Genetic and Functional Characterisation of *Immunoglobulin Heavy Chain Locus-CCAAT Enhancer Binding Protein* B-Cell Acute Lymphoblastic Leukaemia

A thesis submitted in part requirement for the degree of

Doctor of Philosophy from the Faculty of Medical Sciences

Dino Mašić

Leukaemia Research Cytogenetic Group

Northern Institute for Cancer Research

Faculty of Medical Sciences

Newcastle University

Newcastle-Upon-Tyne

February 2016

Abstract

B-cell precursor acute lymphoblastic leukaemia (BCP-ALL) is a heterogeneous disease in which patient outcome is influenced by genetic lesions. This outcome has improved due to increasingly tailored treatment regimens, selected through risk stratification by use of cytogenetic, copy number alterations (CNAs) classifiers, genomic, and molecular data.

Translocations involving the Immunoglobulin Heavy Chain Locus (*IGH*) comprise 5% of BCP-ALL and lead to overexpression of juxtaposed genes, due to the powerful *IGH* enhancer elements. Multiple *IGH* partner genes have been described in BCP-ALL, including five members of the Ccaat Enhancer-Binding Protein (*CEBP*) transcription factor family. A cohort of 33 *IGH-CEBP* BCP-ALL patients was identified including 11 *IGH-CEBPD*, 10 *IGH-CEBPA*, 8 *IGH-CEBPB*, 3 *IGH-CEBPE* and 1 *IGH-CEBPG* patients, comprising 19% of the *IGH* cohort, and 1% of ALL as a whole. The patients displayed variation between individual *CEBP* subgroups, with *IGH-CEBPB* patients showing higher white blood cell counts (WBC), higher relapse rates, higher number of CNAs and older age than other *CEBP* patients. The *CEBPD* subgroup included mostly younger patients, under the age of 10 years, and had the lowest number of CNAs per patient. Deletions of *CDKN2A*/*B* were the most commonly occurring CNA followed by intragenic exon 4-7 deletions of *IKZF1,* which were found exclusively in the *IGH-CEBPB* and *IGH-CEBPD* subgroups (p=0.04). A novel intragenic deletion of the tyrosine kinase gene, *ABL2,* was found in four patients in the cohort, which may represent a deletion unique to this subgroup. This finding in combination with the *IKZF1* deletions is suggestive of a *BCR-ABL1*-like profile.

Retroviral expression of the *CEBPD* gene in CD34+ cells was found to hinder proliferation in transduced cells, potentially through cell cycle arrest via the *RB*/*E2F* pathway. RNA sequencing analysis of two *IGH-CEBP* patients showed very different expression profiles, suggesting two mechanisms of oncogenesis in *IGH-CEBP* patients: one through inactivation of the *CEBP* function, leading to deregulation of cell cycle and differentiation control, and another through upregulation of inflammatory factors.

Acknowledgements

Creating this body of work has been incredibly challenging process, I could not have done it were it not for a huge amount of support from my family and friends. I will try to convey my thanks below, although it will hardly be sufficient.

First of all I want to thank my parents Sead and Marijana, for their unwavering moral support throughout these last four (and a bit) years. They have both set an amazing example for me and my sister, through their hard work and sacrifice. They are my heroes for a hundred more reasons which I will not list here as I feel my word count is already pretty high! Next I want to thank the first Dr. Mašić in the family, my little sister Una. Seeing what she has achieved has both made me incredibly proud, and has inspired me greatly. She has cheered me up and kept me going through the more difficult periods of this PhD. Ovo je za vas, volim vas najviše na svetu!

I want to thank all of my work friends that have made the last few years so worthwhile. I want to thank Lisa Jones for being a fellow founding member of 'Team Dino', we've been through a lot together, from malfunctioning microscopes to karaoke duets, it's been great. I want to thank Dr. Dan Coleman (we started from the bottom now we're here), my fellow day one starter buddy Dr. Jake Clayton, and Matt Selby for all of the tough love (hate). Thanks to Claire Schwab for lots of good advice, artisan sausage rolls, and Glastonbury adventures. Stacey Richardson for always being there when I needed a whine, James Murray for allowing me to vent about Newcastle United, and Ruth Cranston for somehow manging to be a PhD student and a rock star at the same time. Many thanks to Katie Dormon and George Diggory, for feeding me once week and allowing me to ignore the stresses of thesis writing through D&D!

Huge thanks also to the end office gang, Dr. Amir Enshaei for a massive amount of moral support and for attempting to teach me R and Stata! Dr. Debbie Hicks for all the enthusiastic waving in the halls of the SJS, Dr. Al Gabriel on knowledge for both biology AND hunting, Dr. Matt Bashton and Dr. Reza Raifee.

I want to thank Shanel Swartz for always letting me into the building every other weekend, and for being unfailingly cheerful. Double doctor Becky Hill for

keeping me sane in the last few weeks before this hand in, Dr. Stefano Tonin a new member of ‘Team Dino’, who will no doubt continue my legacy, and Dr. Dan Williamson for all the Big John Carew references.

I want to thank Alannah Elliott, Dr. Lucy Chilton, Dr. Janet Lindsey, Prof. Anthony Moorman, Dr. Sarra Ryan, Dr. Paul Sinclair, Dr. Jeyanthy Eswaran, Dr. Sirintra Nakjang, Dr. Tim Barrow, Dr. Carmela Ciardullo, Yuri Gabrosky, Shawn Hollern, Fang Ju, Azira Ramli, and Joanna Horabin. Thank you for all of your friendship and support (both lab and pub based).

I had a wonderful time in Cincinnati because of the people I met and worked with there. I would genuinely like to thank Dr. James Mulloy for kindly allowing me to work as part of his group for six months, and for all of the guidance and advice throughout my time there, I’ve been working on being more optimistic Jim! I want to thank Christina Sexton, Eric O’Brien, Mark Wunderlich, Shan Lin, Susumu Goyama and Mahesh Shrestha. I also want to thank Rushi Panchal for giving me a place to stay and being a great friend, and Sam Taylor for showing that you can be a respectable scientist whilst wearing a stars and stripes bandana.

Last but not least I want to thank my supervisors. Professor Christine Harrison for all of your advice regarding both my scientific endeavours and my terrible grammar! And to Dr Lisa Russell for giving me the chance to work with her for the last four years, which have been incredible. I would like to genuinely thank you for all of your support both in the lab and out of it, and all of the opportunities you gave me. I aim to emulate the hard work, high standards and genuine support you have provided in my future career. Thank you so much for everything, I’m hugely grateful. Bet you can’t wait to see the back of me!

One more time, I’ve written it down 19 times already (I checked) but thank you each and every one for all of your support. I wouldn’t be here without all of your help and advice, standing on the shoulders of giants.

Statement of Work Undertaken

The retroviral vectors used in this project were kindly gifted by Dr. James Mulloy and Dr. Eric Clambey. The MiT-IK6 construct was produced by Dr. Eric Clambey.

Support was provided by the staff of the Comparative Biology Centre for the maintenance of NSG mice. Intra-femoral injections were performed by Dr. Helen Blair. Mouse health was tracked on a regular basis by Dr. Helen Blair, Dr. Paul Sinclair, Jake Clayton and Dr. Stefano Tonin, in addition to myself. In Cincinnati mouse injections and flow sorting were performed by Mark Wunderlich. Figure 4.25 was created from Mr. Wunderlich's work. Mouse health was tracked by Mark Wunderlich, Christina Sexton and Eric O'Brien.

FISH slide analysis was performed by myself with checks by Dr. Lisa Russell, Lisa Jones, and Dr. Amy Erhorn.

Support was provided by Dr. David McDonald and Dr. Andrew Fuller for FACS optimisation and FACS sorting of CD34+ cells in Newcastle University.

Figure 3.3 was created by Illumina staff Fiona Nielsen, Jennifer Becq and Adrian Alexa. Images from figure 4.27 were created from immunohistochemistry analysis of mouse skulls, this work was performed by Dr. Chris Halsey and her team at the University of Glasgow.

All other work and data analysis was performed by myself.

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables

Abbreviations

AML	Acute Myeloid Leukaemia
BAC	Bacterial Artificial Chromosome
BM	Bone Marrow
CGH	Comparative Genomic Hybridisation
CLL	Chronic lymphocytic Leukaemia
CML	Chronic Myeloid Leukaemia
CMP	Committed Myeloid Progenitor
CNA	Copy Number Alterations
CNS	Central Nervous System
DNA	Deoxyribonucleic Acid
ds	Double Stranded
EGFP	Enhanced Green Fluorescent Protein
ETS	Erythroblast Transformation Specific
FAB	French-American-British
FACS	Fluorescence Activated Cell Sorting
FCS	Foetal Calf Serum
FISH	Fluorescence in situ hybridisation
G-banding	Giemsa Banding
GEO	Gene Expression Omnibus
GM	Geometric Mean
HES	Hydroxyethyl starch
HSC	Haematopoietic Stem Cell
IRES	Internal Ribosome Entry Site
ITD	Internal Tandem Duplication
LOH	Loss of Heterozygosity
LPS	Lipopolysaccharides
LRCG	Leukaemia Research Cytogenetics Group
LSC	Leukaemic Stem Cell
LTR	Long Terminal Repeat
MLP	Immature Lymphoid Progenitor

MLPA	Multiplex Ligation Dependent Probe Amplification
MLV	Murine Leukaemia Virus
MPP	Multipotent Progenitor
MSCV	Murine Stem Cell Virus
MRD	Minimal Residual Disease
NK	Natural Killer
NHEJ	Non Homologous End Joining
NOD	Non Obese Diabetic
NSG	NOD Scid Gamma
PB	Peripheral Blood
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
ROS	Reactive Oxygen Species
RSS	Recombination Signal Sequences
Scid	Severe Combined Immune Deficient
SH	Src Homology Domain
SNP	Single Nucleotide Polymorphism
ss	Single Stranded
SUMO	Small Ubiquitin-related Modifier
TAMs	Tumour Activated Macrophages
TD	Transactivating Domain
TE	Tris-EDTA
TKI	Tyrosine Kinase Inhibitor
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization

Chapter 1 Introduction

Cancer is a highly heterogeneous disease triggered by deregulation of controlled cellular processes. Classified by initiating cell of origin, the disease is propagated by genetic aberrations, which modify the function of key genetic pathways involved in kinase/cytokine signalling, cell cycle control, and cell death. The number of genetic lesions necessary to initiate tumourigenesis is highly variable and heavily dependent upon the cancer type and the pathway affected.

Leukaemias are cancers of the haematopoietic system, which account for 3% of all cancers (*Cancer Research Worldwide Cancer Stats* 2011). They are characterised by recurring cytogenetic rearrangements and genetic lesions. Acute lymphoblastic leukaemia (ALL) is the second most common haematological malignancy. It is the most common cancer in children between the ages of 1 to 4 years, and the second most prevalent in children under 12 months. The incidence of paediatric ALL in Britain is between 35-40 children for every 10^6 per year and accounts for 8% of all leukaemias in the UK (*Cancer Research Worldwide Cancer Stats* 2011) (Parkin *et al.*, 1988). Incidence generally decreases with age. The number of patients has increased at roughly 1% per year in affluent societies (Draper *et al.*, 1994). ALL is comprised of two disease subtypes; B-cell precursor ALL (BCP-ALL) occurring in ~85% and T-lineage ALL (T-ALL) making up the remaining 15% (Chiaretti and Foa, 2009).

1.1 Haematopoiesis

Leukaemic development is the result of aberrant haematopoiesis, which occurs in the bone marrow. Haematopoietic stem cells (HSCs) undergo sequential rounds of differentiation and fate restriction to give rise to all haematopoietic lineages (Figure 1.1). Haematopoietic stem cell fate is dependent upon the interplay between the master regulator genes: *PU.1* and *GATA1*. Expression of *GATA1* leads to *PU.1* inhibition and commitment to the erythroid lineage, while *PU.1* expression results in lymphoid or myeloid commitment depending upon expression levels of the gene, with lower expression leading to production of lymphoid progenitors and higher expression to myeloid progenitors (Nerlov and Graf, 1998).

Further cell maturation is controlled by several master regulator genes, which include *CEBPA*, *CEBPB*, *CEBPE*, *GFI1*, *IRF8*, *RUNX1*, *SCL*, *JUNB*, *IKZF1* and *MYC*. These genes activate multiple downstream targets, leading to activation of diverse functions in the cell, some of the most important being the activation/deactivation of lineage specific genes and the downregulation of self-renewal genes. These genes exert control on a myriad of downstream targets, including expression of receptor genes *G-CSF* and *M-CSF*, and chromatin remodelling genes *SWI*/*SNF*, which prime sections of the chromosome for transcription (Rosenbauer and Tenen, 2007). Initial B-lymphoid commitment is dictated by *IKZF1*, *PU.1* and *TCF3*. After the first round of differentiation directed by *PU.1* and *GATA1, IKZF1* recruits *SWI*/*SNF* chromatin remodelling genes, exposing focal sections of chromosomes in preparation for B-cell proliferation. *TCF3* and *IKZF1* continue the priming process through expression of multiple lymphoid lineage genes including *EBF1*, while supressing HSC and other lineage genes (Mansson *et al.*, 2007; Dias *et al.*, 2008). *TCF3* maintains the expression of *EBF1* throughout the differentiation process, leading to maturation into common lymphoid progenitors (CLPs), and is required for the initiation and maintenance of *PAX5* and *MB1* expression in later developmental stages (Kwon *et al.*, 2008) (Sigvardsson *et al.*, 2002). *PAX5* then maintains B lymphoid identity through activation and deactivation of further gene sets (Cobaleda *et al.*, 2007). As the cell begins to move toward maturation, V(D)J recombination is initiated, where *PAX5* (Fuxa *et al.*, 2004) and *IKZF1* both function to expose the participating chromosomal regions containing the antibody chain sequences (Thompson *et al.*, 2007). *PAX5* also directly activates genes coding for important components of pre-BCR signalling (Nutt *et al.*, 1997), while *IKZF1* modulates the expression of early B-cell specific genes, such as *IGLL1,* which code for the pre-B receptor. After rearrangement of the V(D)J segments and successful BCR signalling, the naive B-cells then migrate out of the bone marrow to circulate throughout the body in a dormant state until binding with a novel antigen, which initiates differentiation into memory B-cells and plasma cells.

