12	9956648	44516094
Rücker	SNP6.0	850790	12p13.31 - p11.21	9121580	31094615
Rücker	SNP6.0	850814	12p13.31 - p12.2	8812160	20662827

Rücker	SNP6.0	850817	12p13.31	6930639	34792469
Rücker	SNP6.0	850818	12pter - p12.1	0	26132366
Rücker	SNP6.0	850820	12p13.1	9750671	12990836
Rücker	SNP6.0	850822	12pter - p12.2	0	20662827
Rücker	500K	850723	12p - p12.2	6246808	21079711
Rücker	500K	850724	12pter - p12.1	0	22283997
Rücker	500K	850749	12p13.31 - p11.23	9118434	26827280
Rücker	500K	850759	12p13.31 - p13.1	9394380	13653732
Rücker	500K	850761	12p13.2	11660780	32156516
Rücker	500K	850764	12p13.2	12120652	13495401
Rücker	500K	850785	12p13.31 - p12.1	7241153	21816913
Rücker	500K	850793	12pter - p11.23	0	26982620
Rücker	500K	850804	12pter - p12.1	0	23565147
Rücker	500K	850805	12p13.31	9895254	14822242
Rücker	500K	850815	12pter - p13.11	0	48315759
Rücker	500K	850821	12p13.2 - p11.23	10019903	26431458

Table 34. Details of 12p deletion in all patients studied. Coordinates displayed for 100K and 500K platforms have been converted from hg18 to hg19.

7.1.3 Table of 11q Gains Detected by SNP Microarray – Copy Number 3

Series Author	Platform	Sample GSM	Chromosomal Location	Copy Number	Start	End
Radtke	500K	393614	11q22.3 - qter	3	105826003	134789366
Radtke	500K	393697	11q24.2 - q24.3	3	127575608	129041737
Bullinger	500K	472593	11q23.3	3	116894496	117088420
Bullinger	500K	472616	11q23.3	3	118023622	118462667
Bullinger	500K	472684	11q24.3	3	130695261	130817259
Bullinger	500K	472689	11q23.3	3	118065299	118307939
Bullinger	100K	473055	11q23.3	3	118354166	118805656
Bullinger	100K	473063	11q23.3	3	118384917	119162891
Gupta	SNP6.0	518882	11q23.3	3	115405407	134853273
Parkin	SNP6.0	631107	11p11.12 - qter	3	48201061	134853273
Kühn	SNP6.0	803061	11q22.3 - qter	3	109254589	134853273
Kühn	SNP6.0	803325	11q22.3 - qter	3	105415417	134853273
Kühn	SNP6.0	805581	11q22.1 - qter	3	98168969	134853273
Kühn	SNP6.0	806093	11q21 - qter	3	94257452	134853273
Rücker	SNP6.0	850719	11q13.3 - qter	3	69132499	134853273
Rücker	SNP6.0	850726	11p12 - qter	3	43222827	134853273
Rücker	SNP6.0	850739	Entire 11	3	0	134853273
Rücker	SNP6.0	850748	11q22.1 - qter	3	98688110	134853273
Rücker	SNP6.0	850769	11q24.2 – q25	3	127691866	132547504
Rücker	SNP6.0	850808	Entire 11q	3	50952568	134853273
Rücker	500K	850761	Entire 11q	3	48244485	134789366
Rücker	500K	850718	11q24.1 - qter	3	122375795	134789366

Table 35. Details of 11q gains in all patients studied. Coordinates displayed for 100K and 500K platforms have been converted from hg18 to hg19.

7.1.4 Table of 11q Gains Detected by SNP Microarray – Copy Number 4

Series Author	Platform	Sample GSM	Chromosomal Location	Copy Number	Start	End
Parkin	SNP6.0	630959	11p11.12 - qter	4	51176560	134944800
Parkin	SNP6.0	631113	11p11.12 - qter	4	50090974	134944800
Rücker	SNP6.0	850809	Entire 11	3	0	134449982
				4	66070303	134449982

Table 36. Details of 11q gains with a copy number of 4 in all patients studied.

7.1.5 Table of 11q Amplifications Detected by SNP Microarray

Series Author	Platform	Sample GSM	Chromosomal Location	Copy Number	Start	End
Parkin	SNP6.0	631097	11q24.2 - q25	>4	126953865	130834263
Rücker	SNP6.0	850743	11q12.1	3	57175157	59616210
			11q12.2 - q23.1	3	60763739	110965409
			11q23.1 - q23.3	4	110965443	121521200
			11q23.3 - q25	>4	121521200	132159600
			11q25	>4*	132168501	132570489
			11q25	>4	132570489	134161639
Rücker	500K	850721	11q23.3 - qter	>4	117536581	134944770
Rücker	500K	850724	11q13.4 - q14.1	3	74649956	79536128
			11q23.3	3	117844008	119955430
			11q23.3	>4	118373919	118800198
Rücker	500K	850821	11q13.4 - q14.1	3	74320458	77499823
			11q14.1	4	77499823	78685950
			11q23.3	>4	118121958	119161489
			11q24.2	>4	126846631	131148234

*Table 37. Details of 11q amplifications in all patients studied. Coordinates displayed for 100K and 500K platforms have been converted from hg18 to hg19. * = very high signal strength.*

7.2 Supplementary Figures

7.2.1 PCR Optimisation of NGS Library Preparation

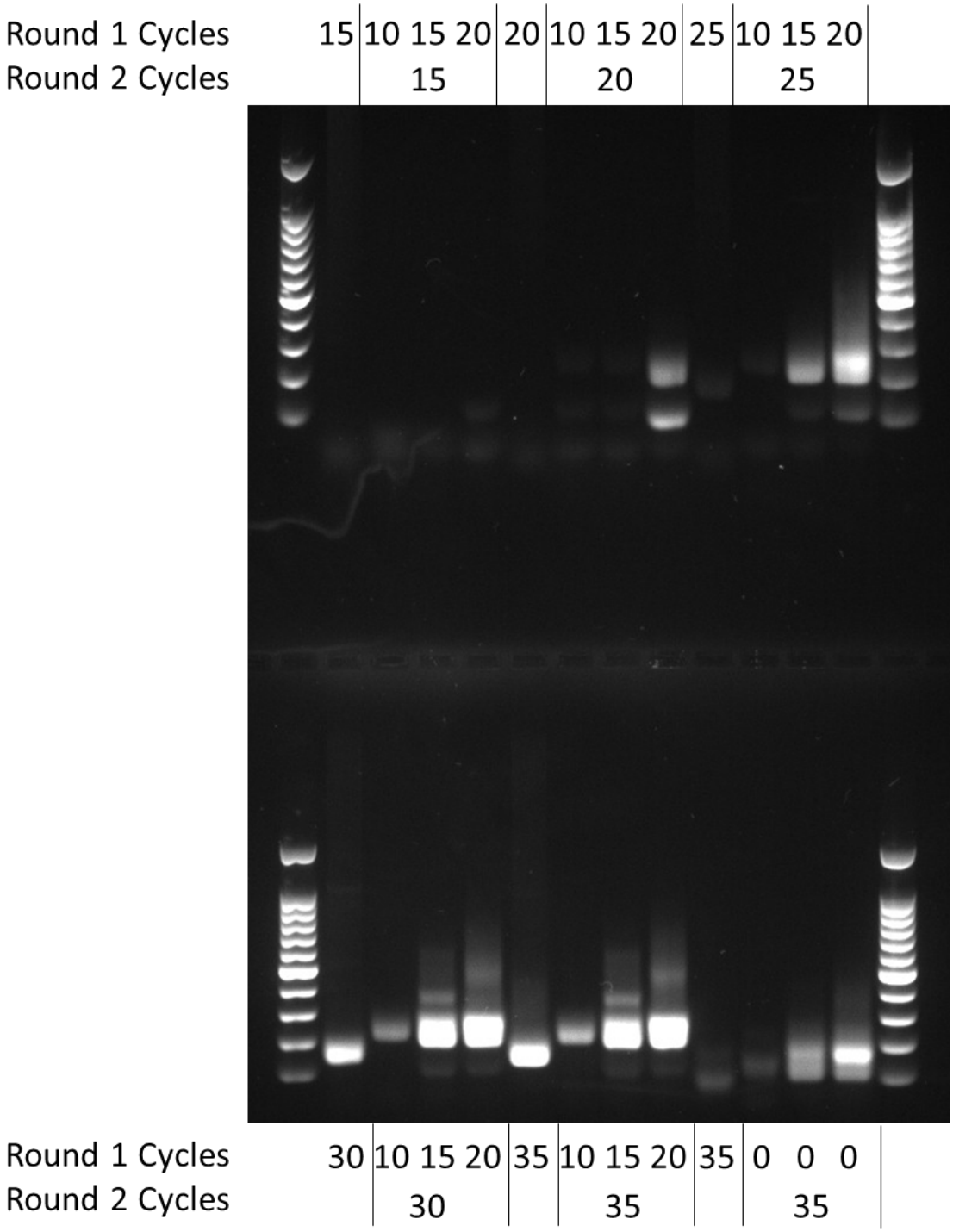


Figure 55. The PCR optimisation process to select numbers of first and second round amplification for sequencing library preparation. Various combinations of round 1 and round 2 cycle numbers were tested to find the optimal combination producing PCR products that are not in the plateau phase of amplification. Round 1 only samples were included in undiluted form for size references as successful round 2 PCR increases product size through the addition of barcode and adapter sequences. In the left and right most lanes, 100bp ladder was loaded for analysis of PCR product sizes.

7.2.2 *FLI1* Negative Selection

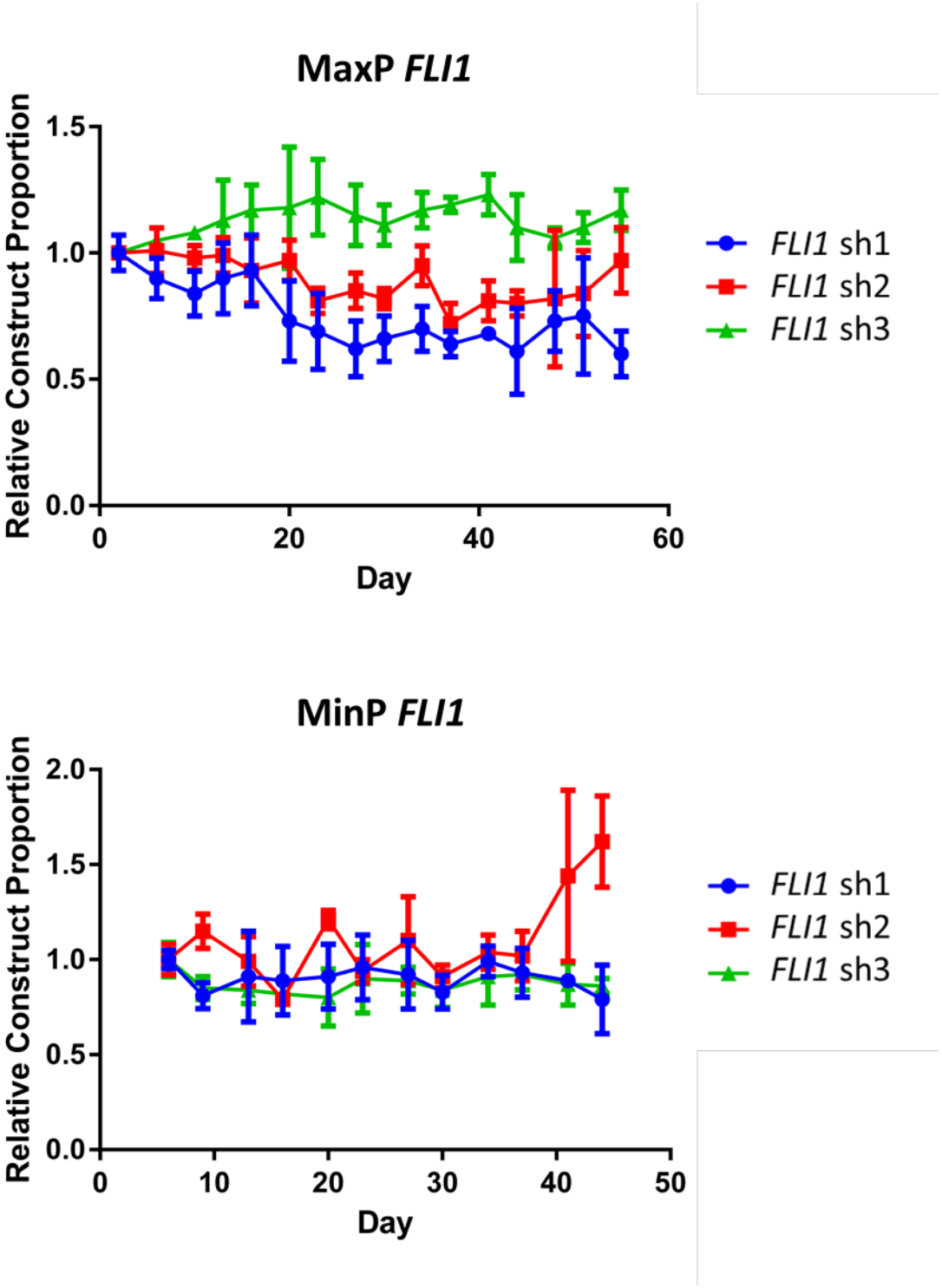


Figure 56. Mean individual results for negative selection using shRNAs targeted to FLI1. sh1 in the MaxP experiment was identified as a hit (p=<0.001).