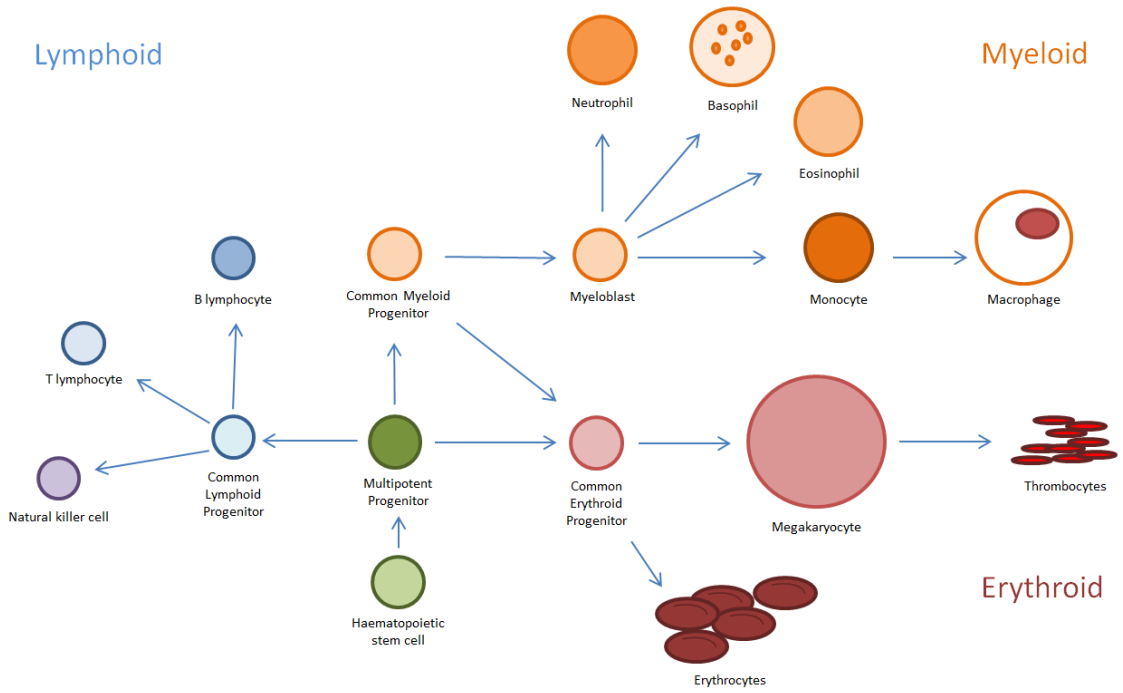


Figure 1.1. Diagram of the three main haematopoietic lineages in humans; erythroid, myeloid and lymphoid.

Haematopoietic stem cells (HSCs) differentiate into multipotent progenitors (MPPs) which in turn branch into common lymphoid progenitors (CLPs) and terminally differentiate into B-cells, T-cells and natural killer (NK) cells. Common myeloid progenitors (CMPs) give rise to both granulocytes, and erythroid progenitors which can also be propagated by MPPs. CMPs differentiate into myeloblasts, which differentiate into a host of cells; neutrophils, basophils, eosinophils and macrophages, forming the bulk of the innate immune system. MPPs and CMPs also differentiate into erythrocytes, megakaryocytes and thrombocytes.

1.1.1 V(D)J Rearrangement

V(D)J rearrangements are the hallmark of B-cell maturation. The process is a complex multi-step method creating between 10^6-10^7 variable antibody molecules. These are subsequently secreted by plasma cells, acting as free roaming markers of antigens, or as membrane bound immunoglobulins on activated memory B-cells.

Antibodies consist of two immunoglobulin heavy chains (IgH) and two identical immunoglobulin light chains (IgL), which are either lambda (IgLλ) or kappa (IgLκ), these polypeptides are formed by combinations between the variable (V), diversity (D), joining (J) and constant (C) gene segments. The heavy chain is produced first, the initial joining taking place in pro-B cells where the D and J segment are spliced together. This function is facilitated by RAG1 and RAG2 enzymes, which recognise recombination signal sequences (RSS) flanking the V

and D segments (Figure 1.2). RAG1 and RAG2 enzymes cleave the sequence between the RSS sites, creating double stranded DNA breaks that are subsequently joined via non-homologous end joining (NHEJ) (Mani and Chinnaiyan, 2010). This is initiated by heterodimerisation between the Ku70 and Ku80 proteins, the binding formed by these proteins attracts DNA-dependent protein kinase catalytic subunit (DNA-PKcs) activating protein kinase activity. This complex brings the DNA ends together and recruits and phosphorylates the DNA nuclease, Artemis, which degrades any single stranded DNA overhangs. Finally a heterodimer consisting of Ligase IV and XRCC4 in the presence of DNA ligases binds the complex and repairs the break (Figure 1.2 A).

Following this initial recombination, the variable sequence is spliced to the DJ segment forming a VDJ segment attached to the remaining constant regions. This is expressed as an mRNA transcript which is spliced together excising all but one constant M segment. The mRNA is then transcribed into a complete heavy chain, which upon expression classifies the cell as a pre-B cell. The heavy chain regions provide different antibody variants through expression of different C chains of which there are five; alpha, beta, gamma, epsilon and delta, however the pre-BCR receptor is always created with a functionally rearranged IgM heavy chain. The membrane bound IgM heavy chain is expressed with either the IgLλ or IgLκ surrogate chains. If signalling is successful, pre-BCR expression leads to several cycles of pre-B cell proliferation. During this process, the IgL chain begins to assemble, combining VJ segments with constant region kappa or lambda chains. During the VJ recombination process of both light and heavy chains, new nucleotides can be added to increase both diversity and specificity of the antibody. If functional, the IgL chain is expressed and assembles with the IgH chain on the cell surface. This creates the fully functioning BCR receptor, which is expressed on the cell membrane, classifying the cell as a surface membrane bound immunoglobulin (smIg+) immature B-cell. The cell will then migrate into the periphery (Figure 1.2 B). B-cells undergo both positive and negative selection during the differentiation process. Positive selection includes successful signalling with the surrogate light chain, it is estimated that as many as half of IgH chain expressing pre-B cells fail to pass this critical checkpoint, with these cells undergoing apoptosis (Vettermann *et al.*, 2006). Negative selection includes failure to successfully undergo V(D)J recombination. Should this occur in one

allele the process begins again in the second allele, if this process fails again the cell will undergo apoptosis. Successful V(D)J recombination results in silencing of the second allele, leading to monospecificity in B-cells (Vettermann and Schlissel, 2010). The majority of autoreactive cells, which have undergone successful V(D)J recombination, undergo apoptosis or anergy (Nemazee, 2006).

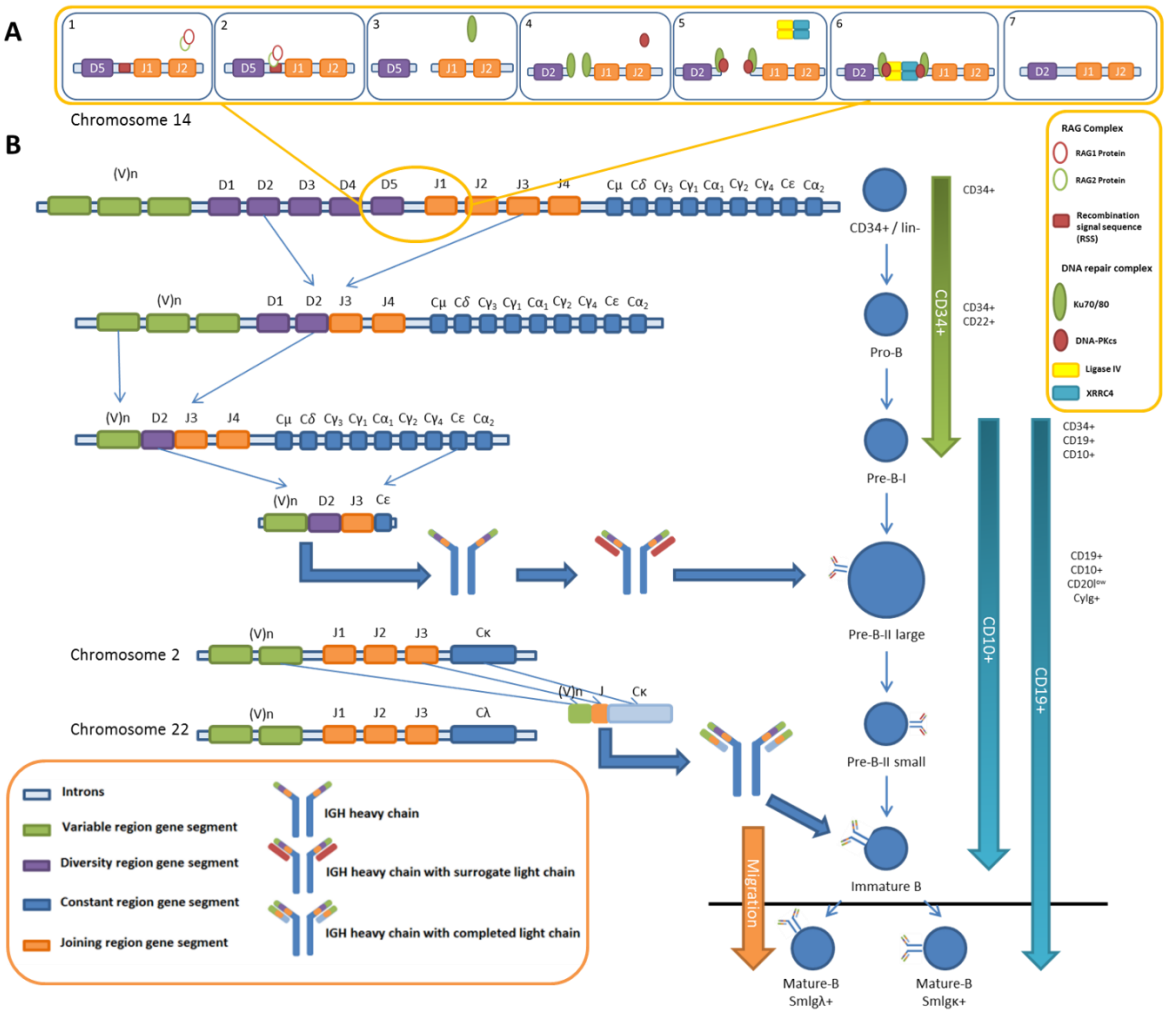


Figure 1.2. Lineage maturation of B-lymphoid cells.

A. NHEJ for the development of mature B-lymphocytes. (1-2) RAG1 and RAG2 proteins create a double stranded DNA break between conserved Recombination Signal Sequences (RSS). (3-5) The Ku70/80 protein heterodimers bind to the exposed DNA ends, this attracts the catalytic subunit of the DNA-dependent protein kinase (DNA-PKCS) and activates kinase activity, the main action of which is to regulate NHEJ. (5-7) Compatible DNA ends are then joined by the Ligase IV/XRRC4 complex, in a reaction stimulated by the XLF/Cernunnos protein. Adapted from Adapted from (van Gent and van der Burg, 2007) and (Pastwa and Blasiak, 2003). B. Sequential V(D)J recombination and B-cell maturation with CD marker expression of maturing lymphoid lineage, adapted from (van Zelm et al., 2005).

1.2 Etiology

Leukaemic development is driven by multiple genetic insults to the cell of origin before full oncogenic transformation is achieved. This theory was originally coined by Alfred G. Knudson whose 'multiple-hit' hypothesis postulated that cancer was the result of multiple genetic aberrations, acquired over time, leading to a pre-cancerous phenotype that is rarely diagnosed. These pre-cancerous cells gain an aberration, described as a driver, which leads to oncogenesis and deregulated cellular proliferation. This theory has been supported by studies in

monozygotic twins, showing several prenatal aberrations which pre-dispose new-borns to leukaemia and create pre-leukaemic phenotypes (Ford *et al.*, 1993; Bateman *et al.*, 2010; Cazzaniga *et al.*, 2011). However, there appears to be some exceptions to the multi-hit theory between leukaemic subtypes. Generally, the initiating mechanisms between infant and childhood leukaemia are believed to be different, with infant ALL resulting from one or more mutations *in utero* and childhood leukaemias being a mix of *in utero* and post-natal genetic lesions. An example in infant ALL is the leukaemia subgroup with *KMT2A* (*MLL*) translocations, which can be initiated from a single hit *in utero,* during foetal haematopoiesis (Greaves *et al.*, 2003). In childhood ALL, the multi-hit model is more relevant, where an initial genetic lesion occurs *in utero*, and at least one secondary postnatal hit initiates the disease. An example is the prevalence of the *ETV6-RUNX1* translocation in new born infants, thought to be at 1% (Mori *et al.*, 2002), which increases susceptibility to ALL but is not sufficient to initiate the disease itself.

The causes of these transforming mutations has been investigated over many years, yet few have been associated with disease initiation with any statistical significance. One clear link was the exposure to high levels of ionizing radiation (~200mSv) of the survivors of the Hiroshima and Nagasaki atom bombings, leading to the development of ALL (Preston *et al.*, 1994). However research on the influence of background ionizing radiation on children, normally at around 2-3mSv, has shown no convincing correlation (Investigators, 2002). The same lack of evidence has been found for the causative effects of electromagnetic field radiation (Skinner *et al.*, 2002). Various other potential causes have been speculated, including car exhaust fumes, pesticides, and parental cigarette smoking, but none with any significant findings (Greaves, 2006). Infections have long been considered a trigger of leukaemia, initially as an effect from specific pathogens, such as the T-cell lymphotrophic virus 1, which has been shown to cause infections leading to T-cell leukaemia and lymphoma (Jaffe *et al.*, 1984). Such direct links however are rare. Later theories, such as the ‘Population Mixing’ hypothesis by Kinlen, built on the infection premise. The theory came from observations of increased incidences of paediatric ALL in populations from initially isolated small towns, which subsequently grew for economic reasons. Influxes of multicultural workers in these towns, where herd immunity was lower

than dense urban centres, led to an increase in viral infections and a threefold increased incidence of ALL. Kinlen theorised that the heightened incidence of ALL was caused by the lack of an immune response to novel pathogens (Kinlen, 1995). This theory is now less favourable as a transforming pathogen is yet to be discovered.

Other theories include Greaves 'delayed-infection' hypothesis, which theorises that infections lead to increased immune system function and in turn cell proliferation, leading to increased likelihood of oncogenic aberrations due to rapid expansion of immune cells (Greaves, 1997). Such data is supported by the lower incidence of ALL in the less affluent developing world (Parkin *et al.*, 1988).

1.3 Treatment and Survival

Survival in ALL has dramatically improved over the last decades as a result of risk stratification of patients for treatment. Patients are typically divided into three main subgroups; high, intermediate and low risk groups, with treatment tailored for each group. Risk stratification is based on several risk factors; white blood cell count (WBC), age, cytogenetics and response to initial therapy. Low WBC is normally a good predictor of outcome with higher blast counts associated with progressively worse outcome. Patients with WBC higher than 50 x 10^9/L are classified as high risk. Patient age is also an indicator of outcome, with patients over the age of 10 years exhibiting worse outcome than those under 10 years. Cytogenetic alterations are a major factor in stratification, typically older patients are more likely to exhibit high risk cytogenetics (Moorman, 2012a). An excellent example of risk stratification leading to improved outcome is the intrachromosomal amplification of chromosome 21 (iAMP21), in which reclassification of patients as high risk was successful in improving outcome (Moorman *et al.*, 2014) (Figure 1.1). Targeted therapies have also improved outcome, such as the tyrosine kinase inhibitor (TKI), Imatinib, which has dramatically improved the dismal survival rates of *BCR-ABL1* positive patients.

Treatment is separated into three stages; 1) induction therapy given upon diagnosis of the disease, aimed at reducing white cell counts and inducing remission, 2) consolidation therapy, which intensifies treatment to remove remaining drug resistant leukaemic cells, and 3) maintenance treatment. The

duration of each stage is dependent upon the risk group and response to treatment (Schrappe *et al.*, 2000). Typically treatment lasts for 2-3 years, with boys being treated for longer than girls due to the risk of testicular relapse. In paediatric ALL, induction therapy typically consists of dexamethasone, vincristine, L-asparaginase and daunorubicin producing complete remission in 95% of children (Moricke *et al.*, 2010). Upon remission, consolidation therapy is a mixture of methotrexate, cytarabine, mercaptopurine, etoposide, and L-asparaginase. Maintenance therapy includes daily doses of 6-mercaptopurine and weekly doses of methotrexate. High risk patients who fail to achieve complete remission are candidates for allogenic stem cell transplantation (Leung *et al.*, 2011). Prevention of central nervous system (CNS) relapse is facilitated by use of systemic intrathecal methotrexate (Pui and Howard, 2008). These modifications have achieved an overall 5 year event free survival rate of around 80-90% (Moorman *et al.*, 2014). Adult treatment is similar, with complete remission achieved in >90% of patients. However due to relapse and drug toxicity, only around 50% of patients reach a 5 year event free survival (Rowe *et al.*, 2005). Older patients, especially >60 years have an unfavourable outcome due to a higher incidence of poor risk cytogenetics, and increased drug toxicity.

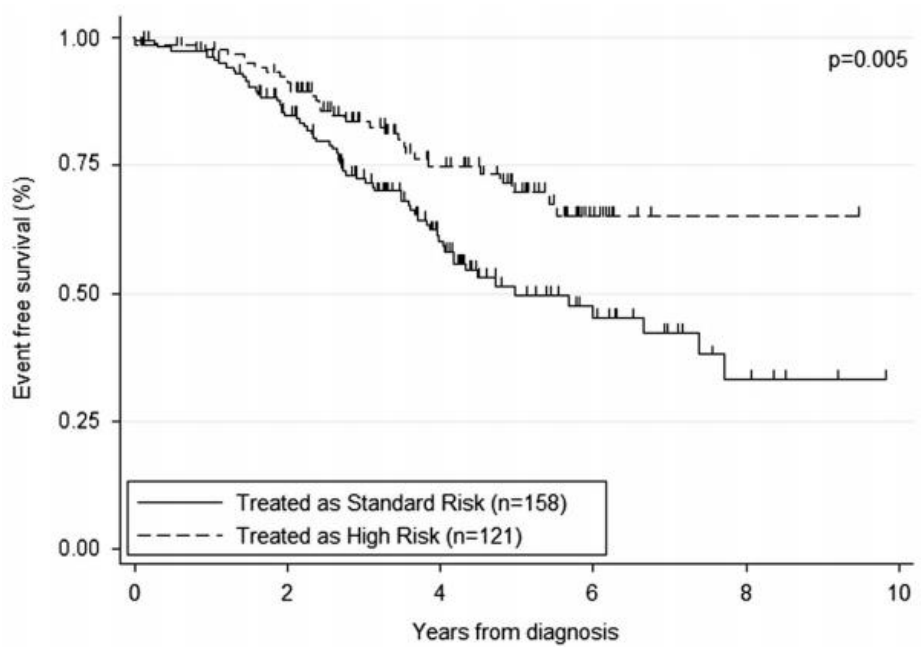


Figure 1.3 Kaplan-Meier survival curves of iAMP21 patient outcome on standard and high risk treatment regimens, showing improved survival of the previously unidentified subgroup. Taken from (Harrison et al., 2014).

1.4 Leukaemia

Leukaemia is broadly divided into four subgroups described by the differentiation stage and lineage of the affected cells. Acute leukaemia arises from haematopoietic progenitors, which have lost the ability to differentiate, leading to accumulation of immature blasts in the bone marrow. This accumulation rapidly spreads throughout the body, including the central nervous system (CNS), requiring early diagnosis and treatment. Chronic leukaemia progresses more slowly, can remain benign for years and comprises more mature blasts. The second classifier is based upon the haematopoietic lineage propagating the malignancy, with the most common leukaemias originating from lymphoid or myeloid lineages (Figure 1.1).

Cytogenetic techniques have been used to identify chromosomal abnormalities in leukaemia for decades. Karyotype analysis led to the identification of the first leukaemia associated chromosomal abnormality in 1960, when the Philadelphia chromosome was described in chronic myeloid leukaemia (CML) (Nowell and Hungerford, 1960). These techniques were further developed including Giemsa banding (G-banding), when it was recognised that Giemsa would stain GC rich chromosomal regions more intensely than the AT regions of less condensed chromatin, creating distinct banding patterns along each chromosome allowing for their precise identification (Seabright, 1971). The development of 'molecular cytogenetics' technique: fluorescence *in situ* hybridization (FISH), has allowed detailed analysis of deletions, gains and rearrangements in interphase nuclei. Here complementary DNA probes are bound to fluorescent tags, which target specific loci. These techniques allowed identification of cryptic translocations and aberrations too small to be visible by conventional cytogenetics alone. FISH first identified the cryptic translocation, t(12;21)(p13;q22), giving rise to the *ETV6-RUNX1* fusion in ALL (Romana *et al.*, 1994). It is also used to screen patients for known oncogenic lesions to aid in treatment stratification, which will be discussed below.

1.4.1 Chronic Leukaemias

Chronic lymphocytic leukaemia (CLL) is a B-cell malignancy affecting the mature B-cells, affecting mainly adults, with over 75% of patients being over the age of 60 years at diagnosis (Chiorazzi *et al.*, 2005).

Chronic myeloid leukaemia (CML) arises from uncontrolled proliferation of mature granulocytes; eosinophils, basophils and neutrophils. It is most prevalent in the middle aged and the elderly population, accounting for 15-20% of all adult leukaemia in the West (Faderl *et al.*, 1999). Over 90% of CML patients express the *BCR-ABL1* fusion gene (Faderl *et al.*, 1999). The disease progresses in three phases; the chronic phase, accelerated phase and blast crisis. At this final stage, CML transforms into an aggressive acute leukaemia with a high mortality rate (Faderl *et al.*, 1999). Recently treatment of CML has focused on arresting the leukaemia in the chronic phase, by using TKIs, which halt disease progression (Gambacorti-Passerini *et al.*, 2011).

1.4.2 Acute Leukaemias

Acute myeloid leukaemia (AML) is the most common acute leukaemia in adults, rarely affecting children. The disease has a median age of 63 years and accounts for ~90% of acute leukaemias in adults (Jemal *et al.*, 2002). It is highly heterogeneous, which has resulted in the creation of complex classification systems. The first system, the French-American-British (FAB) classification system, created in 1976, was based on the originating cell type, the maturity of the abnormal clone, and the cytogenetic profile (Table 1.1). More recently, in 2008, a system was introduced by the World Health Organisation taking into account genetics and AML prognostic factors (Table 1.2).

The French-American-British AML Classification system

Types	Name
M0	Minimally differentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia (APL)
M4	Acute myelomonocytic leukaemia
M4eo	Acute myelomonocytic leukaemia with eosinophilia
M5	Acute monocytic leukaemia
M6	Acute erythroid leukaemia
M7	Acute megakaryoblastic leukaemia

Table 1.1.

The World Health Organisation AML Classification System Adapted from (Vardiman et al., 2009)

Acute myeloid leukaemia with recurrent genetic abnormalities
AML with (8;21)(q22;q22); *RUNX1-RUNT1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
Acute promelocytic leukaemia (APL) with t(15;17)(q22;q12); *PML-RARA*
AML with t(9;11)(p22;q23); *KMT2AT3-KMT2A*
AML with t(6;9)(p23;q34); *DEK-NUP214*
AML with inv(3)(q21q26.2) or t(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 (similar to previous FAB classification)
AML with minimal differentiation (M0)
AML without maturation (M1)
AML with maturation (M2)
Acute myelomonocytic leukaemia (M4)
Acute monoblastic/monocytic leukaemia (M5)
Acute erythroid leukaemias (M6) Pure erythroid leukaemia Erythroleukaemia, erythroid/myeloid
Acute megakaryoblastic leukaemia (M7)
Acute basophilic leukaemia
Acute panmyelosis with fibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis
Myeloid leukaemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasms

Table 1.2.

Approximately 70% of patients diagnosed with AML have defined chromosomal abnormalities, which are strongly linked to outcome, as indicated in Figure 1.4. AML chemotherapy is divided into induction and post-remission phases. HSC

transplantation is used if a patient fails induction therapy, or relapses (Pulte *et al.*, 2008).

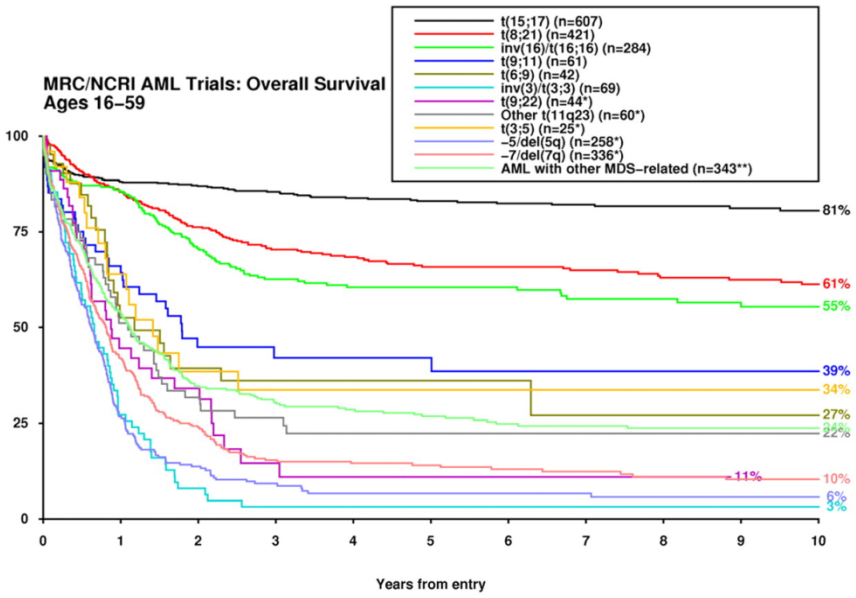


Figure 1.4. Kaplan Meier curve of AML patient survival separated by cytogenetic subgroups over ten years taken from (Grimwade et al., 2010).

1.4.3 Acute Lymphoblastic Leukaemia (ALL)

The majority of BCP-ALL are defined by recurrent numerical or structural chromosomal abnormalities, which are linked to outcome (Figure 1.5) (Harrison, 2013). There are clearly defined good and poor risk subgroups (Figure 1.6) (Moorman, 2012a). The relative incidences of the most significant cytogenetic subgroups in BCP-ALL are shown in Figure 1.4 and their variable outcome within childhood and adult cohorts are shown in Figure 1.5.