7.2.3 CDKN1B Expression in AML Cell Lines

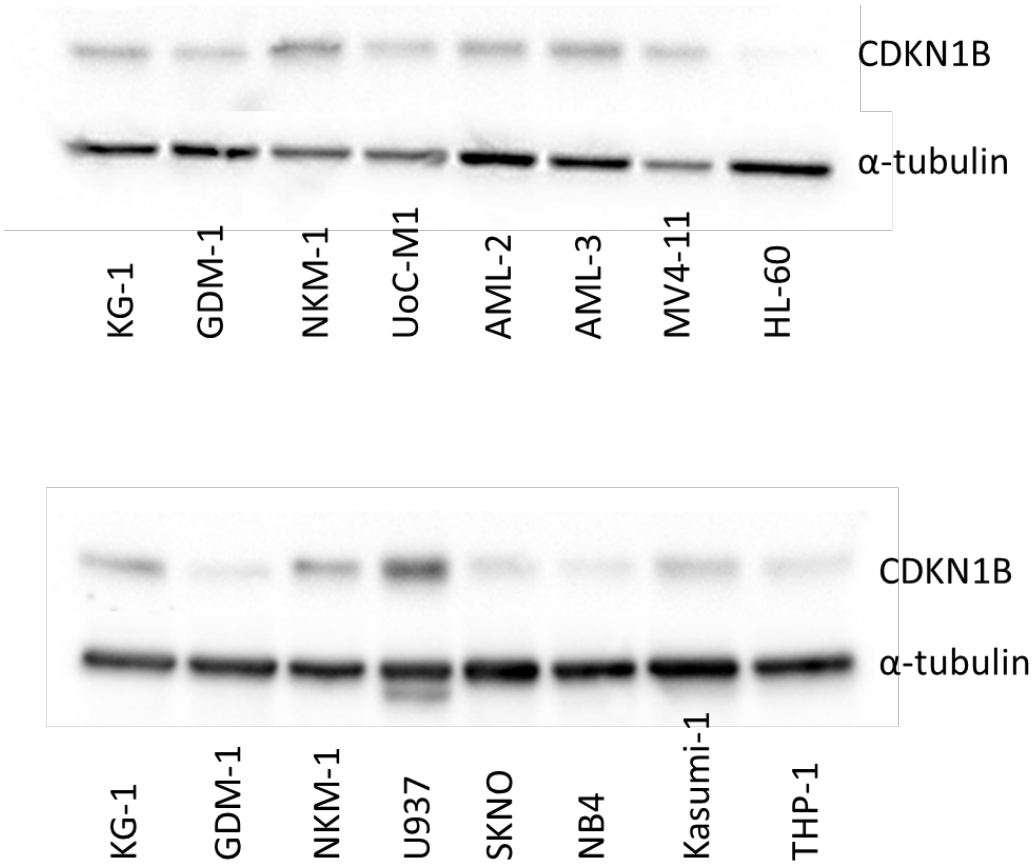


Figure 57. Western Immunoblotting performed using a range of AML cell line protein extracts. Primary antibodies raised to CDKN1B *and α-tubulin were used to bind target proteins, and secondary antibodies conjugated with HRP were used to detect bound protein bands. Samples KG-1, GDM-1 and NKM-1 were used to normalise between blots and band densitometry using ImageJ allowed quantification of protein bands as shown in Figure 47.*

Chapter 8 Bibliography

Acevedo-Arozena, A., Wells, S., Potter, P., Kelly, M., Cox, R.D. and Brown, S.D. (2008) 'ENU mutagenesis, a way forward to understand gene function', *Annu Rev Genomics Hum Genet*, 9, pp. 49-69.

Adams, B.D., Guo, S., Bai, H., Guo, Y., Megyola, C.M., Cheng, J., Heydari, K., Xiao, C., Reddy, E.P. and Lu, J. (2012) 'An in vivo functional screen uncovers miR-150-mediated regulation of hematopoietic injury response', *Cell Rep*, 2(4), pp. 1048-60.

Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D. and Brinster, R.L. (1985) 'The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice', *Nature*, 318(6046), pp. 533-8.

Agarwal, A., Mackenzie, R.J., Besson, A., Jeng, S., Carey, A., LaTocha, D.H., Fleischman, A.G., Duquesnes, N., Eide, C.A., Vasudevan, K.B., Loriaux, M.M., Firpo, E., Cortes, J.E., McWeeney, S., O'Hare, T., Roberts, J.M., Druker, B.J. and Deininger, M.W. (2014) 'BCR-ABL1 promotes leukemia by converting p27 into a cytoplasmic oncoprotein', *Blood*, 124(22), pp. 3260-73.

Ahmad, E.I., Gawish, H.H., Al Azizi, N.M. and Elhefni, A.M. (2011) 'The prognostic impact of K-RAS mutations in adult acute myeloid leukemia patients treated with high-dose cytarabine', *Onco Targets Ther*, 4, pp. 115-21.

Allen, R.J., Smith, S.D., Moldwin, R.L., Lu, M.M., Giordano, L., Vignon, C., Suto, Y., Harden, A., Tomek, R., Veldman, T., Ried, T., Larson, R.A., Le Beau, M.M., Rowley, J.D. and Zeleznik-Le, N. (1998) 'Establishment and characterization of a megakaryoblast cell line with amplification of MLL', *Leukemia*, 12(7), pp. 1119-27.

Anderlini, P., Luna, M., Kantarjian, H.M., O'Brien, S., Pierce, S., Keating, M.J. and Estey, E.H. (1996) 'Causes of initial remission induction failure in patients with acute myeloid leukemia and myelodysplastic syndromes', *Leukemia*, 10(4), pp. 600-8.

Andersen, M.K., Christiansen, D.H., Kirchhoff, M. and Pedersen-Bjergaard, J. (2001) 'Duplication or amplification of chromosome band 11q23, including the unrearranged MLL gene, is a recurrent abnormality in therapy-related MDS and AML, and is closely related to mutation of the TP53 gene and to previous therapy with alkylating agents', *Genes Chromosomes Cancer*, 31(1), pp. 33-41.

Andreasson, P., Johansson, B., Arheden, K., Billstrom, R., Mitelman, F. and Hoglund, M. (1997) 'Deletions of CDKN1B and ETV6 in acute myeloid leukemia and myelodysplastic syndromes

without cytogenetic evidence of 12p abnormalities', *Genes Chromosomes Cancer*, 19(2), pp. 77-83.

Andreu, E.J., Lledo, E., Poch, E., Ivorra, C., Albero, M.P., Martinez-Climent, J.A., Montiel-Duarte, C., Rifon, J., Perez-Calvo, J., Arbona, C., Prosper, F. and Perez-Roger, I. (2005) 'BCR-ABL induces the expression of Skp2 through the PI3K pathway to promote p27Kip1 degradation and proliferation of chronic myelogenous leukemia cells', *Cancer Res*, 65(8), pp. 3264-72.

Arabi, A., Wu, S., Ridderstrale, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlen, S., Hydbring, P., Soderberg, O., Grummt, I., Larsson, L.G. and Wright, A.P. (2005) 'c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription', *Nat Cell Biol*, 7(3), pp. 303-10.

Arhel, N. (2010) 'Revisiting HIV-1 uncoating', *Retrovirology*, 7, p. 96.

Arteaga, C.L. and Baselga, J. (2003) 'Clinical trial design and end points for epidermal growth factor receptor-targeted therapies: implications for drug development and practice', *Clin Cancer Res*, 9(5), pp. 1579-89.

Askew, D.S., Ashmun, R.A., Simmons, B.C. and Cleveland, J.L. (1991) 'Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis', *Oncogene*, 6(10), pp. 1915-22.

Augustin, M., Sedlmeier, R., Peters, T., Huffstadt, U., Kochmann, E., Simon, D., Schoniger, M., Garke-Mayerthaler, S., Laufs, J., Mayhaus, M., Franke, S., Klose, M., Graupner, A., Kurzmann, M., Zinser, C., Wolf, A., Voelkel, M., Kellner, M., Kilian, M., Seelig, S., Koppius, A., Teubner, A., Korthaus, D., Nehls, M. and Wattler, S. (2005) 'Efficient and fast targeted production of murine models based on ENU mutagenesis', *Mamm Genome*, 16(6), pp. 405-13.

Avet-Loiseau, H., Godon, C., Li, J.Y., Daviet, A., Mellerin, M.P., Talmant, P., Harousseau, J.L. and Bataille, R. (1999) 'Amplification of the 11q23 region in acute myeloid leukemia', *Genes Chromosomes Cancer*, 26(2), pp. 166-70.

Avruch, J., Khokhlatchev, A., Kyriakis, J.M., Luo, Z., Tzivion, G., Vavvas, D. and Zhang, X.F. (2001) 'Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade', *Recent Prog Horm Res*, 56, pp. 127-55.

Baens, M., Wlodarska, I., Corveleyn, A., Hoornaert, I., Hagemeijer, A. and Marynen, P. (1999) 'A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p)', *Genomics*, 56(1), pp. 40-50.

Balgobind, B.V., Hollink, I.H., Arentsen-Peters, S.T., Zimmermann, M., Harbott, J., Beverloo, H.B., von Bergh, A.R., Cloos, J., Kaspers, G.J., de Haas, V., Zemanova, Z., Stary, J., Cayuela, J.M., Baruchel, A., Creutzig, U., Reinhardt, D., Pieters, R., Zwaan, C.M. and van den Heuvel-Eibrink, M.M. (2011) 'Integrative analysis of type-I and type-II aberrations underscores the genetic heterogeneity of pediatric acute myeloid leukemia', *Haematologica*, 96(10), pp. 1478-87.

Barjesteh van Waalwijk van Doorn-Khosrovani, S., Spensberger, D., de Knegt, Y., Tang, M., Lowenberg, B. and Delwel, R. (2005) 'Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia', *Oncogene*, 24(25), pp. 4129-37.

Barnes, A., Pinder, S.E., Bell, J.A., Paish, E.C., Wencyk, P.M., Robertson, J.F., Elston, C.W. and Ellis, I.O. (2003) 'Expression of p27kip1 in breast cancer and its prognostic significance', *J Pathol*, 201(3), pp. 451-9.

Barzi, A. and Sekeres, M.A. (2010) 'Myelodysplastic syndromes: a practical approach to diagnosis and treatment', *Cleve Clin J Med*, 77(1), pp. 37-44.

Bayreuther, K. (1960) 'Chromosomes in primary neoplastic growth', *Nature*, 186, pp. 6-9.

Ben-Bassat, H., Korkesh, A., Voss, R., Leizerowitz, R. and Polliack, A. (1982) 'Establishment and characterization of a new permanent cell line (GDM-1) from a patient with myelomonoblastic leukemia', *Leuk Res*, 6(6), pp. 743-52.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. and Sultan, C. (1976) 'Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group', *Br J Haematol*, 33(4), pp. 451-8.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. and Sultan, C. (1985) 'Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group', *Ann Intern Med*, 103(4), pp. 620-5.

Bernstein, A., MacCormick, R. and Martin, G.S. (1976) 'Transformation-defective mutants of avian sarcoma viruses: the genetic relationship between conditional and nonconditional mutants', *Virology*, 70(1), pp. 206-9.

Besson, A., Assoian, R.K. and Roberts, J.M. (2004) 'Regulation of the cytoskeleton: an oncogenic function for CDK inhibitors?', *Nat Rev Cancer*, 4(12), pp. 948-55.

Betts, D.R., Ammann, R.A., Hirt, A., Hengartner, H., Beck-Popovic, M., Kuhne, T., Nobile, L., Caflisch, U., Wacker, P. and Niggli, F.K. (2007) 'The prognostic significance of cytogenetic aberrations in childhood acute myeloid leukaemia. A study of the Swiss Paediatric Oncology Group (SPOG)', *Eur J Haematol*, 78(6), pp. 468-76.

Beurlet, S., Omidvar, N., Gorombei, P., Krief, P., Le Pogam, C., Setterblad, N., de la Grange, P., Leboeuf, C., Janin, A., Noguera, M.E., Hervatin, F., Sarda-Mantel, L., Konopleva, M., Andreeff, M., Tu, A.W., Fan, A.C., Felsher, D.W., Whetton, A., Pla, M., West, R., Fenaux, P., Chomienne, C. and Padua, R.A. (2013) 'BCL-2 inhibition with ABT-737 prolongs survival in an NRAS/BCL-2 mouse model of AML by targeting primitive LSK and progenitor cells', *Blood*, 122(16), pp. 2864-76.

Bhayat, F., Das-Gupta, E., Smith, C., McKeever, T. and Hubbard, R. (2009) 'The incidence of and mortality from leukaemias in the UK: a general population-based study', *BMC Cancer*, 9, p. 252.

Bloom, J. and Pagano, M. (2003) 'Deregulated degradation of the cdk inhibitor p27 and malignant transformation', *Semin Cancer Biol*, 13(1), pp. 41-7.

Bohlander, S.K. (2005) 'ETV6: a versatile player in leukemogenesis', *Semin Cancer Biol*, 15(3), pp. 162-74.

Bomken, S., Buechler, L., Rehe, K., Ponthan, F., Elder, A., Blair, H., Bacon, C.M., Vormoor, J. and Heidenreich, O. (2013) 'Lentiviral marking of patient-derived acute lymphoblastic leukaemic cells allows in vivo tracking of disease progression', *Leukemia*, 27(3), pp. 718-21.

Bondar, T., Kalinina, A., Khair, L., Kopanja, D., Nag, A., Bagchi, S. and Raychaudhuri, P. (2006) 'Cul4A and DDB1 associate with Skp2 to target p27Kip1 for proteolysis involving the COP9 signalosome', *Mol Cell Biol*, 26(7), pp. 2531-9.

Bosma, G.C., Custer, R.P. and Bosma, M.J. (1983) 'A severe combined immunodeficiency mutation in the mouse', *Nature*, 301(5900), pp. 527-30.

Boveri, T. (1914) 'Zur Frage der Entstehung maligner Tumoren ', *Gustav Fischer*.

Bowen, D.T., Frew, M.E., Hills, R., Gale, R.E., Wheatley, K., Groves, M.J., Langabeer, S.E., Kottaridis, P.D., Moorman, A.V., Burnett, A.K. and Linch, D.C. (2005) 'RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years', *Blood*, 106(6), pp. 2113-9.

Burnett, A.K., Hills, R.K., Milligan, D.W., Goldstone, A.H., Prentice, A.G., McMullin, M.F., Duncombe, A., Gibson, B. and Wheatley, K. (2010) 'Attempts to optimize induction and consolidation treatment in acute myeloid leukemia: results of the MRC AML12 trial', *J Clin Oncol*, 28(4), pp. 586-95.

Byrd, J.C., Mrozek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S., Koduru, P.R., Moore, J.O., Stone, R.M., Mayer, R.J., Feldman, E.J., Davey, F.R., Schiffer, C.A., Larson, R.A., Bloomfield, C.D., Cancer and Leukemia

Group, B. (2002) 'Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461)', *Blood*, 100(13), pp. 4325-36.

Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H. and et al. (1990) 'Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus', *Cell*, 60(3), pp. 509-20.

Calzone, L., Gelay, A., Zinovyev, A., Radvanyi, F. and Barillot, E. (2008) 'A comprehensive modular map of molecular interactions in RB/E2F pathway', *Mol Syst Biol*, 4, p. 173.

Cao, T., Racz, P., Szauter, K.M., Groma, G., Nakamatsu, G.Y., Fogelgren, B., Pankotai, E., He, Q.P. and Csiszar, K. (2007) 'Mutation in Mpzl3, a novel [corrected] gene encoding a predicted [corrected] adhesion protein, in the rough coat (rc) mice with severe skin and hair abnormalities', *J Invest Dermatol*, 127(6), pp. 1375-86.

Cardis, E., Gilbert, E.S., Carpenter, L., Howe, G., Kato, I., Armstrong, B.K., Beral, V., Cowper, G., Douglas, A., Fix, J. and et al. (1995) 'Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries', *Radiat Res*, 142(2), pp. 117-32.

Caren, H., Holmstrand, A., Sjoberg, R.M. and Martinsson, T. (2006) 'The two human homologues of yeast UFD2 ubiquitination factor, UBE4A and UBE4B, are located in common neuroblastoma deletion regions and are subject to mutations in tumours', *Eur J Cancer*, 42(3), pp. 381-7.

Caspersson, T., Farber, S., Foley, G.E., Kudynowski, J., Modest, E.J., Simonsson, E., Wagh, U. and Zech, L. (1968) 'Chemical differentiation along metaphase chromosomes', *Exp Cell Res*, 49(1), pp. 219-22.

Caspersson, T., Zech, L. and Johansson, C. (1970) 'Differential binding of alkylating fluorochromes in human chromosomes', *Exp Cell Res*, 60(3), pp. 315-9.

Catzavelos, C., Tsao, M.S., DeBoer, G., Bhattacharya, N., Shepherd, F.A. and Slingerland, J.M. (1999) 'Reduced expression of the cell cycle inhibitor p27Kip1 in non-small cell lung carcinoma: a prognostic factor independent of Ras', *Cancer Res*, 59(3), pp. 684-8.

Chen, C., Liu, Y., Rappaport, A.R., Kitzing, T., Schultz, N., Zhao, Z., Shroff, A.S., Dickins, R.A., Vakoc, C.R., Bradner, J.E., Stock, W., LeBeau, M.M., Shannon, K.M., Kogan, S., Zuber, J. and

Lowe, S.W. (2014) 'MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia', *Cancer Cell*, 25(5), pp. 652-65.

Chen, Y., Cortes, J., Estrov, Z., Faderl, S., Qiao, W., Abruzzo, L., Garcia-Manero, G., Pierce, S., Huang, X., Kebriaei, P., Kadia, T., De Lima, M., Kantarjian, H. and Ravandi, F. (2011) 'Persistence of cytogenetic abnormalities at complete remission after induction in patients with acute myeloid leukemia: prognostic significance and the potential role of allogeneic stem-cell transplantation', *J Clin Oncol*, 29(18), pp. 2507-13.

Cheson, B.D., Cassileth, P.A., Head, D.R., Schiffer, C.A., Bennett, J.M., Bloomfield, C.D., Brunning, R., Gale, R.P., Grever, M.R., Keating, M.J. and et al. (1990) 'Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia', *J Clin Oncol*, 8(5), pp. 813-9.

Chipuk, J.E., Moldoveanu, T., Llambi, F., Parsons, M.J. and Green, D.R. (2010) 'The BCL-2 family reunion', *Mol Cell*, 37(3), pp. 299-310.

Cho, S., Kim, J.H., Back, S.H. and Jang, S.K. (2005) 'Polypyrimidine tract-binding protein enhances the internal ribosomal entry site-dependent translation of p27Kip1 mRNA and modulates transition from G1 to S phase', *Mol Cell Biol*, 25(4), pp. 1283-97.

Christiansen, D.H., Andersen, M.K. and Pedersen-Bjergaard, J. (2001) 'Mutations with loss of heterozygosity of p53 are common in therapy-related myelodysplasia and acute myeloid leukemia after exposure to alkylating agents and significantly associated with deletion or loss of 5q, a complex karyotype, and a poor prognosis', *J Clin Oncol*, 19(5), pp. 1405-13.

Chu, I., Sun, J., Arnaout, A., Kahn, H., Hanna, W., Narod, S., Sun, P., Tan, C.K., Hengst, L. and Slingerland, J. (2007) 'p27 phosphorylation by Src regulates inhibition of cyclin E-Cdk2', *Cell*, 128(2), pp. 281-94.

Chu, S., McDonald, T. and Bhatia, R. (2010) 'Role of BCR-ABL-Y177-mediated p27kip1 phosphorylation and cytoplasmic localization in enhanced proliferation of chronic myeloid leukemia progenitors', *Leukemia*, 24(4), pp. 779-87.

Cribbs, A.P., Kennedy, A., Gregory, B. and Brennan, F.M. (2013) 'Simplified production and concentration of lentiviral vectors to achieve high transduction in primary human T cells', *BMC Biotechnol*, 13, p. 98.

Cristobal, I., Blanco, F.J., Garcia-Orti, L., Marcotegui, N., Vicente, C., Rifon, J., Novo, F.J., Bandres, E., Calasanz, M.J., Bernabeu, C. and Odero, M.D. (2010) 'SETBP1 overexpression is a

novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia', *Blood*, 115(3), pp. 615-25.

Czernilofsky, A.P., Levinson, A.D., Varmus, H.E., Bishop, J.M., Tischer, E. and Goodman, H.M. (1980) 'Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product', *Nature*, 287(5779), pp. 198-203.

Czyzyk, T.A., Andrews, J.L., Coskun, T., Wade, M.R., Hawkins, E.D., Lockwood, J.F., Varga, G., Sahr, A.E., Chen, Y., Brozinick, J.T., Kikly, K. and Statnick, M.A. (2013) 'Genetic ablation of myelin protein zero-like 3 in mice increases energy expenditure, improves glycemic control, and reduces hepatic lipid synthesis', *Am J Physiol Endocrinol Metab*, 305(2), pp. E282-92.

Dash, A. and Gilliland, D.G. (2001) 'Molecular genetics of acute myeloid leukaemia', *Best Pract Res Clin Haematol*, 14(1), pp. 49-64.

Davies, J.A., Ladomery, M., Hohenstein, P., Michael, L., Shafe, A., Spraggon, L. and Hastie, N. (2004) 'Development of an siRNA-based method for repressing specific genes in renal organ culture and its use to show that the Wt1 tumour suppressor is required for nephron differentiation', *Hum Mol Genet*, 13(2), pp. 235-46.

Davis, H.E., Rosinski, M., Morgan, J.R. and Yarmush, M.L. (2004) 'Charged polymers modulate retrovirus transduction via membrane charge neutralization and virus aggregation', *Biophys J*, 86(2), pp. 1234-42.

DeFeo, D., Gonda, M.A., Young, H.A., Chang, E.H., Lowy, D.R., Scolnick, E.M. and Ellis, R.W. (1981) 'Analysis of two divergent rat genomic clones homologous to the transforming gene of Harvey murine sarcoma virus', *Proc Natl Acad Sci U S A*, 78(6), pp. 3328-32.

Demaison, C., Parsley, K., Brouns, G., Scherr, M., Battmer, K., Kinnon, C., Grez, M. and Thrasher, A.J. (2002) 'High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter', *Hum Gene Ther*, 13(7), pp. 803-13.

Denicourt, C., Saenz, C.C., Datnow, B., Cui, X.S. and Dowdy, S.F. (2007) 'Relocalized p27Kip1 tumor suppressor functions as a cytoplasmic metastatic oncogene in melanoma', *Cancer Res*, 67(19), pp. 9238-43.

Denkert, C., Weichert, W., Pest, S., Koch, I., Licht, D., Kobel, M., Reles, A., Sehouli, J., Dietel, M. and Hauptmann, S. (2004) 'Overexpression of the embryonic-lethal abnormal vision-like

protein HuR in ovarian carcinoma is a prognostic factor and is associated with increased cyclooxygenase 2 expression', *Cancer Res*, 64(1), pp. 189-95.

Dickie, M. (1966) 'Rough coat', *Mouse News Lett*, (34), p. 30.

Dijkers, P.F., Medema, R.H., Pals, C., Banerji, L., Thomas, N.S., Lam, E.W., Burgering, B.M., Raaijmakers, J.A., Lammers, J.W., Koenderman, L. and Coffer, P.J. (2000) 'Forkhead transcription factor FKHR-L1 modulates cytokine-dependent transcriptional regulation of p27(KIP1)', *Mol Cell Biol*, 20(24), pp. 9138-48.

Dolan, M., McGlennen, R.C. and Hirsch, B. (2002) 'MLL amplification in myeloid malignancies: clinical, molecular, and cytogenetic findings', *Cancer Genet Cytogenet*, 134(2), pp. 93-101.

Dolgin, E. (2011) 'Mouse library set to be knockout', *Nature*, 474(7351), pp. 262-3.

Dombret, H. (2011) 'Gene mutation and AML pathogenesis', *Blood*, 118(20), pp. 5366-7.

Doulatov, S., Notta, F., Laurenti, E. and Dick, J.E. (2012) 'Hematopoiesis: a human perspective', *Cell Stem Cell*, 10(2), pp. 120-36.

Dumaz, N., Light, Y. and Marais, R. (2002) 'Cyclic AMP blocks cell growth through Raf-1-dependent and Raf-1-independent mechanisms', *Mol Cell Biol*, 22(11), pp. 3717-28.

Eicher, E.F., S; Reynolds, S (1977) 'Rough coat on chromosome 9', *Mouse News Lett*, (56), p. 42.

Eilers, M. and Eisenman, R.N. (2008) 'Myc's broad reach', *Genes Dev*, 22(20), pp. 2755-66.

Ellis, B.L., Potts, P.R. and Porteus, M.H. (2011) 'Creating higher titer lentivirus with caffeine', *Hum Gene Ther*, 22(1), pp. 93-100.

Engelman, A. and Cherepanov, P. (2012) 'The structural biology of HIV-1: mechanistic and therapeutic insights', *Nat Rev Microbiol*, 10(4), pp. 279-90.

Esposito, V., Baldi, A., De Luca, A., Groger, A.M., Loda, M., Giordano, G.G., Caputi, M., Baldi, F., Pagano, M. and Giordano, A. (1997) 'Prognostic role of the cyclin-dependent kinase inhibitor p27 in non-small cell lung cancer', *Cancer Res*, 57(16), pp. 3381-5.

Estey, E. and Dohner, H. (2006) 'Acute myeloid leukaemia', *Lancet*, 368(9550), pp. 1894-907.

Fanales-Belasio, E., Raimondo, M., Suligoi, B. and Butto, S. (2010) 'HIV virology and pathogenetic mechanisms of infection: a brief overview', *Ann Ist Super Sanita*, 46(1), pp. 5-14.

Fathi, A.T. and Abdel-Wahab, O. (2012) 'Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy', *Adv Hematol*, 2012, p. 469592.

Fathi, A.T. and Chen, Y.B. (2011) 'Treatment of FLT3-ITD acute myeloid leukemia', *Am J Blood Res*, 1(2), pp. 175-89.

Felix, C.A., Megonigal, M.D., Chervinsky, D.S., Leonard, D.G., Tsuchida, N., Kakati, S., Block, A.M., Fisher, J., Grossi, M., Salhany, K.I., Jani-Sait, S.N. and Aplan, P.D. (1998) 'Association of germline p53 mutation with MLL segmental jumping translocation in treatment-related leukemia', *Blood*, 91(12), pp. 4451-6.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A., Kwon, J.S., Guan, Y., Sinha, N. and Zuber, J. (2013) 'An optimized microRNA backbone for effective single-copy RNAi', *Cell Rep*, 5(6), pp. 1704-13.