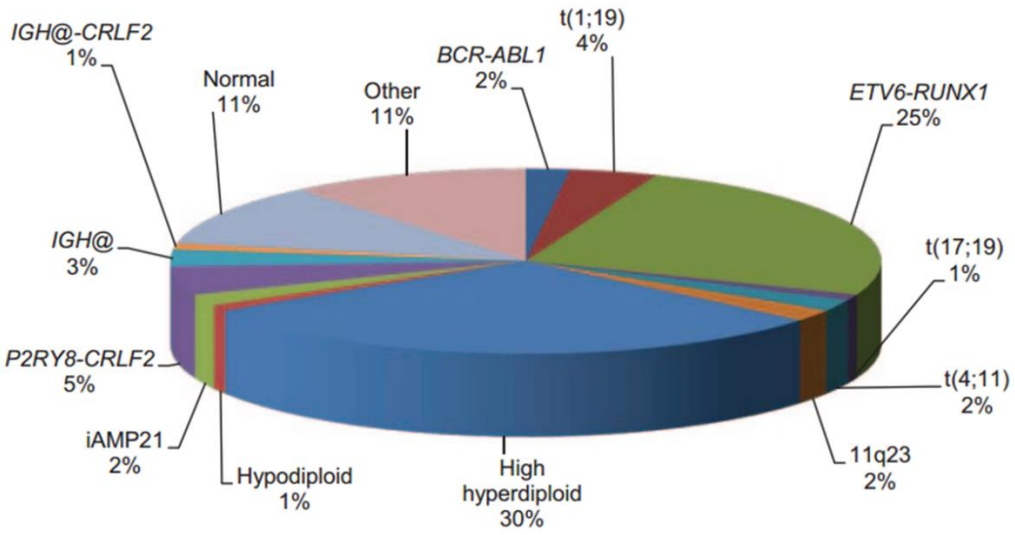


Figure 1.5. Incidence of established cytogenetic abnormalities among childhood ALL, data from UK ALL trials (Harrison, 2013).

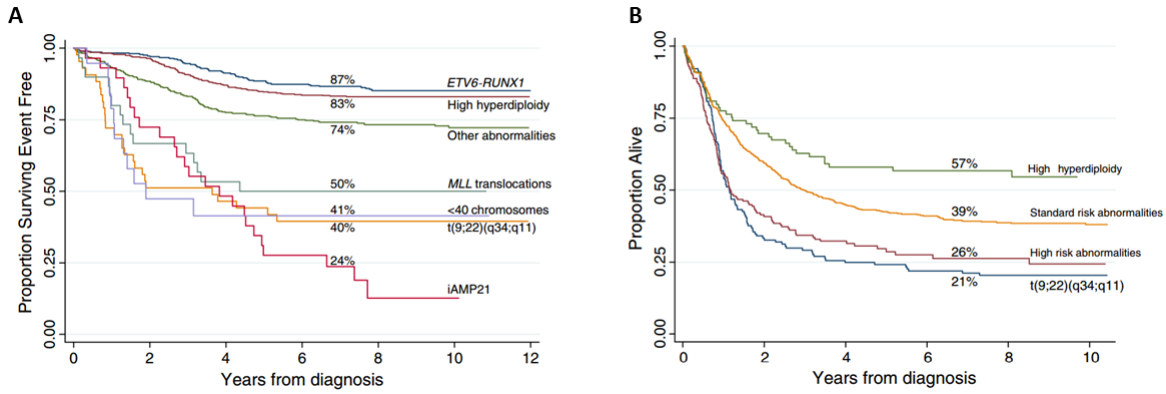


Figure 1.6. Kaplan Meier curves of childhood and adult survival divided by subgroups, indicating the impact of subgroup on overall survival.

A. Kaplan Meier curve of a childhood trial (ALL97) event free survival of the main cytogenetic subgroups in paediatric ALL. B. Kaplan Meier curve of an adult trial (UKALLXII) event free survival for risk groups in adult ALL, taken from (Moorman, 2012a).

1.4.4 Established BCP-ALL Subgroups

The majority of BCP-ALL subgroups are defined by recurrent cytogenetic alterations, involving either aneuploidy or translocations. Cytogenetic classification is vitally important in determining patient treatment options. Below are some of the most common BCP-ALL subgroups.

1.4.4.1 *ETV6-RUNX1*/t(12;21)(p13;q22)

The t(12;21) translocation forming the *ETV6-RUNX1* gene fusion is the most commonly occurring translocation in BCP-ALL at ~30% (Golub *et al.*, 1995; Romana *et al.*, 1995), and is found almost exclusively in paediatric patients (Codrington *et al.*, 2000). The translocation is cytogenetically cryptic, requiring FISH or Reverse Transcription Polymerase Chain Reaction (RT-PCR) to be identified. The translocation creates a fusion gene which contains the protein dimerization domains of *ETV6* and the majority of the DNA binding and activating regions of *RUNX1*, allowing *ETV6* to de-regulate the expression of multiple *RUNX1* target genes, a number of which are critical for normal haematopoietic development (Fenrick *et al.*, 1999). This process leads to enhanced proliferation

and heightened self-renewal of BCP cells, which are unable to differentiate (Mori *et al.*, 2002). The translocation is present in early B-cell progenitors *in utero*, creating a pre-leukaemic clone, which may reside in the bone marrow for several years (Mori *et al.*, 2002). Patients with *ETV6-RUNX1* gene fusion tend to have an average of 11 additional genetic lesions, which is higher than many other subgroups (Papaemmanuil *et al.*, 2014). Almost all patient trials show excellent overall survival and low incidence of relapse (Figure 1.6 A) (Conter *et al.*, 2010; Moorman *et al.*, 2010b).

1.4.4.2 *BCR-ABL1*/t(9;22)(q34;q11)

The *BCR-ABL1* fusion arises from the translocation, t(9;22)(q34;q11) or a variant rearrangement (Nowell and Hungerford, 1960). It results in constitutive activation of the *ABL1* tyrosine kinase, leading to phosphorylation of a number of important downstream genes, activating multiple molecular pathways that lead to proliferative deregulation and decrease in cellular differentiation and adhesion. The fusion is predominant in CML but also present in ALL and AML. In ALL the *BCR-ABL1* fusion is often accompanied by an intragenic deletion of the *IKZF1* gene (Mullighan *et al.*, 2008a). There are three fusion protein variants identified, however all lead to constitutive activation of the ABL1 protein. Incidence increases with age, accounting for 2% of childhood ALL, rising to 40% in adults over 40 years (Moorman *et al.*, 2007a; Burmeister *et al.*, 2008; Moorman *et al.*). Expression of the *BCR-ABL1* fusion protein in ALL was associated with a dismal outcome until the development TKIs, including Imatinib and Dasatinib which have improved three year EFS rates to 80% in children and 60% in adults (Ottmann and Pfeifer, 2009).

1.4.4.3 *KMT2A* (*MLL*) Translocations

KMT2A regulates haematopoiesis partly through transcriptional activation of *HOX* genes. The gene has over 100 fusion partners seen in ALL and AML and was first identified as an oncogene through use of FISH and southern blotting (Ziemin-van der Poel *et al.*, 1991). The general mode of deregulation is due to the fusion of the 5' part of the *KMT2A* gene with the 3' end of the partner gene (Marschalek, 2011). *KMT2A* translocations are unique in being able to initiate

leukaemogenesis as a seemingly sole genetic lesion. They have been shown to be initiated *in utero*, with a high predominance in infants (Mullighan *et al.*, 2007). The most common *KMT2A* translocation in ALL is *KMT2A-AFF1*/t(4;11)(q21;q23) at an incidence of 50% of all paediatric *KMT2A* translocations (Meyer *et al.*, 2009). *KMT2A* translocations are predominant in infants with an incidence of 70-80%, dropping to ~2% in children >2 years and is associated with a poor outcome across all ages (Pieters *et al.*; Moorman *et al.*, 2010b).

1.4.4.4 High Hyperdiploidy (HeH)

HeH is defined as a non-random gain of 5-19 whole chromosomes. The most frequently gained are chromosomes 4, 6, 10, 14, 17, 18, 21 and X (Paulsson *et al.*, 2010). HeH is one of the most common genetic subgroups of ALL in ~30% of children and ~10% of adults (Moorman *et al.*, 2010a). Prognosis is good, however, due to the high number of cases, relapse remains a significant problem (Sutcliffe *et al.*, 2005; Moorman *et al.*, 2010b). Recently whole genome sequencing (WGS) and whole exome sequencing (WES) was performed on a total of 55 HeH patients. Analysis showed involvement of the RTK-RAS pathway and of histone modifiers in the majority of these patients, identifying potentially new therapeutic targets for the subgroup. There was also a strong indication that the chromosomal gains in the subgroup were early events, supporting the idea that HeH is the main oncogenic driver in these leukaemias (Paulsson *et al.*, 2015).

1.4.4.5 *TCF3-PBX1*/t(1;19)(q23;p13)

The *TCF3-PBX1* fusion gene occurs as a result of either a balanced t(1;19) translocation or an unbalanced translocation, where only the derived 19 is observed (Barber *et al.*, 2007). Patients with the translocation show a pre-B immunophenotype expressing cytoplasmic μ. The occurrence of the subtype does not vary with age and comprises 3-5% of ALL (Mancini *et al.*, 2005). *TCF3* is required for the development of B and T lymphocytes. The *TCF3-PBX1* gene fusion transcript places the DNA targets of *PBX1* under the transcriptional control of *TCF3*, leading to deregulation of affected genes and leukaemogenesis

(Mellentin *et al.*, 1989). More recently, treatment intensification for this subgroup has resulted in improved outcome (Felice *et al.*, 2011).

1.4.5 Impact and Incidence of Genetic Aberrations

The importance of collaborating aberrations in ALL is dependent upon the subgroup; *ETV6-RUNX1* positive ALL tends to have a higher incidence of associated secondary abnormalities of 6-8 genetic lesions per patient. In contrast, patients with translocations involving *KMT2A* have none or few associated secondary changes, with the deregulation of *KMT2A* alone being sufficient to initiate leukaemogenesis (Marschalek, 2011). Up to 25% of childhood ALL patients show no established chromosomal rearrangements. These so-called B-other ALLs are presumably driven by acquired submicroscopic abnormalities (Harrison, 2009). The secondary aberrations most frequently observed occur in genes controlling B-cell differentiation, cell cycle and those involved in key signalling pathways:, including *PAX5*, *IKZF1*, *CDKN2A*/*B*, *RB1*, *EBF1*, *ETV6* and *BTG1*. The most common abnormalities are discussed below in further detail.

1.4.5.1 PAX5

PAX5 is a vital early regulator of B-cell development and is one of the most recurrently abnormal genes in BCP-ALL, occurring in ~30% of patients (Familiades *et al.*, 2009) (Mullighan *et al.*, 2007). Multiple mutations of the *PAX5* gene lead to haplosufficiency. Focal internal deletions also occur leading to truncated proteins, or truncated DNA binding and transactivation domains. *PAX5* is also translocated with multiple partners (Mullighan *et al.*, 2007; Nebral *et al.*, 2009). Genetic lesions of *PAX5* however have not been linked to patient outcome.

1.4.5.2 IKZF1

IKZF1 aberrations occur in 15% of BCP-ALL patients and have been associated with a poor outcome (Kuiper *et al.*, 2010). They are more common in high risk ALL subgroups such as *BCR-ABL1* and *BCR-ABL1*-like, with up to a third of patients exhibiting genetic lesions (Mullighan *et al.*, 2008a; Harvey *et al.*, 2010b).

Aberrations are varied, the predominant form of de-regulation in BCP-ALL is through whole gene or exon 4-7 deletions. The latter leads to the loss of 4 N-terminal zinc fingers that mediate DNA binding, resulting in the formation of the dominant negative protein isoform, Ik6 (Mullighan *et al.*, 2008a). Aberrations in this gene are also associated with a threefold risk of treatment failure (Mullighan *et al.*, 2009b; Kuiper *et al.*, 2010). The mechanism behind *IKZF1* loss and leukaemogenesis is unclear, although it has been proposed that *IKZF1* exerts cell cycle control through inhibition of *MYC* (Ma *et al.*, 2010), as such, loss of *IKZF1* would lead to deregulated proliferation in affected cells.

1.4.5.3 CDKN2A/B

CDKN2A/*B* deletions are the most commonly occurring genetic lesion in BCP-ALL (Schwab *et al.*, 2013). The genes lie within the *INK4*/*ARF* tumour suppressor locus on human chromosome 9p21.3, which codes for several tumour suppressors (Sherr, 2012). *CDKN2A* codes for the p16 protein, and *CDKN2B* for the p15 protein, both of which act by slowing cell cycle progression from G1 to S phase by preventing activation of *CDK4* and *CDK6* kinases by *CCND1*. This inhibition halts phosphorylation of the RB1 protein and in turn prevents the nuclear localisation of *E2F1,* which is instrumental in the progression of cell cycle from G1 to S phase (Rayess *et al.*, 2012).

1.4.6 B-Other Subgroups

BCP-ALL is a disease characterised by cytogenetic abnormalities, however approximately 25% of BCP-ALL patients show no consistent cytogenetic aberrations, and remain unclassified. Without specific characteristics these patients have been grouped together as intermediate risk. In recent years however, cytogenetic and molecular analysis has revealed several recurring genetic lesions among patients within this subgroup. Both *IGH* translocations and the iAMP21 subgroup were initially classified as B-other, prior to independent classification (Harewood *et al.*, 2003; Chapiro *et al.*, 2006; Akasaka *et al.*, 2007; Robinson *et al.*, 2007; Russell *et al.*, 2008; Russell *et al.*, 2009). Gene expression profiling identified a subgroup among these patients, with a similar profile to those expressing the *BCR-ABL1* fusion protein (Den Boer *et al.*, 2009; Harvey *et al.*,

2010b), termed *BCR-ABL1*-like/Philadelphia-like (Ph-like) ALL, this group is discussed below. Approximately 12% of B-other patients exhibit intragenic *ERG* deletions. They have a distinctive gene expression profile, linked to a good outcome (Clappier *et al.*, 2014). Other abnormalities in the B-other group include *PAX5* abnormalities (Forestier *et al.*, 2008; Nebral *et al.*, 2009; Moorman *et al.*, 2010b), and the *IGH* translocations with unknown partner genes (Russell *et al.*, 2014). Around a quarter of B-other patients however remain unclassified, with no recurring genetic lesions observed to date.

1.4.6.1 Intrachromosomal Amplification of Chromosome 21 (iAMP21)

iAMP21 is an abnormality involving the long arm of chromosome 21 with multiple regions of deletion, inversion and amplification, occurring in 2% of childhood BCP-ALL (Moorman *et al.*, 2013). This abnormality was identified during routine FISH screening for the *ETV6-RUNX1* fusion, when, in the absence of the fusion, multiple copies of clustered *RUNX1* signals were observed in a number of patients (Harewood *et al.*, 2003). This abnormality has since been defined as three or more additional copies of *RUNX1* seen by FISH (Moorman *et al.*, 2007b; Robinson *et al.*, 2007; Rand *et al.*, 2011). In the majority of cases, iAMP21 is the primary event, remaining constant in both diagnostic and relapse samples (Rand *et al.*, 2011). Copy number arrays of iAMP21 patients have revealed a common region of amplification of 5.1Mb on the long arm of chromosome 21, including *RUNX1.* While the *RUNX1* gene is located within this region, there is no evidence to indicate that it plays a role in the development of ALL, supported by the observation that fold change of *RUNX1* mRNA remain the same in iAMP21 patients when compared to other BCP-ALL patients with chromosome 21 abnormalities (Strefford *et al.*, 2006). iAMP21 is associated with a host of secondary aberrations including gain of chromosomes 10, 14 and X, monosomy 7, deletions of chromosomal arms 7p, 11q and genes *ETV6*, *RB1,* and the *P2RY8-CRLF2* fusion (Harrison *et al.*, 2014). Interestingly, despite sharing the common region of amplification with the Down Syndrome (DS) critical region, only one DS patient has been discovered with iAMP21 (Harrison *et al.*, 2014). Recently individuals with the constitutional Robertsonian translocation, rob(15;21)(q10;q10)c, have been shown to have an extremely high risk of developing iAMP21-ALL (Li *et al.*, 2014). iAMP21 patients had a high relapse risk

on standard therapy (5 year EFS 29%), this improved dramatically when they were treated as high risk (5 year EFS 78%). Relapse risk also decreased from 70% to 16% and overall survival rose from 67% to 89% (Moorman *et al.*, 2013). Patients with iAMP21 are now treated as high risk.

1.4.6.2 *BCR-ABL1*-Like/Ph-Like

Originally part of the B-other group, *BCR-ABL1*-like/Ph-like patients were identified from the similarity of their gene expression profiles to *BCR-ABL1* positive patients. This *BCR-ABL1*-like subgroup comprises 10% of childhood ALL (Den Boer *et al.*, 2009; Harvey *et al.*, 2010b). The most common abnormalities are *IKZF1* deletions and deregulated expression of *CRLF2*, both occurring in around 50%. The remaining 50% of non-*CRLF2* patients show rearrangements affecting other kinases including *PDGFRB*, *ABL1*, *ABL2, JAK2*, *EPOR*, *IL-7R* and *SH2B3*. Several of these aberrations activate the *JAK-STAT* pathway (Figure 1.7). Currently it is proposed that the outcome for this subgroup is very poor; however as a novel subgroup these patients are yet to be incorporated into risk stratification algorithms (Den Boer *et al.*, 2009). Recently is has been shown that patients in this subgroup who respond poorly to initial treatment can be salvaged through use of minimal residual disease (MRD) based risk directed therapy (Roberts *et al.*, 2014b).

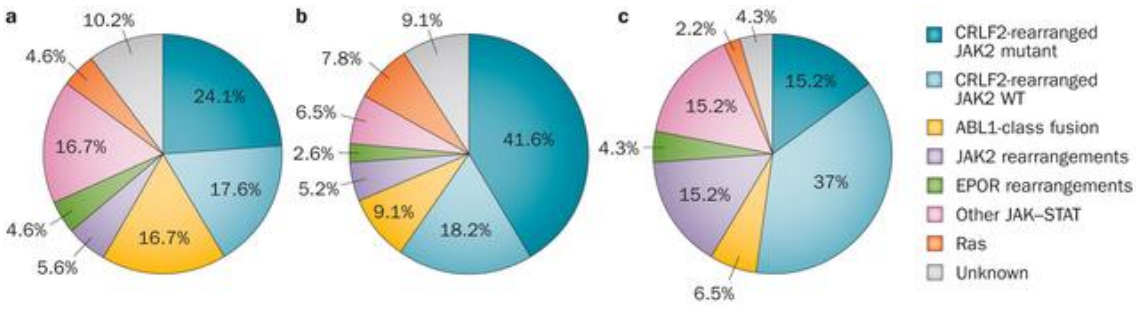


Figure 1.7 BCR-ABL1-like ALL subgroup divided by associating abnormalities in a. Children, b. Adolescents, and c. Young Adults. Taken from (Roberts and Mullighan, 2015), originally adapted from (Roberts et al., 2014a).

1.4.6.3 *Immunoglobulin Heavy Chain Locus* (*IGH*) Translocations

Translocations within the *IGH* locus lead to the overexpression of juxtaposed genes. This is due to the presence of powerful enhancer elements located in the region, which drive expression of multiple gene segments. The pathogenesis of these translocations has been well characterised in several subtypes of mature B-cell malignancies including *IGH-MYC* in Burkitt lymphoma, *IGH-BCL2* in follicular lymphoma and *IGH-CCND1* in mantle cell lymphoma (Willis and Dyer, 2000). A recent analysis of the *IGH* subgroup showed that they comprised 5% of BCP-ALL, with involvement in both B and T-ALL (Russell *et al.*, 2014). The most common partner genes are *CRLF2* at 22%, the *CEBP* gene family at 11% and *ID4* at 7%. The remainder occur at very low levels and are a mix of established oncogenes (*EPOR*, *mir-125b*, *BCL* gene family), new partners (*LHX4*), or antibody specific loci (*IGK*, *TCRA*/*D*) (Russell *et al.*, 2014). For over half of *IGH* translocations the partner gene is as yet unidentified, providing scope for further investigation. Incidence of *IGH* translocations varies according to partner gene; in patients with *CRLF2*, *CEBPA*, *CEBPB* and *BCL2* as partners, the majority of blasts harbour the translocation, suggesting that these genetic aberrations occur early in the development of the leukaemia. In contrast, patients with *IGH-CEBPE* exhibit low level populations suggesting a secondary role in the development of BCP-ALL. HeH is the most common associating aberration (n=9/148, 6%), followed by *ETV6-RUNX1* (n=8/148, 5%), *BCR-ABL1* (n=6/148 4%) and *KMT2A* (n= 2/148, 1%) rearrangements (Russell *et al.*, 2014). Other genetic insults include *CDKN2A*/*B* and *IKZF1* deletions occurring in ~40% of patients. The incidence of *IGH* translocations increases with age, peaking in 20-24 year olds, who comprise 11% of all *IGH* patients with those younger than 5 years accounting for only 2%. Older patients peak in the 30-34 and 45-49 age ranges (Figure 1.8). DS patients account for 16% of cases. Survival analysis showed that children with *IGH* translocations were more likely to belong to the NCI high risk group when compared to *IGH* negative children, 68% and 42%, respectively, and were more likely to be MRD positive at day 28, 72% vs 47%. The association with poor outcome was not due to increased relapse rates, but due to higher death rates following relapse. In young adults and adults the translocation was found to lead to an adverse outcome over an eight year follow up period (P = 0.002) (Russell *et al.*, 2014).

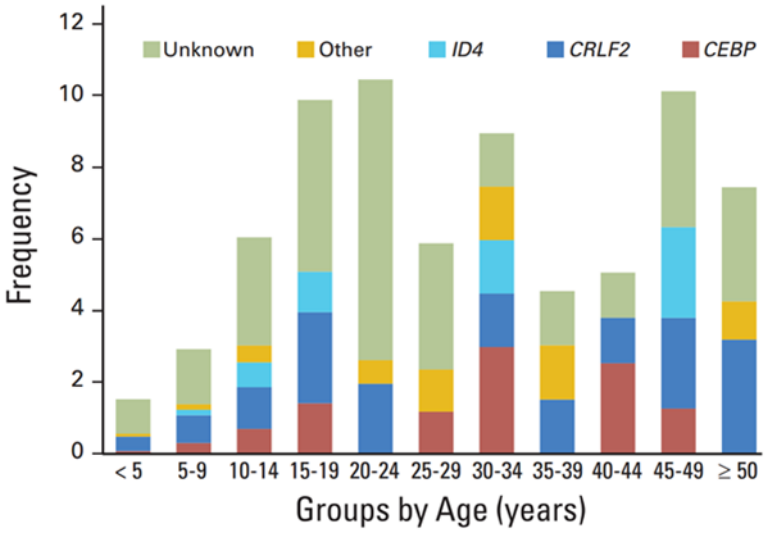


Figure 1.8. Incidence of IGH translocations by age and partner gene, taken from (Russell et al., 2014).

1.4.6.4 CRLF2

CRLF2 codes for the receptor bound by thymic stromal lymphopoietin (TSLP), which along with IL-7 leads to activation of the *JAK-STAT* pathway. The most common form of *CRLF2* deregulation occurs from an interstitial deletion within the Pseudoautosomal Region 1 (*PAR1*), located to the p arm of the X and Y chromosomes, which juxtaposes the *P2RY8* promoter next to the *CRLF2* gene. The second common method is the t(X;14)(p22;q32) or t(Y;14)(p11;q32)/*IGH-CRLF2* translocation, with the ratio between these two forms of deregulation of 5:1 *P2RY8*:*IGH* (Russell *et al.*, 2009). The third rarely occurring method of deregulation is through activating mutations of *CRLF2*, being an F-C amino acid substitution at position 232 (Chapiro *et al.*, 2010; Hertzberg *et al.*, 2010), which was confirmed to lead to *IL-3* independent growth *in vitro* and increased colony formation *in vivo* (Chapiro *et al.*, 2010). *CRLF2* overexpression constitutively activates the *JAK-STAT* pathway leading to expression of a large panel of genes including inflammatory factors *JUN*, *FOS* and *NF-κB*. These abnormalities occur in ~5% of both children and adults, with the incidence of *IGH-CRLF2* subtype increasing with age. There are conflicting reports regarding outcome in *CRLF2* deregulated patients. Data from the UK show 5 year survival figures consistent with intermediate risk (Ensor *et al.*, 2011). Research from the US indicates a poor outcome, however these data may be skewed due to the large numbers of Hispanic patients in the study cohort, which have been shown to have higher

incidence of BCP-ALL (Harvey *et al.*, 2010a; Dores *et al.*, 2012). *CRLF2* deregulation is the most common cytogenetic abnormality in DS ALL with 50% of patients having the abnormality (Mullighan *et al.*, 2009a; Russell *et al.*, 2009; Chapiro *et al.*, 2010; Hertzberg *et al.*, 2010). More recent analysis of *CRLF2* deregulation suggested that the abnormality negatively impacts on EFS and overall survival (OS) of adolescents and adults (Moorman, 2012a). More recently the subgroup has been characterised at the molecular level. Primary *CRLF2* rearranged ALL samples were stimulated *in vitro* with the *CRLF2* ligand TSLP, which showed signal transduction through the *JAK-STAT* and *PI3K*/*mTOR* pathways. These primary samples were then treated with JAK inhibitor: Ruxolitinib, *mTOR* inhibitor: Rapamycin, and *PI3k*/*mTOR* inhibitor: PI103. Ruxolitinib was shown to inhibit both the *JAK-STAT* and the *PI3K*/*mTOR* pathways, while the other two inhibitors showed specificity for the *PI3K*/*mTOR* pathway, giving scope for multiple inhibitor molecules in future treatment protocols (Tasian *et al.*, 2012). Further targeted treatment of the subgroup has not been published, however the use of *JAK* inhibitors has shown promise in Ph-like patients exhibiting *JAK* mutations and *CRLF2* deregulation in xenograft models (Maude *et al.*, 2012; Roberts *et al.*, 2014a).

1.5 CCAAT Enhancer Binding Proteins (CEBPs)

The *CEBPs* are a family of six multifunctional basic leucine zipper (bZIP) transcription factors, which function by regulating mRNA transcription, cell differentiation, proliferation control, metabolism and immunity. The genes are located on different chromosomes, *CEBPA* and *CEBPG* at 19q13, separated by 71kb, *CEBPB* at 20q13, *CEBPD* at 8q11, *CEBPE* at 14q11 and *CEBPZ* at 2p22. The transcription factors function by binding with the CCAAT (cytosine-cytosine-adenosine-adenosine-thymidine) box motif located within a range of gene promoter regions. The genes are well established in myeloid, adipose and hepatic genesis and differentiation.

1.5.1 IGH-CEBP in BCP-ALL

Translocations involving *IGH* have been observed in five of the six CEBP family members (Akasaka *et al.*, 2007). They comprise the second most common *IGH*

subgroup: 11% of *IGH* and around 0.6% of ALL as a whole (Russell *et al.*, 2014). The *IGH* translocations are: t(14;19)(q32;q13) for *CEBPA/G,* t(14;20)(q32;q13) for *CEBPB* t(8;14)(q11;q32) for *CEBPD* and inv(14)(q11q32)/t(14;14)(q11;q32) for *CEBPE* with the most common translocation being *IGH-CEBPD* (Russell et al., 2014). Median age at diagnosis is 15 years with a low white blood cell count (median $9x10^9$/L) (Akasaka *et al.*, 2007). These translocations are unique to ALL, although the mechanism of leukaemogenesis remains unknown. A study of 44 patients showed DS to be the most commonly associated cytogenetic abnormality with *IGH-CEBPD* in 12 of the 44 patients (Lundin *et al.*, 2009). This link is specific to *IGH-CEBPD*, with one exception of an *IGH-CEBPE* DS patient reported (Akasaka *et al.*, 2007). *CEBPA* is a myeloid differentiation gene exclusively expressed in myeloid and monocytic lineages (Nerlov, 2007). It has been described as a tumour suppressor in this setting, thus expression in the lymphoid lineage and action as a tumour promoter is an unexpected finding. Typically observed as a single phenotype translocation, a *IGH-CEBPE* translocation was discovered in a patient with a complex karyotype, involving unbalanced rearrangements of both chromosomes 14; der(14)t(13;14)(q21;q21) and dup(14)(q11q32), giving rise to two individual *IGH-CEBPE* translocations in the same cells. This patient provides the first evidence of a chromosome duplication and cryptic insertion creating a *IGH-CEBPE* fusion (Pierini *et al.*, 2011).