Felsher, D.W. and Bishop, J.M. (1999) 'Transient excess of MYC activity can elicit genomic instability and tumorigenesis', *Proc Natl Acad Sci U S A*, 96(7), pp. 3940-4.

Feurstein, S., Rucker, F.G., Bullinger, L., Hofmann, W., Manukjan, G., Gohring, G., Lehmann, U., Heuser, M., Ganser, A., Dohner, K., Schlegelberger, B. and Steinemann, D. (2014) 'Haploinsufficiency of ETV6 and CDKN1B in patients with acute myeloid leukemia and complex karyotype', *BMC Genomics*, 15, p. 784.

Finlay, C.A., Hinds, P.W. and Levine, A.J. (1989) 'The p53 proto-oncogene can act as a suppressor of transformation', *Cell*, 57(7), pp. 1083-93.

Frelin, C., Imbert, V., Griessinger, E., Peyron, A-C., Rochet, N., Philip, P., Dageville, C., Sirvent, A., Hummelsberger, M., Bérard, E., Dreano, M., Sirvent, N. and Peyron, J-F. (2005) 'Targeting NF-κB activation via pharmacologic inhibition of IKK2-induced apoptosis of human acute myeloid leukemia cells' *Blood*, 105(2), pp. 804-11.

Frohling, S., Scholl, C., Gilliland, D.G. and Levine, R.L. (2005) 'Genetics of myeloid malignancies: pathogenetic and clinical implications', *J Clin Oncol*, 23(26), pp. 6285-95.

Fullwood, M.J., Wei, C.L., Liu, E.T. and Ruan, Y. (2009) 'Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses', *Genome Res*, 19(4), pp. 521-32.

Furley, A.J., Reeves, B.R., Mizutani, S., Altass, L.J., Watt, S.M., Jacob, M.C., van den Elsen, P., Terhorst, C. and Greaves, M.F. (1986) 'Divergent molecular phenotypes of KG1 and KG1a myeloid cell lines', *Blood*, 68(5), pp. 1101-7.

Gaidzik, V.I., Schlenk, R.F., Moschny, S., Becker, A., Bullinger, L., Corbacioglu, A., Krauter, J., Schlegelberger, B., Ganser, A., Dohner, H., Dohner, K. and German-Austrian, A.M.L.S.G. (2009) 'Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group', *Blood*, 113(19), pp. 4505-11.

Garnis, C., Buys, T.P. and Lam, W.L. (2004) 'Genetic alteration and gene expression modulation during cancer progression', *Mol Cancer*, 3, p. 9.

Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H. and Bruns, G.A. (1990) 'Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping', *Nature*, 343(6260), pp. 774-8.

Gillet, J.P., Varma, S. and Gottesman, M.M. (2013) 'The clinical relevance of cancer cell lines', *J Natl Cancer Inst*, 105(7), pp. 452-8.

Gillett, C.E., Smith, P., Peters, G., Lu, X. and Barnes, D.M. (1999) 'Cyclin-dependent kinase inhibitor p27Kip1 expression and interaction with other cell cycle-associated proteins in mammary carcinoma', *J Pathol*, 187(2), pp. 200-6.

Gilliland, D.G. and Griffin, J.D. (2002) 'The roles of FLT3 in hematopoiesis and leukemia', *Blood*, 100(5), pp. 1532-42.

Golde, A. (1970) 'Radio-induced mutants of the Schmidt-Ruppin strain of rous sarcoma virus', *Virology*, 40(4), pp. 1022-9.

Gomez, C. and Hope, T.J. (2005) 'The ins and outs of HIV replication', *Cell Microbiol*, 7(5), pp. 621-6.

Gomez-Roman, N., Grandori, C., Eisenman, R.N. and White, R.J. (2003) 'Direct activation of RNA polymerase III transcription by c-Myc', *Nature*, 421(6920), pp. 290-4.

Grimmler, M., Wang, Y., Mund, T., Cilensek, Z., Keidel, E.M., Waddell, M.B., Jakel, H., Kullmann, M., Kriwacki, R.W. and Hengst, L. (2007) 'Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases', *Cell*, 128(2), pp. 269-80.

Grimwade, D., Hills, R.K., Moorman, A.V., Walker, H., Chatters, S., Goldstone, A.H., Wheatley, K., Harrison, C.J., Burnett, A.K. and National Cancer Research Institute Adult Leukaemia Working, G. (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), pp. 354-65.

Grimwade, D., Walker, H., Harrison, G., Oliver, F., Chatters, S., Harrison, C.J., Wheatley, K., Burnett, A.K., Goldstone, A.H. and Medical Research Council Adult Leukemia Working, P. (2001) 'The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial', *Blood*, 98(5), pp. 1312-20.

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), pp. 2322-33.

Gundestrup, M. and Storm, H.H. (1999) 'Radiation-induced acute myeloid leukaemia and other cancers in commercial jet cockpit crew: a population-based cohort study', *Lancet*, 354(9195), pp. 2029-31.

Haber, D.A., Buckler, A.J., Glaser, T., Call, K.M., Pelletier, J., Sohn, R.L., Douglass, E.C. and Housman, D.E. (1990) 'An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor', *Cell*, 61(7), pp. 1257-69.

Haferlach, C., Bacher, U., Kohlmann, A., Schindela, S., Alpermann, T., Kern, W., Schnittger, S. and Haferlach, T. (2011) 'CDKN1B, encoding the cyclin-dependent kinase inhibitor 1B (p27), is located in the minimally deleted region of 12p abnormalities in myeloid malignancies and its low expression is a favorable prognostic marker in acute myeloid leukemia', *Haematologica*, 96(6), pp. 829-36.

Haferlach, C., Dicker, F., Herholz, H., Schnittger, S., Kern, W. and Haferlach, T. (2008) 'Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype', *Leukemia*, 22(8), pp. 1539-41.

Haferlach, C., Kern, W., Schindela, S., Kohlmann, A., Alpermann, T., Schnittger, S. and Haferlach, T. (2012) 'Gene expression of BAALC, CDKN1B, ERG, and MN1 adds independent prognostic information to cytogenetics and molecular mutations in adult acute myeloid leukemia', *Genes Chromosomes Cancer*, 51(3), pp. 257-65.

Haferlach, C., Kohlmann, A., Schindela, S., Weiss, T., Kern, W., Schnittger, S. and Haferlach, T. (2009) 'CDKN1B Encoding the Cyclin-Dependent Kinase Inhibitor 1B (p27) but Not ETV6 Is Located in the Minimally Deleted Region of 12p Abnormalities in Myeloid Malignancies and Its Expression Is Associated with Outcome in Acute Myeloid Leukemia (AML)', *Blood*, 114(22), pp. 116-117.

Hansemann, V. (1890) *Virchows Arch. A Pathol. Anat.* , (119), pp. 299–326.

Hanzelmann, P., Stingele, J., Hofmann, K., Schindelin, H. and Raasi, S. (2010) 'The yeast E4 ubiquitin ligase Ufd2 interacts with the ubiquitin-like domains of Rad23 and Dsk2 via a novel and distinct ubiquitin-like binding domain', *J Biol Chem*, 285(26), pp. 20390-8.

Harrison, C.J., Hills, R.K., Moorman, A.V., Grimwade, D.J., Hann, I., Webb, D.K., Wheatley, K., de Graaf, S.S., van den Berg, E., Burnett, A.K. and Gibson, B.E. (2010) 'Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12', *J Clin Oncol*, 28(16), pp. 2674-81.

Hatzimichael, E., Georgiou, G., Benetatos, L. and Briasoulis, E. (2013) 'Gene mutations and molecularly targeted therapies in acute myeloid leukemia', *Am J Blood Res*, 3(1), pp. 29-51.

Hausser, J. and Zavolan, M. (2014) 'Identification and consequences of miRNA-target interactions--beyond repression of gene expression', *Nat Rev Genet*, 15(9), pp. 599-612.

Hayashi, H., Ogawa, N., Ishiwa, N., Yazawa, T., Inayama, Y., Ito, T. and Kitamura, H. (2001) 'High cyclin E and low p27/Kip1 expressions are potentially poor prognostic factors in lung adenocarcinoma patients', *Lung Cancer*, 34(1), pp. 59-65.

Hayashi, K., Cao, T., Passmore, H., Jourdan-Le Saux, C., Fogelgren, B., Khan, S., Hornstra, I., Kim, Y., Hayashi, M. and Csiszar, K. (2004) 'Progressive hair loss and myocardial degeneration in rough coat mice: reduced lysyl oxidase-like (LOXL) in the skin and heart', *J Invest Dermatol*, 123(5), pp. 864-71.

Hayward, W.S., Neel, B.G. and Astrin, S.M. (1981) 'Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis', *Nature*, 290(5806), pp. 475-80.

Hellen, C.U. and Sarnow, P. (2001) 'Internal ribosome entry sites in eukaryotic mRNA molecules', *Genes Dev*, 15(13), pp. 1593-612.

Hohenstein, P. and Hastie, N.D. (2006) 'The many facets of the Wilms' tumour gene, WT1', *Hum Mol Genet*, 15 Spec No 2, pp. R196-201.

Hope, K.J., Cellot, S., Ting, S.B., MacRae, T., Mayotte, N., Iscove, N.N. and Sauvageau, G. (2010) 'An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity', *Cell Stem Cell*, 7(1), pp. 101-13.

Hulea, L. and Nepveu, A. (2012) 'CUX1 transcription factors: from biochemical activities and cell-based assays to mouse models and human diseases', *Gene*, 497(1), pp. 18-26.

International Human Genome Sequencing, C. (2004) 'Finishing the euchromatic sequence of the human genome', *Nature*, 431(7011), pp. 931-45.

Iritani, B.M. and Eisenman, R.N. (1999) 'c-Myc enhances protein synthesis and cell size during B lymphocyte development', *Proc Natl Acad Sci U S A*, 96(23), pp. 13180-5.

Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., Ueyama, Y., Koyanagi, Y., Sugamura, K., Tsuji, K., Heike, T. and Nakahata, T. (2002) 'NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells', *Blood*, 100(9), pp. 3175-82.

Ivanov X., M.Z., Nedyalkov S., Todorov T.G. (1964) 'Experimental investigations into avian leucoses. V. Transmission, haematology and morphology of avian myelocytomatosis.', *Bull. Inst. Pathol. Comp. Anim. Acad. Bulg. Sci.* , 10:5–38.

Ivics, Z., Hackett, P.B., Plasterk, R.H. and Izsvak, Z. (1997) 'Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells', *Cell*, 91(4), pp. 501-10.

Jaiswal, S., Jamieson, C.H., Pang, W.W., Park, C.Y., Chao, M.P., Majeti, R., Traver, D., van Rooijen, N. and Weissman, I.L. (2009) 'CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis', *Cell*, 138(2), pp. 271-85.

Jin, Z., Maiti, S., Huls, H., Singh, H., Olivares, S., Mates, L., Izsvak, Z., Ivics, Z., Lee, D.A., Champlin, R.E. and Cooper, L.J. (2011) 'The hyperactive Sleeping Beauty transposase SB100X improves the genetic modification of T cells to express a chimeric antigen receptor', *Gene Ther*, 18(9), pp. 849-56.

Johansen, L.M., Iwama, A., Lodie, T.A., Sasaki, K., Felsher, D.W., Golub, T.R. and Tenen, D.G. (2001) 'c-Myc is a critical target for c/EBPalpha in granulopoiesis', *Mol Cell Biol*, 21(11), pp. 3789-806.

Joshua F. Zeidner, M.C.F., Amanda Blackford, Mark Robert Litzow, Lawrence Morris, Stephen Anthony Strickland, Jeffrey E. Lancet, Prithviraj Bose, M. Yair Levy, Raoul Tibes, Ivana Gojo, Christopher D Gocke, Gary L. Rosner, Jacqueline Greer, Joan M Cain, Richard F. Little, John Joseph Wright, L. Austin Doyle, B Douglas Smith, Judith E. Karp (2014) 'Randomized multicenter phase II trial of timed-sequential therapy with flavopiridol (alvocidib), cytarabine, and mitoxantrone (FLAM) versus “7+3” for adults with newly diagnosed acute myeloid leukemia (AML).', *J Clin Oncol*, (32), p. 5s.

Justice, M.J. (2000) 'Capitalizing on large-scale mouse mutagenesis screens', *Nat Rev Genet*, 1(2), pp. 109-15.

Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J.C., Valent, A., Minty, A., Chalon, P., Lelias, J.M., Dumont, X., Ferrara, P., McKeon, F. and Caput, D. (1997) 'Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers', *Cell*, 90(4), pp. 809-19.

Kane, E.V., Roman, E., Cartwright, R., Parker, J. and Morgan, G. (1999) 'Tobacco and the risk of acute leukaemia in adults', *Br J Cancer*, 81(7), pp. 1228-33.

Karakas, T., Maurer, U., Weidmann, E., Miething, C.C., Hoelzer, D. and Bergmann, L. (1998) 'High expression of bcl-2 mRNA as a determinant of poor prognosis in acute myeloid leukemia', *Ann Oncol*, 9(2), pp. 159-65.

Karnoub, A.E. and Weinberg, R.A. (2008) 'Ras oncogenes: split personalities', *Nat Rev Mol Cell Biol*, 9(7), pp. 517-31.

Kataoka, T., Morishita, Y., Ogura, M., Morishima, Y., Towatari, M., Kato, Y., Inoue, H. and Saito, H. (1990) 'A novel human myeloid leukemia cell line, NKM-1, coexpressing granulocyte colony-stimulating factor receptors and macrophage colony-stimulating factor receptors', *Cancer Res*, 50(23), pp. 7703-9.

Keays, D.A., Clark, T.G., Campbell, T.G., Broxholme, J. and Valdar, W. (2007) 'Estimating the number of coding mutations in genotypic and phenotypic driven N-ethyl-N-nitrosourea (ENU) screens: revisited', *Mamm Genome*, 18(2), pp. 123-4.

Keller, U.B., Old, J.B., Dorsey, F.C., Nilsson, J.A., Nilsson, L., MacLean, K.H., Chung, L., Yang, C., Spruck, C., Boyd, K., Reed, S.I. and Cleveland, J.L. (2007) 'Myc targets Cks1 to provoke the suppression of p27Kip1, proliferation and lymphomagenesis', *EMBO J*, 26(10), pp. 2562-74.

Kelly, L.M. and Gilliland, D.G. (2002) 'Genetics of myeloid leukemias', *Annu Rev Genomics Hum Genet*, 3, pp. 179-98.

Klampfer, L., Zhang, J., Zelenetz, A.O., Uchida, H. and Nimer, S.D. (1996) 'The AML1/ETO fusion protein activates transcription of BCL-2', *Proc Natl Acad Sci U S A*, 93(24), pp. 14059-64.

Knudson, A.G., Jr. (1971) 'Mutation and cancer: statistical study of retinoblastoma', *Proc Natl Acad Sci U S A*, 68(4), pp. 820-3.

Knudson, A.G. (2001) 'Two genetic hits (more or less) to cancer', *Nat Rev Cancer*, 1(2), pp. 157-62.

Knutsen, T., Pack, S., Petropavlovskaja, M., Padilla-Nash, H., Knight, C., Mickley, L.A., Ried, T., Elwood, P.C. and Roberts, S.J. (2003) 'Cytogenetic, spectral karyotyping, fluorescence in situ hybridization, and comparative genomic hybridization characterization of two new secondary leukemia cell lines with 5q deletions, and MYC and MLL amplification', *Genes Chromosomes Cancer*, 37(3), pp. 270-81.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H.D., Mayer, T.U. and Jentsch, S. (1999) 'A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly', *Cell*, 96(5), pp. 635-44.

Kohl, T.M., Hellinger, C., Ahmed, F., Buske, C., Hiddemann, W., Bohlander, S.K. and Spiekermann, K. (2007) 'BH3 mimetic ABT-737 neutralizes resistance to FLT3 inhibitor treatment mediated by FLT3-independent expression of BCL2 in primary AML blasts', *Leukemia*, 21(8), pp. 1763-72.

Komander, D. (2009) 'The emerging complexity of protein ubiquitination', *Biochem Soc Trans*, 37(Pt 5), pp. 937-53.

Komander, D. and Rape, M. (2012) 'The ubiquitin code', *Annu Rev Biochem*, 81, pp. 203-29.

Kool, J. and Berns, A. (2009) 'High-throughput insertional mutagenesis screens in mice to identify oncogenic networks', *Nat Rev Cancer*, 9(6), pp. 389-99.

Korkmaz, H., Du, W., Yoo, G.H., Enamorado, II, Lin, H.S., Adsay, V., Kewson, D., Ensley, J.F., Shibuya, T.Y., Jacobs, J.R. and Kim, H. (2005) 'Prognostic significance of G1 cell-cycle inhibitors in early laryngeal cancer', *Am J Otolaryngol*, 26(2), pp. 77-82.

Kornblau, S.M., Thall, P.F., Estrov, Z., Walterscheid, M., Patel, S., Theriault, A., Keating, M.J., Kantarjian, H., Estey, E. and Andreeff, M. (1999) 'The prognostic impact of BCL2 protein expression in acute myelogenous leukemia varies with cytogenetics', *Clin Cancer Res*, 5(7), pp. 1758-66.

Kress, M., May, E., Cassingena, R. and May, P. (1979) 'Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum', *J Virol*, 31(2), pp. 472-83.

Kronke, J., Schlenk, R.F., Jensen, K.O., Tschurtz, F., Corbacioglu, A., Gaidzik, V.I., Paschka, P., Onken, S., Eiwen, K., Habdank, M., Spath, D., Lubbert, M., Wattad, M., Kindler, T., Salih, H.R., Held, G., Nachbaur, D., von Lilienfeld-Toal, M., Germing, U., Haase, D., Mergenthaler, H.G., Krauter, J., Ganser, A., Gohring, G., Schlegelberger, B., Dohner, H. and Dohner, K. (2011) 'Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group', *J Clin Oncol*, 29(19), pp. 2709-16.

Kullmann, M., Gopfert, U., Siewe, B. and Hengst, L. (2002) 'ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5'UTR', *Genes Dev*, 16(23), pp. 3087-99.

Kumar, C.C. (2011) 'Genetic abnormalities and challenges in the treatment of acute myeloid leukemia', *Genes Cancer*, 2(2), pp. 95-107.

Kundu, M. and Liu, P.P. (2001) 'Function of the inv(16) fusion gene CBFB-MYH11', *Curr Opin Hematol*, 8(4), pp. 201-5.

Kutner, R.H., Zhang, X.Y. and Reiser, J. (2009) 'Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors', *Nat Protoc*, 4(4), pp. 495-505.

Ladomery, M., Sommerville, J., Woolner, S., Slight, J. and Hastie, N. (2003) 'Expression in Xenopus oocytes shows that WT1 binds transcripts in vivo, with a central role for zinc finger one', *J Cell Sci*, 116(Pt 8), pp. 1539-49.

LaFramboise, T. (2009) 'Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances', *Nucleic Acids Res*, 37(13), pp. 4181-93.

Lane, D.P. and Crawford, L.V. (1979) 'T antigen is bound to a host protein in SV40-transformed cells', *Nature*, 278(5701), pp. 261-3.

Langer, R., Von Rahden, B.H., Nahrig, J., Von Weyhern, C., Reiter, R., Feith, M., Stein, H.J., Siewert, J.R., Hofler, H. and Sarbia, M. (2006) 'Prognostic significance of expression patterns of c-erbB-2, p53, p16INK4A, p27KIP1, cyclin D1 and epidermal growth factor receptor in oesophageal adenocarcinoma: a tissue microarray study', *J Clin Pathol*, 59(6), pp. 631-4.

Langer-Safer, P.R., Levine, M. and Ward, D.C. (1982) 'Immunological method for mapping genes on Drosophila polytene chromosomes', *Proc Natl Acad Sci U S A*, 79(14), pp. 4381-5.

le Sage, C., Nagel, R. and Agami, R. (2007) 'Diverse ways to control p27Kip1 function: miRNAs come into play', *Cell Cycle*, 6(22), pp. 2742-9.

Lee, T., Shah, C. and Xu, E.Y. (2007) 'Gene trap mutagenesis: a functional genomics approach towards reproductive research', *Mol Hum Reprod*, 13(11), pp. 771-9.

Li, Z. and Hann, S.R. (2009) 'The Myc-nucleophosmin-ARF network: a complex web unveiled', *Cell Cycle*, 8(17), pp. 2703-7.

Liang, D.C., Liu, H.C., Yang, C.P., Jaing, T.H., Hung, I.J., Yeh, T.C., Chen, S.H., Hou, J.Y., Huang, Y.J., Shih, Y.S., Huang, Y.H., Lin, T.H. and Shih, L.Y. (2013) 'Cooperating gene mutations in childhood acute myeloid leukemia with special reference on mutations of ASXL1, TET2, IDH1, IDH2, and DNMT3A', *Blood*, 121(15), pp. 2988-95.

Limaye, A., Hall, B. and Kulkarni, A.B. (2009) 'Manipulation of mouse embryonic stem cells for knockout mouse production', *Curr Protoc Cell Biol*, Chapter 19, pp. Unit 19 13 19 13 1-24.

Linzer, D.I. and Levine, A.J. (1979) 'Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells', *Cell*, 17(1), pp. 43-52.

Loda, M., Cukor, B., Tam, S.W., Lavin, P., Fiorentino, M., Draetta, G.F., Jessup, J.M. and Pagano, M. (1997) 'Increased proteasome-dependent degradation of the cyclin-dependent kinase inhibitor p27 in aggressive colorectal carcinomas', *Nat Med*, 3(2), pp. 231-4.

Lopez de Silanes, I., Fan, J., Yang, X., Zonderman, A.B., Potapova, O., Pizer, E.S. and Gorospe, M. (2003) 'Role of the RNA-binding protein HuR in colon carcinogenesis', *Oncogene*, 22(46), pp. 7146-54.

Lorenz, E., Uphoff, D., Reid, T.R. and Shelton, E. (1951) 'Modification of irradiation injury in mice and guinea pigs by bone marrow injections', *J Natl Cancer Inst*, 12(1), pp. 197-201.

Lund, K., Adams, P.D. and Copland, M. (2014) 'EZH2 in normal and malignant hematopoiesis', *Leukemia*, 28(1), pp. 44-9.

Luo, S., Yu, K., Yan, Q.X., Shen, Z.J., Wu, J.B., Chen, H.M. and Gao, S.M. (2014) 'Analysis of WT1 mutations, expression levels and single nucleotide polymorphism rs16754 in de novo non-M3 acute myeloid leukemia', *Leuk Lymphoma*, 55(2), pp. 349-57.

Mai, S., Hanley-Hyde, J. and Fluri, M. (1996) 'c-Myc overexpression associated DHFR gene amplification in hamster, rat, mouse and human cell lines', *Oncogene*, 12(2), pp. 277-88.

Maitta, R.W., Cannizzaro, L.A. and Ramesh, K.H. (2009) 'Association of MLL amplification with poor outcome in acute myeloid leukemia', *Cancer Genet Cytogenet*, 192(1), pp. 40-3.

Manne, U., Jhala, N.C., Jones, J., Weiss, H.L., Chatla, C., Meleth, S., Suarez-Cuervo, C. and Grizzle, W.E. (2004) 'Prognostic significance of p27(kip-1) expression in colorectal adenocarcinomas is associated with tumor stage', *Clin Cancer Res*, 10(5), pp. 1743-52.

Mao, X., Young, B.D. and Lu, Y.J. (2007) 'The application of single nucleotide polymorphism microarrays in cancer research', *Curr Genomics*, 8(4), pp. 219-28.

Mardis, E.R. (2008) 'Next-generation DNA sequencing methods', *Annu Rev Genomics Hum Genet*, 9, pp. 387-402.

Martin, G.S. (2001) 'The hunting of the Src', *Nat Rev Mol Cell Biol*, 2(6), pp. 467-75.

Masuda, H., Miller, C., Koeffler, H.P., Battifora, H. and Cline, M.J. (1987) 'Rearrangement of the p53 gene in human osteogenic sarcomas', *Proc Natl Acad Sci U S A*, 84(21), pp. 7716-9.

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z. and Izsvak, Z. (2009) 'Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates', *Nat Genet*, 41(6), pp. 753-61.

Mbisa, J.L., Barr, R., Thomas, J.A., Vandegraaff, N., Dorweiler, I.J., Svarovskaia, E.S., Brown, W.L., Mansky, L.M., Gorelick, R.J., Harris, R.S., Engelman, A. and Pathak, V.K. (2007) 'Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration', *J Virol*, 81(13), pp. 7099-110.

McCune, J.M., Namikawa, R., Kaneshima, H., Shultz, L.D., Lieberman, M. and Weissman, I.L. (1988) 'The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function', *Science*, 241(4873), pp. 1632-9.

McNerney, M.E., Brown, C.D., Wang, X., Bartom, E.T., Karmakar, S., Bandlamudi, C., Yu, S., Ko, J., Sandall, B.P., Stricker, T., Anastasi, J., Grossman, R.L., Cunningham, J.M., Le Beau, M.M. and White, K.P. (2013) 'CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia', *Blood*, 121(6), pp. 975-83.

Medema, R.H., Kops, G.J., Bos, J.L. and Burgering, B.M. (2000) 'AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27kip1', *Nature*, 404(6779), pp. 782-7.

Melikyan, G.B., Markosyan, R.M., Hemmati, H., Delmedico, M.K., Lambert, D.M. and Cohen, F.S. (2000) 'Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion', *J Cell Biol*, 151(2), pp. 413-23.

Meng, Q., Hagemeier, S.R., Kuny, C.V., Kalejta, R.F. and Kenney, S.C. (2010) 'Simian virus 40 T/t antigens and lamin A/C small interfering RNA rescue the phenotype of an Epstein-Barr virus protein kinase (BGLF4) mutant', *J Virol*, 84(9), pp. 4524-33.

Metzger, M.B. and Weissman, A.M. (2010) 'Working on a chain: E3s ganging up for ubiquitylation', *Nat Cell Biol*, 12(12), pp. 1124-6.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., Delabesse, E., de Oliveira, M.P., Cave, H., Clappier, E., van Dongen, J.J., Balgobind, B.V., van den Heuvel-Eibrink, M.M., Beverloo, H.B., Panzer-Grumayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L.C., Yip, S.F., Krzywinski, M., Eckert, C., Moricke, A., Schrappe, M., Alonso, C.N., Schafer, B.W., Krauter, J., Lee, D.A., Zur Stadt, U., Te Kronnie, G., Sutton, R.,

Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsaur, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dingermann, T., Klingebiel, T. and Marschalek, R. (2009) 'New insights to the MLL recombinome of acute leukemias', *Leukemia*, 23(8), pp. 1490-9.