1.5.2 Structure and Function

The C-terminus of the *CEBP* genes is highly conserved, with a sequence similarity of >90%. This region contains the leucine zipper and basic DNA binding domains and the sequence similarity translates into similar structural and functional properties of all family members (Figure 1.9). The leucine zipper consists of a heptad of leucine repeats, which form an alpha helix (Vinson *et al.*, 1993). This protein structure allows bZIP family members to homo and heterodimerise with each other; a crucial step in the action of the *CEBPs,* which stabilises the proteins interaction with target DNA (Cooper *et al.*, 1995; Parkin *et al.*, 2002). The DNA binding domain is located upstream of the bZIP, the function of which is determined by approximately 20 amino acids in the region (Johnson, 1993). The N-terminus of the CEBP proteins contain the transactivating and

regulatory domains, which allow protein interactions. This region is more variable than the C-terminus. Different *CEBP* genes and isoforms express different numbers of transactivating domains (TD). These domains have been shown to exert varying potencies of transactivation and have been observed to bind different targets, but also act synergistically (Nerlov and Ziff, 1995). Differing isoform function is in part conferred by variation in which TD is expressed. For example, *CEBPA* contains three TDs: TDI, TDII and TDIII, with different *CEBPA* isoforms expressing different TDs. Isoform p42 expresses the most powerful TD, TDI, which is lacking in the shorter p30 isoform (Figure 1.9). *CEBPG* contains no TDs, giving rise to the theory that *CEBPG* is a dominant negative isoform functioning by inactivation of other *CEBP* and bZIP family members. The other *CEBPs* contain varying numbers of TDs depending upon the gene and protein isoform (Figure 1.9).

Regulatory domains are also present in the N-terminus, with four of the *CEBP* genes expressing these domains (Figure 1.9), which function in an inhibitory manner. Regulatory domain I has been shown to have binding sequences for the small ubiquitin-related modifier (SUMO) protein, which post transcriptionally modifies its targets and generally functions by SUMOylating and inhibiting transcription factors. Further research has since discovered that SUMOylation exerts a regulatory effect on specific CEBP isoforms in several different settings, exerted by multiple *SUMO* family genes (Eaton and Sealy, 2003; Wang *et al.*, 2006; Khanna-Gupta, 2008).

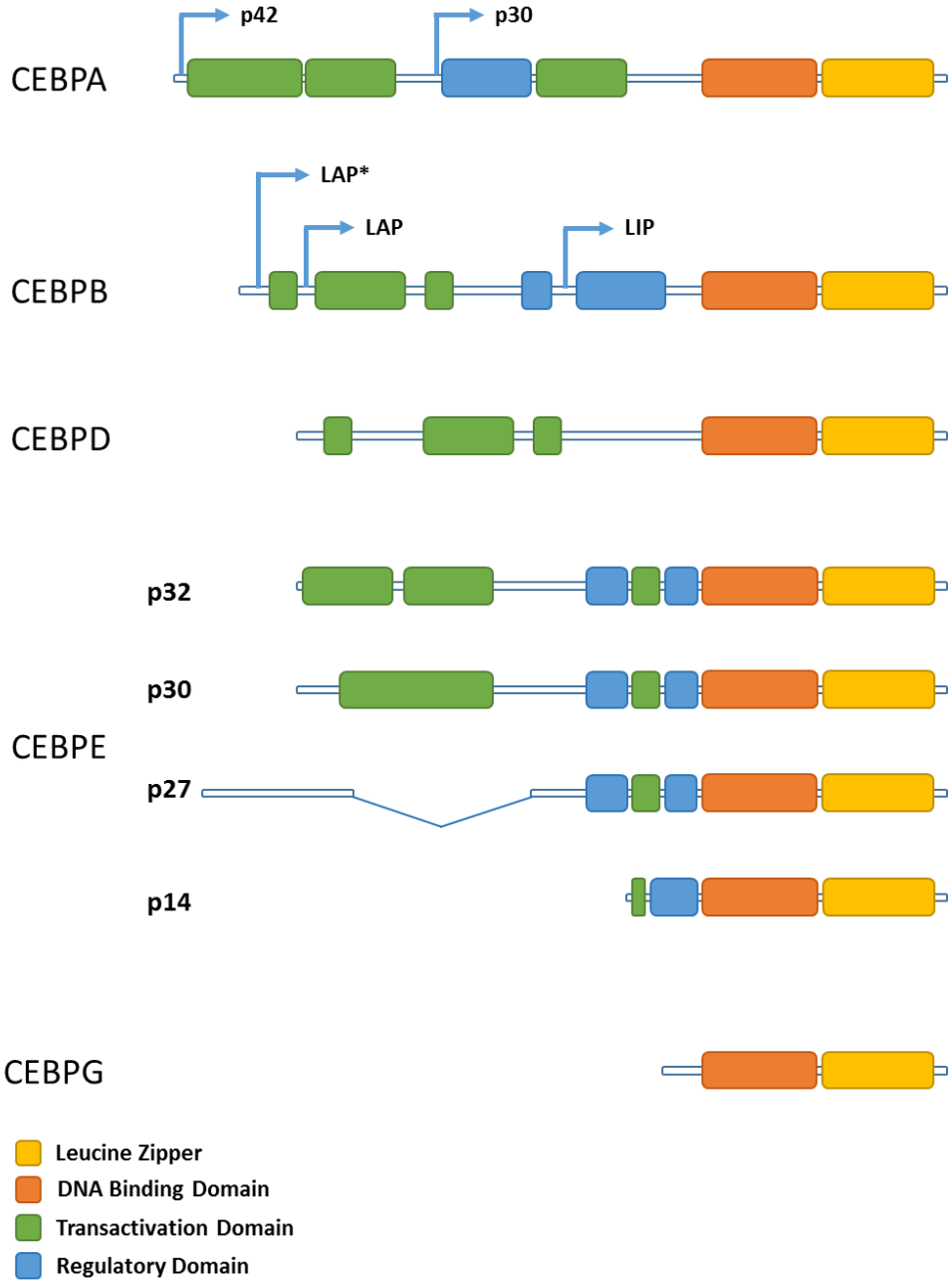


Figure 1.9. IGH partnered CEBP protein domain loci, with protein isoforms.

The proteins show a conserved C-terminal region with the leucine zipper domain, allowing for protein heterodimerisation, and a DNA binding domain. The N-terminus shows more variation with different combinations of transactivating and regulatory domains present, dependent upon gene and protein isoform. CEBPA is predominantly expressed as two isoforms, the full length p42 isoform and the regulatory p30 isoform. CEBPB is expressed as three isoforms, the activating LAP and LAP isoforms, and the negative LIP isoform. CEBPD is expressed as a single isoform. CEBPE is expressed as four p32, p30, p27 and p14 isoforms whose interplay controls terminal granulocytic differentiation. CEBPG shows no activating or regulatory domains and is believed to mainly function as a dominant negative protein for the other CEBP members. Adapted from (Schrem et al., 2004).*

CEBP isoforms are created by altered ribosomal translation points, which begin at codons downstream of the mRNA sequence (Lin *et al.*, 1993; Ossipow *et al.*,

1993). This translation is controlled by the *mTOR* pathway, which determines *CEBP* isoform ratios through the *eIF4E* gene (Calkhoven *et al.*, 2000).

Isoform expression is particularly important for *CEBPA, CEBPB* and *CEBP*E, all of which have dominant negative isoforms influencing the function of their full length CEBP proteins (Figure 1.9). The CEBP proteins exhibit largely identical DNA binding specificities *in vitro*, making it difficult to predict which CEBP protein will bind to which sequences (Tsukada *et al.*, 2011). Although DNA binding specificity is generally unique, the CEBP proteins can function in place of each other. For example, in haematopoiesis and adipocyte differentiation, *CEBPA* can function in place of *CEBPB* (Jones *et al.*, 2002; Chiu *et al.*, 2004). Synergy has also been observed in CEBP heterodimers, with *CEBPB* and *CEBPD* acting in concert in inflammatory signalling (Yan *et al.*, 2012). In contrast *CEBPA* and *CEBPD* play opposing roles in hypoxia signalling through the *HIF-1α* gene, with *CEBPA* inhibiting its function and *CEBPD* supporting *HIF-1α* expression (Balamurugan and Sterneck, 2013). Other examples will be discussed below.

As well as forming hetero- and homodimers within the CEBP family, the proteins can also form heterodimers with other bZIP family members, such as the *CREB*/*ATF* and *FOS*/*JUN* family, generally to facilitate existing CEBP functions, such as immune signalling and response, as well as differentiation control (Tsukada *et al.*, 1994; Newman and Keating, 2003). They also repress *CEBP* function (Podust *et al.*, 2001).

While the *CEBP* genes act in concert in haematopoiesis (Figure 1.10) and in the development of other tissues, there is a level of complexity in their interplay, both in specific tissues and with each other. It is understood that the proteins can and do heterodimerize, but the results of such dimerization are largely unclear. Even known functions can be inverted depending upon context. Below is a brief description of some of the many functions of the *CEBP* gene family members (Figure 1.10).

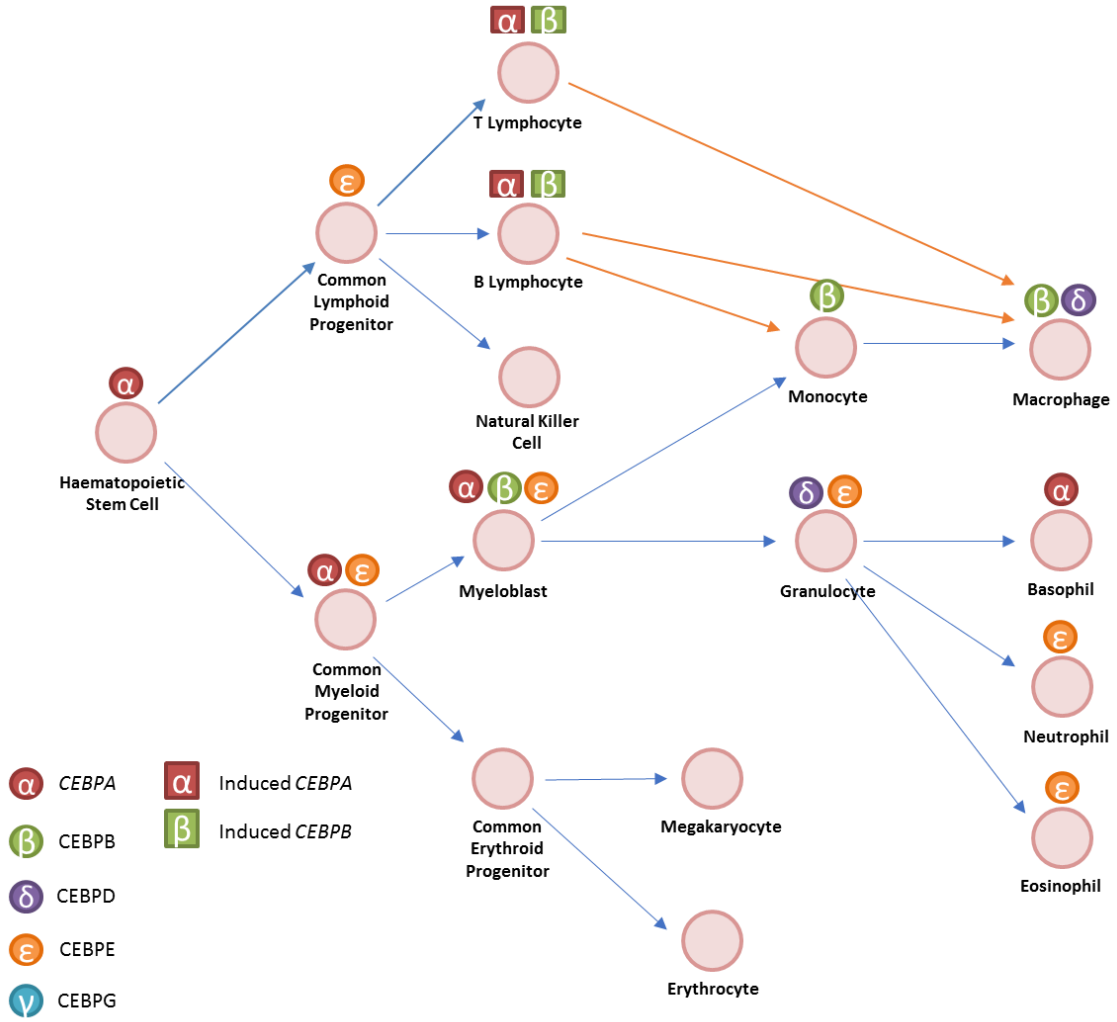


Figure 1.10. Natural and induced CEBP expression in haematopoiesis from published reports (Tsukada et al., 2011).

Blue arrows indicate natural lineage progression, orange arrows indicate experimentally induced differentiation (Heavey et al., 2003; Xie et al., 2004). CEBP expression is mostly sequential, beginning with CEBPA directing haematopoiesis down the myeloid lineage, and continuing with interplay between the CEBP genes in CMPs and myeloblasts and ending with specific CEBPs contributing to the terminal differentiation of the branched myeloid lineages. CEBPB and CEBPD expressed in monocytes and macrophages, and CEBPE and CEBPA expressed in terminal granulocytes. CEBPs importance in myeloid differentiation was underlined upon induced expression in lymphocytes leading to forced differentiation into myeloid cells.

1.5.3 CEBPA

CEBPA is an intronless gene located on chromosome 19q13.1. It has two major protein isoforms, one 42kDa containing three transactivating domains, coding from the second AUG site in the gene (Nerlov and Ziff, 1995), and a 30kDa protein, with altered transactivating potential, due to a truncated transactivating domain coded from the third AUG site (Lin *et al.*, 1993; Ossipow *et al.*, 1993; Calkhoven *et al.*, 2000). More recently it has been discovered that there is a third

larger *CEBPA* isoform coding from a GUG site, this isoform functions specifically within the nucleolus to enhance ribosomal DNA transcription (Muller *et al.*, 2010).