Meyer, L.H. and Debatin, K.M. (2011) 'Diversity of human leukemia xenograft mouse models: implications for disease biology', *Cancer Res*, 71(23), pp. 7141-4.

Michaux, L., Wlodarska, I., Stul, M., Dierlamm, J., Mugneret, F., Herens, C., Beverloo, B., Verhest, A., Verellen-Dumoulin, C., Verhoef, G., Selleslag, D., Madoe, V., Lecomte, M., Deprijck, B., Ferrant, A., Delannoy, A., Marichal, S., Duhem, C., Dicato, M. and Hagemeijer, A. (2000) 'MLL amplification in myeloid leukemias: A study of 14 cases with multiple copies of 11q23', *Genes Chromosomes Cancer*, 29(1), pp. 40-7.

Miller, P.G., Al-Shahrour, F., Hartwell, K.A., Chu, L.P., Jaras, M., Puram, R.V., Puissant, A., Callahan, K.P., Ashton, J., McConkey, M.E., Poveromo, L.P., Cowley, G.S., Kharas, M.G., Labelle, M., Shterental, S., Fujisaki, J., Silberstein, L., Alexe, G., Al-Hajj, M.A., Shelton, C.A., Armstrong, S.A., Root, D.E., Scadden, D.T., Hynes, R.O., Mukherjee, S., Stegmaier, K., Jordan, C.T. and Ebert, B.L. (2013) 'In Vivo RNAi screening identifies a leukemia-specific dependence on integrin beta 3 signaling', *Cancer Cell*, 24(1), pp. 45-58.

Mohamed, A. (2011) 'MLL amplification in leukemia', *Atlas Genet Cytogenet Oncol Haematol*.

Moller, M.B., Skjodt, K., Mortensen, L.S. and Pedersen, N.T. (1999) 'Clinical significance of cyclin-dependent kinase inhibitor p27Kip1 expression and proliferation in non-Hodgkin's lymphoma: independent prognostic value of p27Kip1', *Br J Haematol*, 105(3), pp. 730-6.

Montpetit, A., Larose, J., Boily, G., Langlois, S., Trudel, N. and Sinnett, D. (2004) 'Mutational and expression analysis of the chromosome 12p candidate tumor suppressor genes in pre-B acute lymphoblastic leukemia', *Leukemia*, 18(9), pp. 1499-504.

Moore, A.W., Schedl, A., McInnes, L., Doyle, M., Hecksher-Sorensen, J. and Hastie, N.D. (1998) 'YAC transgenic analysis reveals Wilms' tumour 1 gene activity in the proliferating coelomic epithelium, developing diaphragm and limb', *Mech Dev*, 79(1-2), pp. 169-84.

Moore, M.D., Nikolaitchik, O.A., Chen, J., Hammarskjold, M.L., Rekosh, D. and Hu, W.S. (2009) 'Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses', *PLoS Pathog*, 5(10), p. e1000627.

Mosier, D.E., Gulizia, R.J., Baird, S.M. and Wilson, D.B. (1988) 'Transfer of a functional human immune system to mice with severe combined immunodeficiency', *Nature*, 335(6187), pp. 256-9.

Mosier, D.E., Gulizia, R.J., Baird, S.M., Wilson, D.B., Spector, D.H. and Spector, S.A. (1991) 'Human immunodeficiency virus infection of human-PBL-SCID mice', *Science*, 251(4995), pp. 791-4.

Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., et al. (2002) 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420(6915), pp. 520-62.

Mowat, M., Cheng, A., Kimura, N., Bernstein, A. and Benchimol, S. (1985) 'Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed by Friend virus', *Nature*, 314(6012), pp. 633-6.

Muller-Tidow, C., Steffen, B., Cauvet, T., Tickenbrock, L., Ji, P., Diederichs, S., Sargin, B., Kohler, G., Stelljes, M., Puccetti, E., Ruthardt, M., deVos, S., Hiebert, S.W., Koeffler, H.P., Berdel, W.E. and Serve, H. (2004) 'Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells', *Mol Cell Biol*, 24(7), pp. 2890-904.

Murphy, D.J., Junttila, M.R., Pouyet, L., Karnezis, A., Shchors, K., Bui, D.A., Brown-Swigart, L., Johnson, L. and Evan, G.I. (2008) 'Distinct thresholds govern Myc's biological output in vivo', *Cancer Cell*, 14(6), pp. 447-57.

Myllykangas, S., Himberg, J., Bohling, T., Nagy, B., Hollmen, J. and Knuutila, S. (2006) 'DNA copy number amplification profiling of human neoplasms', *Oncogene*, 25(55), pp. 7324-32.

Nakanishi, M., Tanaka, K., Shintani, T., Takahashi, T. and Kamada, N. (1999) 'Chromosomal instability in acute myelocytic leukemia and myelodysplastic syndrome patients among atomic bomb survivors', *J Radiat Res*, 40(2), pp. 159-67.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M. and Trono, D. (1996) 'In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector', *Science*, 272(5259), pp. 263-7.

Neubauer, A., Maharry, K., Mrozek, K., Thiede, C., Marcucci, G., Paschka, P., Mayer, R.J., Larson, R.A., Liu, E.T. and Bloomfield, C.D. (2008) 'Patients with acute myeloid leukemia and RAS mutations benefit most from postremission high-dose cytarabine: a Cancer and Leukemia Group B study', *J Clin Oncol*, 26(28), pp. 4603-9.

Nguyen, N., Judd, L.M., Kalantzis, A., Whittle, B., Giraud, A.S. and van Driel, I.R. (2011) 'Random mutagenesis of the mouse genome: a strategy for discovering gene function and the molecular basis of disease', *Am J Physiol Gastrointest Liver Physiol*, 300(1), pp. G1-11.

Nomdedeu, J.F., Hoyos, M., Carricondo, M., Bussaglia, E., Estivill, C., Esteve, J., Tormo, M., Duarte, R., Salamero, O., de Llano, M.P., Garcia, A., Bargay, J., Heras, I., Marti-Tutusaus, J.M., Llorente, A., Ribera, J.M., Gallardo, D., Aventin, A., Brunet, S., Sierra, J. and Group, C. (2013) 'Bone marrow WT1 levels at diagnosis, post-induction and post-intensification in adult de novo AML', *Leukemia*, 27(11), pp. 2157-64.

Nowell, P.C. (1985) 'Citation Classic - a Minute Chromosome in Human Chronic Granulocytic-Leukemia', *Current Contents/Life Sciences*, (8), pp. 19-19.

O'Donnell, K.A., Keng, V.W., York, B., Reineke, E.L., Seo, D., Fan, D., Silverstein, K.A., Schrum, C.T., Xie, W.R., Mularoni, L., Wheelan, S.J., Torbenson, M.S., O'Malley, B.W., Largaespada, D.A. and Boeke, J.D. (2012) 'A Sleeping Beauty mutagenesis screen reveals a tumor suppressor role for Ncoa2/Src-2 in liver cancer', *Proc Natl Acad Sci U S A*, 109(21), pp. E1377-86.

O'Hagan, R.C., Ohh, M., David, G., de Alboran, I.M., Alt, F.W., Kaelin, W.G., Jr. and DePinho, R.A. (2000) 'Myc-enhanced expression of Cul1 promotes ubiquitin-dependent proteolysis and cell cycle progression', *Genes Dev*, 14(17), pp. 2185-91.

Oka, Y., Nakajima, K., Nagao, K., Miura, K., Ishii, N. and Kobayashi, H. (2010) '293FT cells transduced with four transcription factors (OCT4, SOX2, NANOG, and LIN28) generate aberrant ES-like cells', *J Stem Cells Regen Med*, 6(3), pp. 149-56.

Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 mutations in human cancers: origins, consequences, and clinical use', *Cold Spring Harb Perspect Biol*, 2(1), p. a001008.

Oster, S.K., Ho, C.S., Soucie, E.L. and Penn, L.Z. (2002) 'The myc oncogene: MarvelouslY Complex', *Adv Cancer Res*, 84, pp. 81-154.

Painter, T.S. (1921) 'The Y-Chromosome in Mammals', *Science*, 53(1378), pp. 503-4.

Pan, R., Hogdal, L.J., Benito, J.M., Bucci, D., Han, L., Borthakur, G., Cortes, J., DeAngelo, D.J., Debose, L., Mu, H., Dohner, H., Gaidzik, V.I., Galinsky, I., Golfman, L.S., Haferlach, T., Harutyunyan, K.G., Hu, J., Leverson, J.D., Marcucci, G., Muschen, M., Newman, R., Park, E., Ruvolo, P.P., Ruvolo, V., Ryan, J., Schindela, S., Zweidler-McKay, P., Stone, R.M., Kantarjian, H., Andreeff, M., Konopleva, M. and Letai, A.G. (2014) 'Selective BCL-2 inhibition by ABT-199 causes on-target cell death in acute myeloid leukemia', *Cancer Discov*, 4(3), pp. 362-75.

Parada, L.F., Land, H., Weinberg, R.A., Wolf, D. and Rotter, V. (1984) 'Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation', *Nature*, 312(5995), pp. 649-51.

Parkin, B., Erba, H., Ouillette, P., Roulston, D., Purkayastha, A., Karp, J., Talpaz, M., Kujawski, L., Shakhan, S., Li, C., Shedden, K. and Malek, S.N. (2010) 'Acquired genomic copy number aberrations and survival in adult acute myelogenous leukemia', *Blood*, 116(23), pp. 4958-67.

Pavletich, N.P., Chambers, K.A. and Pabo, C.O. (1993) 'The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots', *Genes Dev*, 7(12B), pp. 2556-64.

Pierceall, W.E., Warner, S.L., Lena, R.J., Doykan, C., Blake, N., Elashoff, M., Hoff, D.V., Bearss, D.J., Cardone, M.H., Andritsos, L., Byrd, J.C., Lanasa, M.C., Grever, M.R. and Johnson, A.J. (2014) 'Mitochondrial priming of chronic lymphocytic leukemia patients associates Bcl-xL dependence with alvocidib response', *Leukemia*, 28(11), pp. 2251-4.

Pohl, G., Rudas, M., Dietze, O., Lax, S., Markis, E., Pirker, R., Zielinski, C.C., Hausmaninger, H., Kubista, E., Samonigg, H., Jakesz, R. and Filipits, M. (2003) 'High p27Kip1 expression predicts superior relapse-free and overall survival for premenopausal women with early-stage breast cancer receiving adjuvant treatment with tamoxifen plus goserelin', *J Clin Oncol*, 21(19), pp. 3594-600.

Porter, P.L., Barlow, W.E., Yeh, I.T., Lin, M.G., Yuan, X.P., Donato, E., Sledge, G.W., Shapiro, C.L., Ingle, J.N., Haskell, C.M., Albain, K.S., Roberts, J.M., Livingston, R.B. and Hayes, D.F. (2006) 'p27(Kip1) and cyclin E expression and breast cancer survival after treatment with adjuvant chemotherapy', *J Natl Cancer Inst*, 98(23), pp. 1723-31.

Pourdehnad, M., Truitt, M.L., Siddiqi, I.N., Ducker, G.S., Shokat, K.M. and Ruggero, D. (2013) 'Myc and mTOR converge on a common node in protein synthesis control that confers synthetic lethality in Myc-driven cancers', *Proc Natl Acad Sci U S A*, 110(29), pp. 11988-93.

Purton, L.E. and Scadden, D.T. (2007) 'Limiting factors in murine hematopoietic stem cell assays', *Cell Stem Cell*, 1(3), pp. 263-70.

Quwailid, M.M., Hugill, A., Dear, N., Vizor, L., Wells, S., Horner, E., Fuller, S., Weedon, J., McMath, H., Woodman, P., Edwards, D., Campbell, D., Rodger, S., Carey, J., Roberts, A., Glenister, P., Lalanne, Z., Parkinson, N., Coghill, E.L., McKeone, R., Cox, S., Willan, J., Greenfield, A., Keays, D., Brady, S., Spurr, N., Gray, I., Hunter, J., Brown, S.D. and Cox, R.D. (2004) 'A gene-driven ENU-based approach to generating an allelic series in any gene', *Mamm Genome*, 15(8), pp. 585-91.

Raimondi, S.C., Chang, M.N., Ravindranath, Y., Behm, F.G., Gresik, M.V., Steuber, C.P., Weinstein, H.J. and Carroll, A.J. (1999) 'Chromosomal abnormalities in 478 children with acute myeloid leukemia: clinical characteristics and treatment outcome in a cooperative pediatric oncology group study-POG 8821', *Blood*, 94(11), pp. 3707-16.

Rambaut, A., Posada, D., Crandall, K.A. and Holmes, E.C. (2004) 'The causes and consequences of HIV evolution', *Nat Rev Genet*, 5(1), pp. 52-61.

Ranzani, M., Annunziato, S., Adams, D.J. and Montini, E. (2013) 'Cancer gene discovery: exploiting insertional mutagenesis', *Mol Cancer Res*, 11(10), pp. 1141-58.

Renneville, A., Roumier, C., Biggio, V., Nibourel, O., Boissel, N., Fenaux, P. and Preudhomme, C. (2008) 'Cooperating gene mutations in acute myeloid leukemia: a review of the literature', *Leukemia*, 22(5), pp. 915-31.

Ried, T., Schrock, E., Ning, Y. and Wienberg, J. (1998) 'Chromosome painting: a useful art', *Hum Mol Genet*, 7(10), pp. 1619-26.

Ripperger, T., Steinemann, D., Gohring, G., Finke, J., Niemeyer, C.M., Strahm, B. and Schlegelberger, B. (2009) 'A novel pedigree with heterozygous germline RUNX1 mutation causing familial MDS-related AML: can these families serve as a multistep model for leukemic transformation?', *Leukemia*, 23(7), pp. 1364-6.

Robinson, H.L. (2002) 'New hope for an AIDS vaccine', *Nat Rev Immunol*, 2(4), pp. 239-50.

Roboz, G.J. (2011) 'Novel approaches to the treatment of acute myeloid leukemia', *Hematology Am Soc Hematol Educ Program*, 2011, pp. 43-50.

Roboz, G.J. and Guzman, M. (2009) 'Acute myeloid leukemia stem cells: seek and destroy', *Expert Rev Hematol*, 2(6), pp. 663-72.

Rommel, C., Clarke, B.A., Zimmermann, S., Nunez, L., Rossman, R., Reid, K., Moelling, K., Yancopoulos, G.D. and Glass, D.J. (1999) 'Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt', *Science*, 286(5445), pp. 1738-41.

Rosati, G., Chiacchio, R., Reggiardo, G., De Sanctis, D. and Manzione, L. (2004) 'Thymidylate synthase expression, p53, bcl-2, Ki-67 and p27 in colorectal cancer: relationships with tumor recurrence and survival', *Tumour Biol*, 25(5-6), pp. 258-63.

Rowley, J.D., Golomb, H.M. and Dougherty, C. (1977) '15/17 translocation, a consistent chromosomal change in acute promyelocytic leukaemia', *Lancet*, 1(8010), pp. 549-50.

Rubin, H. (1955) 'Quantitative relations between causative virus and cell in the Rous no. 1 chicken sarcoma', *Virology*, 1(5), pp. 445-73.

Rucker, F.G., Bullinger, L., Schwaenen, C., Lipka, D.B., Wessendorf, S., Frohling, S., Bentz, M., Miller, S., Scholl, C., Schlenk, R.F., Radlwimmer, B., Kestler, H.A., Pollack, J.R., Lichter, P., Dohner, K. and Dohner, H. (2006) 'Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization', *J Clin Oncol*, 24(24), pp. 3887-94.

Rucker, F.G., Schlenk, R.F., Bullinger, L., Kayser, S., Teleanu, V., Kett, H., Habdank, M., Kugler, C.M., Holzmann, K., Gaidzik, V.I., Paschka, P., Held, G., von Lilienfeld-Toal, M., Lubbert, M., Frohling, S., Zenz, T., Krauter, J., Schlegelberger, B., Ganser, A., Lichter, P., Dohner, K. and Dohner, H. (2012) 'TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome', *Blood*, 119(9), pp. 2114-21.

Russell, W.L., Kelly, E.M., Hunsicker, P.R., Bangham, J.W., Maddux, S.C. and Phipps, E.L. (1979) 'Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse', *Proc Natl Acad Sci U S A*, 76(11), pp. 5818-9.

Sanderson, R.N., Johnson, P.R., Moorman, A.V., Roman, E., Willett, E., Taylor, P.R., Proctor, S.J., Bown, N., Ogston, S. and Bowen, D.T. (2006) 'Population-based demographic study of karyotypes in 1709 patients with adult acute myeloid leukemia', *Leukemia*, 20(3), pp. 444-50.

Sato, Y., Kobayashi, H., Suto, Y., Olney, H.J., Davis, E.M., Super, H.G., Espinosa, R., 3rd, Le Beau, M.M. and Rowley, J.D. (2001) 'Chromosomal instability in chromosome band 12p13: multiple

breaks leading to complex rearrangements including cytogenetically undetectable sub-clones', *Leukemia*, 15(8), pp. 1193-202.

Sato, Y., Suto, Y., Pietenpol, J., Golub, T.R., Gilliland, D.G., Davis, E.M., Le Beau, M.M., Roberts, J.M., Vogelstein, B., Rowley, J.D. and et al. (1995) 'TEL and KIP1 define the smallest region of deletions on 12p13 in hematopoietic malignancies', *Blood*, 86(4), pp. 1525-33.

Schaich, M., Parmentier, S., Kramer, M., Illmer, T., Stolzel, F., Rollig, C., Thiede, C., Hanel, M., Schafer-Eckart, K., Aulitzky, W., Einsele, H., Ho, A.D., Serve, H., Berdel, W.E., Mayer, J., Schmitz, N., Krause, S.W., Neubauer, A., Baldus, C.D., Schetelig, J., Bornhauser, M. and Ehninger, G. (2013) 'High-dose cytarabine consolidation with or without additional amsacrine and mitoxantrone in acute myeloid leukemia: results of the prospective randomized AML2003 trial', *J Clin Oncol*, 31(17), pp. 2094-102.

Schmitt, C.A., Fridman, J.S., Yang, M., Baranov, E., Hoffman, R.M. and Lowe, S.W. (2002) 'Dissecting p53 tumor suppressor functions in vivo', *Cancer Cell*, 1(3), pp. 289-98.

Schneider, V., Zhang, L., Bullinger, L., Rojewski, M., Hofmann, S., Wiesneth, M., Schrezenmeier, H., Gotz, M., Botzenhardt, U., Barth, T.F., Dohner, K., Dohner, H. and Greiner, J. (2014) 'Leukemic stem cells of acute myeloid leukemia patients carrying NPM1 mutation are candidates for targeted immunotherapy', *Leukemia*, 28(8), pp. 1759-62.

Schoch, C., Dicker, F., Herholz, H., Schnittger, S., Kern, W. and Haferlach, T. (2006) 'Mutations of the TP53 gene occur in 13.4% of acute myeloid leukemia and are strongly associated with a complex aberrant karyotype.', *Blood*, 108(11), pp. 542a-542a.

Schoch, C., Haferlach, T., Bursch, S., Gerstner, D., Schnittger, S., Dugas, M., Kern, W., Loffler, H. and Hiddemann, W. (2002) 'Loss of genetic material is more common than gain in acute myeloid leukemia with complex aberrant karyotype: a detailed analysis of 125 cases using conventional chromosome analysis and fluorescence in situ hybridization including 24-color FISH', *Genes Chromosomes Cancer*, 35(1), pp. 20-9.

Schoch, C., Kohlmann, A., Dugas, M., Kern, W., Hiddemann, W., Schnittger, S. and Haferlach, T. (2005) 'Genomic gains and losses influence expression levels of genes located within the affected regions: a study on acute myeloid leukemias with trisomy 8, 11, or 13, monosomy 7, or deletion 5q', *Leukemia*, 19(7), pp. 1224-8.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F. and Pals, G. (2002) 'Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification', *Nucleic Acids Res*, 30(12), p. e57.

Seabright, M. (1971) 'A rapid banding technique for human chromosomes', *Lancet*, 2(7731), pp. 971-2.

Seifert, H., Mohr, B., Thiede, C., Oelschlagel, U., Schakel, U., Illmer, T., Soucek, S., Ehninger, G., Schaich, M. and Study Alliance, L. (2009) 'The prognostic impact of 17p (p53) deletion in 2272 adults with acute myeloid leukemia', *Leukemia*, 23(4), pp. 656-63.

Seita, J. and Weissman, I.L. (2010) 'Hematopoietic stem cell: self-renewal versus differentiation', *Wiley Interdiscip Rev Syst Biol Med*, 2(6), pp. 640-53.

Sellner, L.N. and Taylor, G.R. (2004) 'MLPA and MAPH: new techniques for detection of gene deletions', *Hum Mutat*, 23(5), pp. 413-9.

Senderowicz, A.M. (1999) 'Flavopiridol: the first cyclin-dependent kinase inhibitor in human clinical trials', *Invest New Drugs*, 17(3), pp. 313-20.

Seymour, J.F., Pierce, S.A., Kantarjian, H.M., Keating, M.J. and Estey, E.H. (1994) 'Investigation of karyotypic, morphologic and clinical features in patients with acute myeloid leukemia blast cells expressing the neural cell adhesion molecule (CD56)', *Leukemia*, 8(5), pp. 823-6.

Shapiro, G.I. (2004) 'Preclinical and clinical development of the cyclin-dependent kinase inhibitor flavopiridol', *Clin Cancer Res*, 10(12 Pt 2), pp. 4270s-4275s.

Sheiness, D., Fanshier, L. and Bishop, J.M. (1978) 'Identification of nucleotide sequences which may encode the oncogenic capacity of avian retrovirus MC29', *J Virol*, 28(2), pp. 600-10.

Shi, J., Shao, Z.H., Liu, H., Bai, J., Cao, Y.R., He, G.S., Tu, M.F., Wang, X.L., Hao, Y.S., Yang, T.Y. and Yang, C.L. (2004) 'Transformation of myelodysplastic syndromes into acute myeloid leukemias', *Chin Med J (Engl)*, 117(7), pp. 963-7.

Shih, T.Y., Papageorge, A.G., Stokes, P.E., Weeks, M.O. and Scolnick, E.M. (1980) 'Guanine nucleotide-binding and autophosphorylating activities associated with the p21src protein of Harvey murine sarcoma virus', *Nature*, 287(5784), pp. 686-91.

Shimizu, K., Goldfarb, M., Suard, Y., Perucho, M., Li, Y., Kamata, T., Feramisco, J., Stavnezer, E., Fogh, J. and Wigler, M.H. (1983) 'Three human transforming genes are related to the viral ras oncogenes', *Proc Natl Acad Sci U S A*, 80(8), pp. 2112-6.

Shultz, L.D., Lyons, B.L., Burzenski, L.M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S.D., King, M., Mangada, J., Greiner, D.L. and Handgretinger, R. (2005) 'Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells', *J Immunol*, 174(10), pp. 6477-89.

Shultz, L.D., Schweitzer, P.A., Christianson, S.W., Gott, B., Schweitzer, I.B., Tennent, B., McKenna, S., Mobraaten, L., Rajan, T.V., Greiner, D.L. and et al. (1995) 'Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice', *J Immunol*, 154(1), pp. 180-91.

Silva, F.P., Morolli, B., Storlazzi, C.T., Zagaria, A., Impera, L., Klein, B., Vrieling, H., Kluin-Nelemans, H.C. and Giphart-Gassler, M. (2008) 'ETV6 mutations and loss in AML-M0', *Leukemia*, 22(8), pp. 1639-43.

Simon, C., Chagraoui, J., Krosl, J., Gendron, P., Wilhelm, B., Lemieux, S., Boucher, G., Chagnon, P., Drouin, S., Lambert, R., Rondeau, C., Bilodeau, A., Lavallee, S., Sauvageau, M., Hebert, J. and Sauvageau, G. (2012) 'A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia', *Genes Dev*, 26(7), pp. 651-6.

Simon-Loriere, E., Rossolillo, P. and Negroni, M. (2011) 'RNA structures, genomic organization and selection of recombinant HIV', *RNA Biol*, 8(2), pp. 280-6.

Slovak, M.L., Ho, J.P., Pettenati, M.J., Khan, A., Douer, D., Lal, S. and Traweek, S.T. (1994) 'Localization of amplified MYC gene sequences to double minute chromosomes in acute myelogenous leukemia', *Genes Chromosomes Cancer*, 9(1), pp. 62-7.

Slovak, M.L., Kopecky, K.J., Cassileth, P.A., Harrington, D.H., Theil, K.S., Mohamed, A., Paietta, E., Willman, C.L., Head, D.R., Rowe, J.M., Forman, S.J. and Appelbaum, F.R. (2000) 'Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study', *Blood*, 96(13), pp. 4075-83.

Smith, S.M., Le Beau, M.M., Huo, D., Karrison, T., Sobecks, R.M., Anastasi, J., Vardiman, J.W., Rowley, J.D. and Larson, R.A. (2003) 'Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series', *Blood*, 102(1), pp. 43-52.

Spataro, V.J., Litman, H., Viale, G., Maffini, F., Masullo, M., Golouh, R., Martinez-Tello, F.J., Grigolato, P., Shilkin, K.B., Gusterson, B.A., Castiglione-Gertsch, M., Price, K., Lindtner, J., Cortes-Funes, H., Simoncini, E., Byrne, M.J., Collins, J., Gelber, R.D., Coates, A.S., Goldhirsch, A. and International Breast Cancer Study, G. (2003) 'Decreased immunoreactivity for p27 protein in patients with early-stage breast carcinoma is correlated with HER-2/neu overexpression and with benefit from one course of perioperative chemotherapy in patients with negative lymph node status: results from International Breast Cancer Study Group Trial V', *Cancer*, 97(7), pp. 1591-600.

Stalfelt, A.M., Brodin, H., Pettersson, S. and Eklof, A. (2001) 'The final phase in acute myeloid leukaemia (AML): a study of cause of death, place of death and type of care during the last week of life', *Leuk Res*, 25(8), pp. 673-80.

Stanford, W.L., Cohn, J.B. and Cordes, S.P. (2001) 'Gene-trap mutagenesis: past, present and beyond', *Nat Rev Genet*, 2(10), pp. 756-68.

Stehelin, D., Varmus, H.E., Bishop, J.M. and Vogt, P.K. (1976) 'DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA', *Nature*, 260(5547), pp. 170-3.