1.5.3.1 Function

CEBPA is best known for its function as a myeloid differentiation gene, controlling haematopoietic progression from CMP to myeloblast cells. It also shows high levels of expression in liver, lung, and adipose cells and is involved in the proliferation and differentiation of these tissues, often acting in concert with other *CEBP* genes (Lekstrom-Himes and Xanthopoulos, 1998). The short isoform, p30, lacks TDI important for interaction with various transcriptional initiating factors and growth arrest through regulation of *MYC* (Johansen *et al.*, 2001). The second transactivating domain shared by the protein isoforms interacts with cell cycle genes *CDK2/4* and chromatin remodelling complexes *SWI/SNF* (Wang *et al.*, 2001) (Figure 1.11).

CEBPA directly activates lineage specific gene promoters, such as the *G-CSF* receptor, basal transcriptional apparatus, histone acetyl transferases, and chromatin remodelling complexes, *SWI/SNF*. It drives tissue specific differentiation through cooperation with specific collaborating transcription factors. In haematopoiesis these adaptors are thought to be the *GATA* factors (McNagny *et al.*, 1998). The role of *CEBPA* as a critical myeloid differentiation gene has been supported both by loss and gain of function experiments. Ectopic expression of *CEBPA* was shown to induce differentiation and loss of proliferation in myeloid cells (Nerlov, 2004; Johnson, 2005), while *CEBPA* knock out mice displayed high levels of undifferentiated myeloid precursor cells (Zhang *et al.*, 1997). While not involved in lymphoid differentiation, over-expression of *CEBPA* and *CEBPB* was observed to transform B-cells into myeloid precursor cells, through upregulation of *PU.1*, antagonism of *PAX5*, and use of endogenous *SPI1* (Heavey *et al.*, 2003; Xie *et al.*, 2004) (Figure 1.10). Expression of *CEBPA* and *CEBPB* also differentiates committed T-cell progenitors into inflammatory macrophages via *PU.1* and down regulation of T-cell lineage genes, such as *Notch-1* and *GATA3* (Hsu *et al.*, 2006). In hepatocytes, expression of *CEBPA* is critically important in glucose homeostasis, as shown by null *CEBPA* mice being unable to store hepatic glycogen, leading to death from hypoglycaemia shortly after birth (Wang *et al.*, 1995). *CEBPA* has also been shown to be modified post

transcriptionally via *FLT3* phosphorylation of serine 21, leading to loss of granulopoietic differentiation (Radomska *et al.*, 2006).

1.5.3.2 Cancer

CEBPA functions as an important driver of myeloid differentiation, giving it a natural role as a tumour suppressor in a haematopoietic setting. The importance of this role is highlighted by the presence of biallelic mutations in both sporadic and familial AML (Pabst *et al.*, 2001; Nerlov, 2004; Smith *et al.*, 2004), where they are associated with a good prognosis (Preudhomme *et al.*, 2002; Fröhling *et al.*, 2004). The majority of patients show single allele frameshift mutations preventing the expression of the full p42 isoform and leading to expression of the dominant negative p30 isoform (Nerlov, 2004). Other patients with biallelic mutations showed that the most common partner mutation of the *CEBPA* p42 frameshift mutation was amino acid insertions within the DNA binding domain. These insertions produce incorrect alignment of bound DNA, effectively leading to inactivation of the protein. These two separate mutations completely knock out *CEBPA* function (Asou *et al.*, 2003). *CEBPA* mutations are most commonly found in the minimally differentiated M1 and M2 FAB types, which support the role of *CEBPA* in differentiation from CMP to myeloblasts and later differentiation of granulocytes. Interestingly, gene expression analysis of patients with *CEBPA* mutations showed upregulation of erythroid differentiation genes and down regulation of myeloid commitment genes (Hackanson *et al.*, 2008). The tumour suppressor role of *CEBPA* is further supported by findings that *CEBPA* and *CEBPB* are both downregulated at the mRNA level in *BCR-ABL1* positive patients, and a number of other haematological malignancies (Guerzoni *et al.*, 2006).

1.5.4 CEBPB

CEBPB is an intronless gene located on chromosome 20q13.13, which is expressed as three protein isoforms whose interplay dictates the functional role of the gene. The largest isoform is known as liver enriched activating protein* (LAP*) of 44kDa, followed by a smaller isoform, LAP, of 42kDa and the smallest isoform, liver inhibitory protein (LIP) at 20kDa. Transactivational activity of

CEBPB is determined by the ratios of LAP*, LAP and LIP, with LIP acting as an inhibitor of the other two isoforms (Ramji and Foka, 2002; Zahnow, 2009; Tsukada *et al.*, 2011). LAP* is a less potent activator of gene activity than LAP. However LAP* has the ability to recruit more co-activators than LAP, due to its extended N-terminus, providing a broader range of functions, including interaction with *MYB* to induce myeloid differentiation (Kowenz-Leutz and Leutz, 1999).

1.5.4.1 Function

Strongly induced by bacterial lipopolysaccharides (LPS), *CEBPB* plays an important part in immune and inflammatory responses. It was identified as a mediator of *IL-6*, binding to its *IL-1* response element. It has also been found to bind to the regulatory regions of acute phase and cytokine genes such as *TNF*, *IL8* and *G-CSF* (Akira *et al.*, 1990; Poli *et al.*, 1990).

CEBPB plays an important role in haematopoietic differentiation, specifically in the myeloid lineage, where it is active in late macrophage cell differentiation. A high LAP/LIP ratio, indicating LAP* and LAP dominance are important in monocytic lineage commitment (Gutsch *et al.*, 2011), with increasing LAP/LIP ratios observed in cells committed to the myeloid and macrophage lineage. The same trend is observed in hepatocyte and adipocyte differentiation (Buck *et al.*, 1994). The importance of *CEBPB* in macrophage differentiation is observed in *CEBPB* knockout mice, where the resulting macrophages show reduced functional activity (Screpanti *et al.*, 1995). LAP and LAP* can also promote granulocyte differentiation in emergency conditions such as infections, where rapid increases of immune cells are required (Popernack *et al.*, 2001; Hirai *et al.*, 2006). Overexpression of the larger isoforms results in decreased myeloid progenitors and increased granulocytes in murine primary bone marrow cells, while LIP overexpression did not affect myeloid progenitor formation (Popernack *et al.*, 2001).

The *CEBPB* LAP/LIP ratio also dictates proliferation and cell cycle, which is controlled by *FLT3* through induced expression of the short LIP isoform, leading to inactivation of LAP* and LAP. Inactivation of the larger protein isoforms removes their inhibition of *MYC* and in turn promotes cell cycle progression.

(Gutsch *et al.*, 2011; Zhang *et al.*, 2011) (Figure 1.11). In mouse fibroblasts and in pre-monocytic cells, *CEBPB* was observed to associate with *RB-E2F* protein complex, preventing transition from G1 to S phase (Johnson, 2005; Sebastian *et al.*, 2005) (Figure 1.11). In later development, *CEBPB* loss enhances differentiation (Screpanti *et al.*, 1995; Gutsch *et al.*, 2011), leading to an increase in the number of cells in S phase (Gutsch *et al.*, 2011; Zhang *et al.*, 2011). Complete absence of *CEBPB* in murine monocytes affects cells at early and late stages of development. In early progenitors, prior to granulocytic commitment, cell proliferation is retarded thus lowering myeloid colony formation in methylcellulose tests (Hirai *et al.*, 2006).

1.5.4.2 Cancer

CEBPB is linked to leukaemia mainly through inactivation of the large isoform proteins by LIP, leading to deregulation of *CEBPB* cell cycle control (Figure 1.11). AML patients with *FLT3-ITD* have high LIP expression, leading to deregulated *CEBPB* function (Haas *et al.*, 2010). This trend is also observed in several *FLT3-ITD* positive monocytic cell lines (Wall *et al.*, 1996) and across several cancer types, including HeLa cells and breast cancer cells (Zahnow *et al.*, 1997; Zahnow *et al.*, 2001; Gomis *et al.*, 2006). Indeed the expression of larger LAP isoforms is generally reduced in highly proliferative cells, such as bone marrow derived CML cells from patients during blast crisis (Guerzoni *et al.*, 2006). The importance of the LAP/LIP ratio is underlined by upregulation of LAP/LAP* which is shown to reduce proliferation and induce differentiation in acute promyelocytic leukaemia (APL) (Duprez *et al.*, 2003) and in primary AML cells (Studzinski *et al.*, 2005), among other examples.

1.5.5 CEBPD

Located on chromosome 8q11.21, the *CEBPD* gene is coded by a single exon and is expressed as a single isoform, coding for a 28kDa protein which readily forms heterodimers with *CEBPA* and *CEBPB* (Tsukada *et al.*, 2011). This gene is an important mediator within inflammatory pathways, as well as adipocyte and granulocyte differentiation. Although important in multiple pathways, findings from *CEBPD* knock out mice has shown that the gene is not crucial for normal

development (Tanaka *et al.*, 1997). This view is supported by the generally low and specific expression of the gene, as well as the overlap of function with CEBP proteins. These observations and the rapid upregulation of *CEBPD* in response to specific stimuli, such as inflammation and corticosteroids, has led to the theory that *CEBPD* functions as a modulator for physiological adaptations (Balamurugan and Sterneck, 2013).

1.5.5.1 Function

CEBPD has been shown to operate as both an activator and repressor of multiple genes, depending upon the context and cell of expression. One of its first observed functions was up-regulation, due to bacterial LPS and pro-inflammatory cytokine stimulation, leading to its discovery as an activator of inflammatory pathways (Ramji and Foka, 2002). Further research into the inflammatory pathways has shown *CEBPD* to act as both a mediator of chronic inflammation and of macrophage function (Wu *et al.*, 2011; Ko *et al.*, 2012).

CEBPD exhibits the same cell cycle control mechanisms as *CEBPA* and *CEBPB*, although it upregulates both *CDKN1B* and *CDKN1C* (Figure 1.11) (O'Rourke *et al.*, 1999; Barbaro *et al.*, 2007; Pawar *et al.*, 2010). It promotes growth control by influencing multiple targets in G0/G1 of the cell cycle and later in S phase. *CEBPD* also drives maturation in both macrophage differentiation and function (Litvak *et al.*, 2009), as well as granulocyte differentiation (Gery *et al.*, 2005). *CEBPD* also promotes cell death through induction of *STAT3*, leading to the down regulation of *CCND1*, upregulation of *IGFBP5* and *P53*, and regulation of the *BCL* family (Thangaraju *et al.*, 2005).

As with other *CEBP* family members, *CEBPD* acts primarily as a transcriptional regulator, however it has also been shown to function as a chaperone protein, guiding the DNA repair gene, *FANCD2,* into the nucleus, which is a unique function for this family of genes (Wang *et al.*, 2010).

Cancer

In cancer, *CEBPD* acts as both a tumour promoter and suppressor. Cell cycle control through *CEBPD* has been observed in many cancer cell lines through

multiple interactions with the *RB*/*E2F* pathway (Figure 1.11). *CEBPD* binds several targets during early cell cycle progression, it has been shown to mitigate the proliferative action of *E2F1* in breast cancer (Pan *et al.*, 2010), and in limbal stem cells. In limbal cells, *CEBPD* mediated downregulation of *E2F1* and upregulation of *CDKN1B* and *CDKN1C* was considered to encourage stemness as cells were halted at the G0 phase, prolonging their lifespan (Barbaro *et al.*, 2007). Conversely in breast cancer cell lines, *CEBPD* is down regulated by the ubiquitin ligase, *SIAH2,* resulting in deregulation of cell cycle control and increased proliferation (Sarkar *et al.*, 2012). Expression of *CEBPD* resulted in the growth arrest and subsequent differentiation of *CEBPD* deficient CML cell lines, KCL22 and K562. Growth arrest was achieved through interaction with RB and E2F1 proteins, down regulation of *MYC* and *CCNE1* and upregulation of *CDKN1B*. Subsequent differentiation occurred through upregulation of *G-CSFR* and collagenase (Gery *et al.*, 2005). This differentiation capability was also observed during overexpression of *CEBPD* in mouse progenitor cells, resulting in AML and differentiation of acute myelogenous leukaemia cell line models down the myeloid pathway (Gery *et al.*, 2005). Supporting the function of *CEBPD* in cell cycle control and differentiation in AML, is the finding that the *CEBPD* promoter is silenced in 35% (28/80) AML patient samples (Agrawal *et al.*, 2007), with reduced protein levels observed in a number of other cancers (Radich *et al.*, 2006). However in prostate cancer cell lines, results are inconsistent, with the androgen dependent LNCAP cell line exhibiting growth inhibition, while the androgen dependent prostate cancer cell line, CWR22, remains unaffected (Ikezoe *et al.*, 2005; Sanford and DeWille, 2005).

These data suggest that the main function of *CEBPD* in tumour suppression is through cell cycle control. As a tumour promoter, *CEBPD* has been shown to upregulate levels of *SOD-1*, a reactive oxygen species scavenger, conferring drug resistance to urothelial carcinoma cell lines (Hour *et al.*, 2010). In a recent breast cancer study, it was observed as both a promoter and a suppressor, protecting from the initiation of oncogenesis, while promoting metastasis (Balamurugan *et al.*, 2010).