Stephens, D.M., Ruppert, A.S., Maddocks, K., Andritsos, L., Baiocchi, R., Jones, J., Johnson, A.J., Smith, L.L., Zhao, Y., Ling, Y., Li, J., Phelps, M.A., Grever, M.R., Byrd, J.C. and Flynn, J.M. (2013) 'Cyclophosphamide, alvocidib (flavopiridol), and rituximab, a novel feasible chemoimmunotherapy regimen for patients with high-risk chronic lymphocytic leukemia', *Leuk Res*, 37(10), pp. 1195-9.

Strachan, T. and Read, A.P. (1999) in *Human Molecular Genetics*. 2nd edn. New York.

Stratton, M.R., Campbell, P.J. and Futreal, P.A. (2009) 'The cancer genome', *Nature*, 458(7239), pp. 719-24.

Streubel, B., Valent, P., Jager, U., Edelhauser, M., Wandt, H., Wagner, T., Buchner, T., Lechner, K. and Fonatsch, C. (2000) 'Amplification of the MLL gene on double minutes, a homogeneously staining region, and ring chromosomes in five patients with acute myeloid leukemia or myelodysplastic syndrome', *Genes Chromosomes Cancer*, 27(4), pp. 380-6.

Takenaka, K., Prasolava, T.K., Wang, J.C., Mortin-Toth, S.M., Khalouei, S., Gan, O.I., Dick, J.E. and Danska, J.S. (2007) 'Polymorphism in Sirpa modulates engraftment of human hematopoietic stem cells', *Nat Immunol*, 8(12), pp. 1313-23.

Tallman, M.S., Gilliland, D.G. and Rowe, J.M. (2005) 'Drug therapy for acute myeloid leukemia', *Blood*, 106(4), pp. 1154-63.

Tansey, W.P. (2014) 'Mammalian MYC Proteins and Cancer', *New Journal of Science*.

Temin, H.M. (1960) 'The control of cellular morphology in embryonic cells infected with rous sarcoma virus in vitro', *Virology*, 10, pp. 182-97.

Temin, H.M. and Rubin, H. (1958) 'Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture', *Virology*, 6(3), pp. 669-88.

Theocharides, A.P., Jin, L., Cheng, P.Y., Prasolava, T.K., Malko, A.V., Ho, J.M., Poeppl, A.G., van Rooijen, N., Minden, M.D., Danska, J.S., Dick, J.E. and Wang, J.C. (2012) 'Disruption of SIRPalpha signaling in macrophages eliminates human acute myeloid leukemia stem cells in xenografts', *J Exp Med*, 209(10), pp. 1883-99.

Thomas, D., Powell, J.A., Vergez, F., Segal, D.H., Nguyen, N.Y., Baker, A., The, T.C., Barry, E.F., Sarry, J.E., Lee, E.M., Nero, T.L., Jabbour, A.M., Pomilio, G., Green, B.D., Manenti, S., Glaser, S.P., Parker, M.W., Lopez, A.F., Ekert, P.G., Lock, R.B., Huang, D.C., Nilsson, S.K., Récher, C., Wei, A.H. and Guthridge, M.A. (2013) 'Targeting acute myeloid leukemia by dual inhibition of PI3K signalling and Cdk9-mediated Mcl-1 transcription', *Blood*, 122(5): pp. 738-48

Toyoshima, K., Friis, R.R. and Vogt, P.K. (1970) 'The reproductive and cell-transforming capacities of avian sarcoma virus B77: inactivation with UV light', *Virology*, 42(1), pp. 163-70.

Travis, L.B., Li, C.Y., Zhang, Z.N., Li, D.G., Yin, S.N., Chow, W.H., Li, G.L., Dosemeci, M., Blot, W., Fraumeni, J.F., Jr. and et al. (1994) 'Hematopoietic malignancies and related disorders among benzene-exposed workers in China', *Leuk Lymphoma*, 14(1-2), pp. 91-102.

Tsihlias, J., Kapusta, L. and Slingerland, J. (1999) 'The prognostic significance of altered cyclin-dependent kinase inhibitors in human cancer', *Annu Rev Med*, 50, pp. 401-23.

Tsujimoto, Y., Finger, L.R., Yunis, J., Nowell, P.C. and Croce, C.M. (1984) 'Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation', *Science*, 226(4678), pp. 1097-9.

Tuna, M., Chavez-Reyes, A. and Tari, A.M. (2005) 'HER2/neu increases the expression of Wilms' Tumor 1 (WT1) protein to stimulate S-phase proliferation and inhibit apoptosis in breast cancer cells', *Oncogene*, 24(9), pp. 1648-52.

Uren, A.G., Kool, J., Berns, A. and van Lohuizen, M. (2005) 'Retroviral insertional mutagenesis: past, present and future', *Oncogene*, 24(52), pp. 7656-72.

Uren, A.G., Kool, J., Matentzoglu, K., de Ridder, J., Mattison, J., van Uitert, M., Lagcher, W., Sie, D., Tanger, E., Cox, T., Reinders, M., Hubbard, T.J., Rogers, J., Jonkers, J., Wessels, L., Adams, D.J., van Lohuizen, M. and Berns, A. (2008) 'Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks', *Cell*, 133(4), pp. 727-41.

Vafa, O., Wade, M., Kern, S., Beeche, M., Pandita, T.K., Hampton, G.M. and Wahl, G.M. (2002) 'c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability', *Mol Cell*, 9(5), pp. 1031-44.

Vardiman, J.W., Harris, N.L. and Brunning, R.D. (2002) 'The World Health Organization (WHO) classification of the myeloid neoplasms', *Blood*, 100(7), pp. 2292-302.

Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellstrom-Lindberg, E., Tefferi, A. and Bloomfield, C.D. (2009) 'The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes', *Blood*, 114(5), pp. 937-51.

Venditti, A., Del Poeta, G., Maurillo, L., Buccisano, F., Del Principe, M.I., Mazzone, C., Tamburini, A., Cox, C., Panetta, P., Neri, B., Ottaviani, L. and Amadori, S. (2004) 'Combined analysis of bcl-2 and MDR1 proteins in 256 cases of acute myeloid leukemia', *Haematologica*, 89(8), pp. 934-9.

Vennstrom, B., Sheiness, D., Zabielski, J. and Bishop, J.M. (1982) 'Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29', *J Virol*, 42(3), pp. 773-9.

Wall, M., Rayeroux, K.C., MacKinnon, R.N., Zordan, A. and Campbell, L.J. (2012) 'ETV6 deletion is a common additional abnormality in patients with myelodysplastic syndromes or acute myeloid leukemia and monosomy 7', *Haematologica*, 97(12), pp. 1933-6.

Walter, R.B., Appelbaum, F.R., Estey, E.H. and Bernstein, I.D. (2012) 'Acute myeloid leukemia stem cells and CD33-targeted immunotherapy', *Blood*, 119(26), pp. 6198-208.

Wang, J., Sun, Q., Morita, Y., Jiang, H., Gross, A., Lechel, A., Hildner, K., Guachalla, L.M., Gompf, A., Hartmann, D., Schambach, A., Wuestefeld, T., Dauch, D., Schrezenmeier, H., Hofmann, W.K., Nakauchi, H., Ju, Z., Kestler, H.A., Zender, L. and Rudolph, K.L. (2012) 'A differentiation checkpoint limits hematopoietic stem cell self-renewal in response to DNA damage', *Cell*, 148(5), pp. 1001-14.

Wang, K. and Bucan, M. (2008) 'Copy Number Variation Detection via High-Density SNP Genotyping', *CSH Protoc*, 2008, p. pdb top46.

Wang, L.H., Duesberg, P.H., Kawai, S. and Hanafusa, H. (1976) 'Location of envelope-specific and sarcoma-specific oligonucleotides on RNA of Schmidt-Ruppin Rous sarcoma virus', *Proc Natl Acad Sci U S A*, 73(2), pp. 447-51.

Wang, X., Krupczak-Hollis, K., Tan, Y., Dennewitz, M.B., Adami, G.R. and Costa, R.H. (2002) 'Increased hepatic Forkhead Box M1B (FoxM1B) levels in old-aged mice stimulated liver regeneration through diminished p27Kip1 protein levels and increased Cdc25B expression', *J Biol Chem*, 277(46), pp. 44310-6.

Watson, M., Buck, G., Wheatley, K., Homewood, J.R., Goldstone, A.H., Rees, J.K., Burnett, A.K. and trial, U.K.M.R.C.A. (2004) 'Adverse impact of bone marrow transplantation on quality of life in acute myeloid leukaemia patients; analysis of the UK Medical Research Council AML 10 Trial', *Eur J Cancer*, 40(7), pp. 971-8.

Weiss, M.M., Hermsen, M.A., Meijer, G.A., van Grieken, N.C., Baak, J.P., Kuipers, E.J. and van Diest, P.J. (1999) 'Comparative genomic hybridisation', *Mol Pathol*, 52(5), pp. 243-51.

Willinger, T., Rongvaux, A., Strowig, T., Manz, M.G. and Flavell, R.A. (2011) 'Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement', *Trends Immunol*, 32(7), pp. 321-7.

Wlodarska, I., Marynen, P., La Starza, R., Mecucci, C. and Van den Berghe, H. (1996) 'The ETV6, CDKN1B and D12S178 loci are involved in a segment commonly deleted in various 12p aberration in different hematological malignancies', *Cytogenet Cell Genet*, 72(2-3), pp. 229-35.

Wong, R., Shahjahan, M., Wang, X., Thall, P.F., De Lima, M., Khouri, I., Gajewski, J., Alamo, J., Couriel, D., Andersson, B.S., Donato, M., Hosing, C., Komanduri, K., Anderlini, P., Molldrem, J., Ueno, N.T., Estey, E., Ippoliti, C., Champlin, R. and Giralt, S. (2005) 'Prognostic factors for outcomes of patients with refractory or relapsed acute myelogenous leukemia or myelodysplastic syndromes undergoing allogeneic progenitor cell transplantation', *Biol Blood Marrow Transplant*, 11(2), pp. 108-14.

Wu, F.Y., Wang, S.E., Sanders, M.E., Shin, I., Rojo, F., Baselga, J. and Arteaga, C.L. (2006) 'Reduction of cytosolic p27(Kip1) inhibits cancer cell motility, survival, and tumorigenicity', *Cancer Res*, 66(4), pp. 2162-72.

Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M.D., Dotsch, V., Andrews, N.C., Caput, D. and McKeon, F. (1998) 'p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities', *Mol Cell*, 2(3), pp. 305-16.

Yang, H., Zhao, R., Yang, H.Y. and Lee, M.H. (2005) 'Constitutively active FOXO4 inhibits Akt activity, regulates p27 Kip1 stability, and suppresses HER2-mediated tumorigenicity', *Oncogene*, 24(11), pp. 1924-35.

Yang, L., Han, Y., Suarez Saiz, F. and Minden, M.D. (2007) 'A tumor suppressor and oncogene: the WT1 story', *Leukemia*, 21(5), pp. 868-76.

Yang, W., Shen, J., Wu, M., Arsura, M., FitzGerald, M., Suldan, Z., Kim, D.W., Hofmann, C.S., Pianetti, S., Romieu-Mourez, R., Freedman, L.P. and Sonenshein, G.E. (2001) 'Repression of

transcription of the p27(Kip1) cyclin-dependent kinase inhibitor gene by c-Myc', *Oncogene*, 20(14), pp. 1688-702.

Yang, Y., Hou, J.Q., Qu, L.Y., Wang, G.Q., Ju, H.W., Zhao, Z.W., Yu, Z.H. and Yang, H.J. (2007) '[Differential expression of USP2, USP14 and UBE4A between ovarian serous cystadenocarcinoma and adjacent normal tissues]', *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*, 23(6), pp. 504-6.

Ylstra, B., van den Ijssel, P., Carvalho, B., Brakenhoff, R.H. and Meijer, G.A. (2006) 'BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH)', *Nucleic Acids Res*, 34(2), pp. 445-50.

Youle, R.J. and Strasser, A. (2008) 'The BCL-2 protein family: opposing activities that mediate cell death', *Nat Rev Mol Cell Biol*, 9(1), pp. 47-59.

Youn, B.S., Mantel, C. and Broxmeyer, H.E. (2000) 'Chemokines, chemokine receptors and hematopoiesis', *Immunol Rev*, 177, pp. 150-74.

Yunis, J.J. (1976) 'High resolution of human chromosomes', *Science*, 191(4233), pp. 1268-70.

Zatkova, A., Merk, S., Wendehack, M., Bilban, M., Muzik, E.M., Muradyan, A., Haferlach, C., Haferlach, T., Wimmer, K., Fonatsch, C. and Ullmann, R. (2009) 'AML/MDS with 11q/MLL amplification show characteristic gene expression signature and interplay of DNA copy number changes', *Genes Chromosomes Cancer*, 48(6), pp. 510-20.

Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., McCombie, W.R., Wigler, M., Hicks, J., Hannon, G.J., Powers, S. and Lowe, S.W. (2008) 'An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer', *Cell*, 135(5), pp. 852-64.

Zhang, Y., Wong, J., Klinger, M., Tran, M.T., Shannon, K.M. and Killeen, N. (2009) 'MLL5 contributes to hematopoietic stem cell fitness and homeostasis', *Blood*, 113(7), pp. 1455-63.

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L. and Trono, D. (1998) 'Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery', *J Virol*, 72(12), pp. 9873-80.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L. and Trono, D. (1997) 'Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo', *Nat Biotechnol*, 15(9), pp. 871-5.