Interestingly, *CEBPD* is located within a fragile site, which is often lost in cancer cell lines due to proximity to the *KIAA0146* gene. However, it is unclear whether this loss directly influences oncogenesis (Brueckner *et al.*, 2013). Sequencing of

CEBPD and *CEBPB* in haematological malignancies found little evidence that mutations in these genes contributed to the development of oncogenesis (Vegesna *et al.*, 2002), unlike *CEBPA*. Although this observation did not include the effects of expression change through gene amplification or copy number loss. The strongest evidence of *CEBPD* as a tumour promoter is found in brain tumours. *CEBPD* mRNA is overexpressed in mesenchymal glioblastoma cells and is linked to a poor outcome (Cooper *et al.*, 2012). The action of *CEBPD* in promoting cell survival in these conditions is through upregulation of *HIF-1α*, leading to improved tumour cell survival in hypoxic conditions. Additionally *CEBPD* inhibits expression of *FBXW7α* in glioblastoma cell lines (Balamurugan *et al.*, 2010), which is a subunit of the SCF E3 ubiquitin ligase complex and a confirmed tumour suppressor in glioblastoma (Hagedorn *et al.*, 2007; Cheng and Li, 2012). Despite this action, there is evidence to suggest that *CEBPD* actively recruits tumour associated macrophages to the hypoxic rich regions in murine tumours (Balamurugan and Sterneck, 2013). Simplified, the two main actions of *CEBPD* as a tumour promoter are through enhanced hypoxic survival in tumour cells, and increased inflammatory signalling, leading to increased DNA damage (Trinchieri, 2012).

1.5.6 CEBPE

Coded by a two exon gene, *CEBPE* is located on chromosome 14q11.2 and expressed as four different protein isoforms; p32, p30, p27, and p14 , with variable transactivating domains and differing transcriptional activities (Chumakov *et al.*, 1997; Yamanaka *et al.*, 1997) (Figure 1.11). *CEBPE* is expressed primarily in myeloid and lymphoid cells (Yamanaka *et al.*, 1997), the highest expression is seen in pro-myelocytes and late myeloblasts (Thomassin *et al.*, 1992; Antonson *et al.*, 1996; Morosetti *et al.*, 1997). The gene and isoforms are weak transcriptional activators, which require coactivators for full functional activity (Yamanaka *et al.*, 1997).

1.5.6.1 Function and Cancer

CEBPE is involved in terminal granulocyte differentiation. This function is supported by the observations in *CEBPE* mutant mice, which die after 2-5 months

due to the build-up of immature granulocytes (Cohen *et al.*, 2008). Research has also shown that lack of *CEBPE* results in neutrophil specific granule deficiency (Lekstrom-Himes and Xanthopoulos, 1998; Gombart *et al.*, 2001), due to its importance in the expression of genes coding for proteins stored in specific neutrophil granules. All four *CEBPE* isoforms are expressed in mature neutrophils and eosinophils, but isoform expression varies between different neutrophil and eosinophil progenitors. The p32/30 isoforms were found to promote eosinophil differentiation, while the smaller p27 isoform inhibited eosinophil differentiation through *GATA-1* inactivation, resulting in neutrophil differentiation. The smallest p14 repressor isoform inhibited eosinophil differentiation, while pushing erythroid commitment (Bedi *et al.*, 2009). *CEBPE* also plays a role in early myeloid differentiation, inducing cell cycle arrest at the myelocytic stage (Nakajima *et al.*, 2006). Like the other *CEBPs* as discussed above, *CEBPE* interacts with *E2F1*, *RB1* and *MYC* to exert cell cycle control (Gery *et al.*, 2004). Genome wide association studies have found *CEBPE* Single Nucleotide Polymorphisms (SNPs) to be among a group of genes, which significantly increase the likelihood of developing paediatric ALL (Papaemmanuil *et al.*, 2009).

1.5.7 CEBPG

CEBPG is a ubiquitously expressed gene initially discovered by its affinity for *cis*-regulatory sites in the immunoglobulin heavy chain promoter and enhancer (Roman *et al.*, 1990), which shows highest expression in progenitor cells (Thomassin *et al.*, 1992; Cooper *et al.*, 1995). Located on chromosome 19q13.11, the two exon gene codes for a single 16.4kDa protein isoform, which does not possess a transcriptional transactivating domain, leading to the theory that the protein acts as a dominant negative regulatory factor for the other CEBP proteins.

1.5.7.1 Function and Cancer

While functioning as the main *CEBP* regulator, *CEBPG* itself is regulated by *CEBPA*. *CEBPA* has been shown to bind to the proximal promoter region of *CEBPG*, repressing expression of the gene (Alberich-Jorda *et al.*, 2012). *CEBPG* shows high expression in lymphoid and myeloid precursors, becoming down

regulated as cells mature (Cooper *et al.*, 1995). In a murine setting, *CEBPG* controls the function of *CEBPB* through heterodimerisation, leading to inactivation of pro-inflammatory cytokines and chemokines, and cell senescence in mouse embryonic fibroblasts (Huggins *et al.*, 2013). *CEBPG* has been shown to inactivate other CEBPs through heterodimerisation in a cell specific manner, successfully inactivating the function of *CEBPB* and *CEBPD* in fibroblastic cells, while failing to do so in HepG2 hepatoma cells. It also completely failed to inactivate the function of *CEBPA* in either cell type (Parkin *et al.*, 2002). *CEBPG* functions as an inhibitor of neutrophil differentiation through blockage of *G-CSF*, and theoretically *CEBPB*, as demonstrated *in vitro*, *in vivo*, and in a specific subset of *CEBPA* downregulated AML patients. However subsequent downregulation of *CEBPG in vitro* resulted in granulocytic differentiation of these primary AML cells (Alberich-Jorda *et al.*, 2012). Depletion of *CEBPG* has been shown to negatively impact A549 lung adenocarcinoma cells. Expression of *CEBPG* mRNA has also been shown to have a statistically significant association with patient outcome in clinical cancer studies, with higher expression linked to poorer outcomes (Huggins *et al.*, 2013). In cancer, the main role of *CEBPG* appears to be the inactivation of other *CEBP* members, resulting in deregulation of cell cycle control and a block in cellular differentiation.

1.5.8 CEBPZ

Also known as *CHOP* and *DDIT3*, *CEBPZ* is the only *CEBP* gene not identified as that a translocation partner of *IGH*. Containing three exons, the gene is located at chromosome 2q13.32 and is a ubiquitously expressed upon DNA damage (Fornace *et al.*, 1989). It readily dimerizes with the other *CEBPs*, but cannot bind to the cognate DNA enhancer element due to mutations in the DNA binding domain (Lekstrom-Himes and Xanthopoulos, 1998). Aberrant methylation of the *CEBPZ* promoter has been described in AML, suggesting a functional role in the development of the leukaemia (Yao *et al.*, 2011). There are suggestions that *CEBPZ* acts as a negative regulator of stress induced genes (Ron and Habener, 1992).

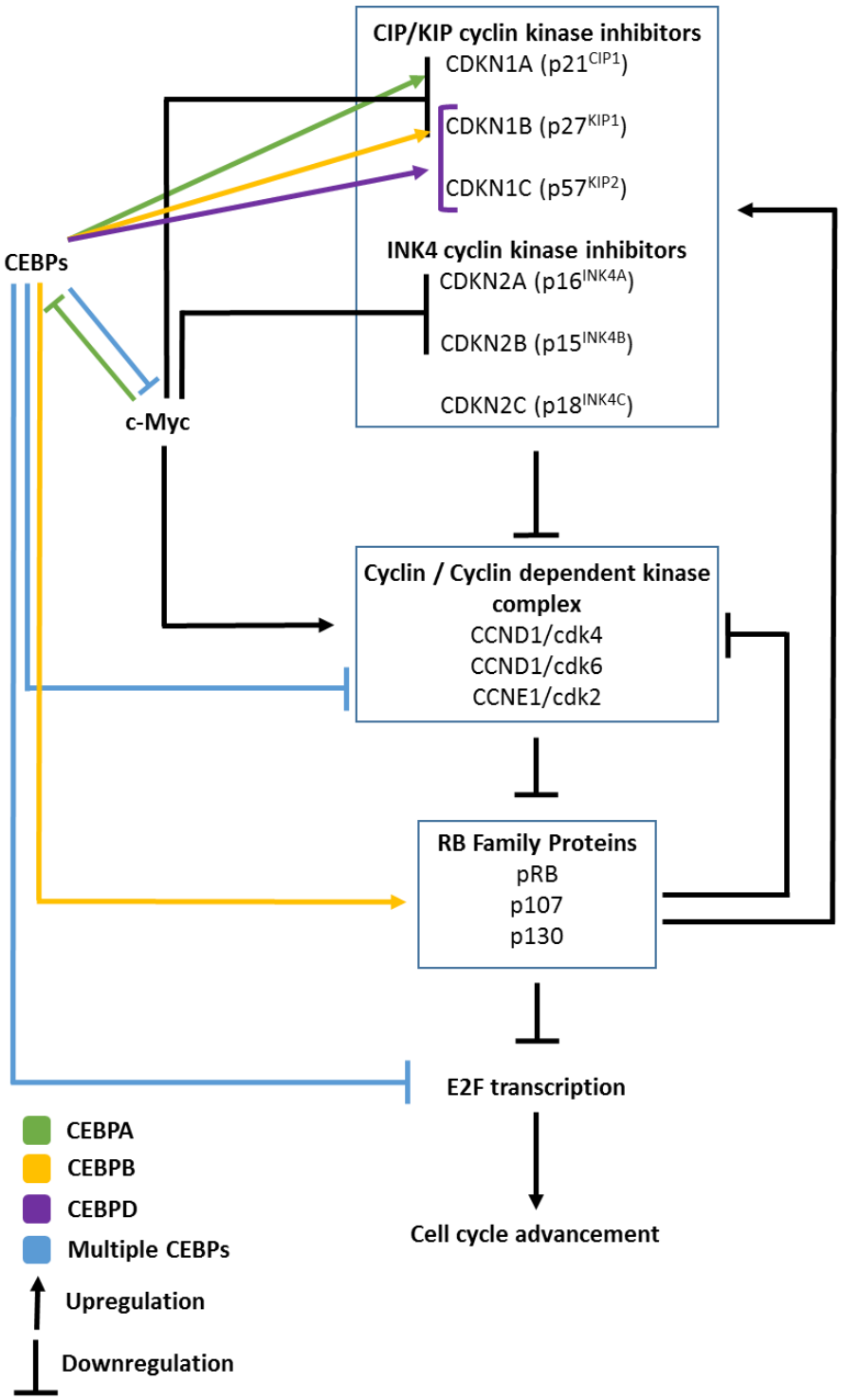


Figure 1.11. CEBP influence on cell cycle control at the G0/1 and S phase.

The capacity to exert cell cycle control is a unifying ability in CEBP genes containing TDs. While some details may differ, broadly CEBPA, CEBPB, CEBPD and CEBPE have been observed to control G0 to G1 progression in several cell types. CEBPs exert this function by controlling the interplay between cyclins, cylin dependent kinases (cdks), cylin inhibitors, the RB protein family and the E2F transcription factor family. The cyclin inhibitors upregulated are dependent upon the individual CEBP gene, CEBPA upregulates CDKN1A (p21CIP1) (Wang et al., 2001), CEBPB upregulates CDKN1B (p27KIP1) (Gutsch et al., 2011) and CEBPD upregulates both CDKN1B (p27KIP1) and CDKN1C (p57KIP2) (O'Rourke et al., 1999). These individual cyclin kinase inhibit all cyclins at the G1 stage. These CEBPs also directly inhibit cyclin function, repressing CCND1/cdk4, CCND1/cdk6 and CCNE1/cdk2. All of the above process function to prevent the phosphorylation of the RB protein family (Johnson, 2005). This is vital as this family binds to E2F transcription factors, repressing the transcription of S phase genes which drive the cell cycle forwards (Classon and Harlow, 2002). The E2F

transcription factors are also affected by CEBP genes through indirect binding which inhibits E2Fs interaction with target genes. CEBP genes also exert an effect later in S phase, where they disrupt E2F-S phase gene complexes, promoting cell cycle withdrawal, and function to stabilise pRB-E2F complexes in combination with the SWI/SNF chromatin remodelling complex (Muchardt and Yaniv, 2001). Indirectly CEBPA, CEBPB, CEBPD and CEBPE downregulate MYC. Nullifying the effects of MYC which acts to drive cell cycle progression through down regulation of cyclin inhibitors and upregulation of cyclin/cdk complexes, with CEBPA shown to have a reciprocal relationship, balancing proliferation with growth arrest (Freytag and Geddes, 1992).

Aims

Detailed characterisation of multiple ALL subgroups has resulted in improved survival as a result of altered treatment stratification and/or the identification of novel therapeutic targets. This PhD aimed to characterise the *IGH-CEBP* subgroup, to identify common clinical characteristics, established and novel genetic lesions, and affected molecular pathways to determine the role of *CEBP* deregulation in BCP-ALL.

This would be achieved by:

- Identifying an *IGH-CEBP* patient cohort through use of FISH screening
- Performing statistical analysis on patient clinical data to identify potential trends
- Genetically characterising the *IGH-CEBP* cohort using Multiplex Ligation-Dependent Probe Amplification (MLPA), SNP and FISH techniques.
- Using functional techniques to model *CEBP* deregulation *in vitro* and *in vivo*.
- Using molecular characterisation techniques on primary patient material and generated functional model material to identify molecular pathways important in *IGH-CEBP* BCP-ALL.

Chapter 2 Materials and Methods

2.1 Approval

2.1.1 Ethical Approval for patient material

Informed consent was given by all patients in accordance with the Declaration of Helsinki and ethics committee approval.

2.1.2 Ethical Approval for mouse studies

Mouse studies were conducted in accordance with the UK Animals (Scientific Procedures) Act (ASPA) 1986 under the project licence PPL60/4552.

2.2 Materials

2.2.1 List of Suppliers

Company	**Location**
Abbott Molecular	Chicago, IL, USA
Affymetrix	Santa Clara, CA, USA
Agilent Technologies	Santa Clara, CA, USA
Beckman Coulter	Brea, CA, USA
BGM Labtech	Ortenberg, Germany
Bio-Rad	Hercules, CA, USA
Clontech	Mountain View, CA, USA
Cytocell	Cambridge, UK
Eppendorf	Hamburg, Germany
Erba Diagnostics	Miami Lakes, FL, USA
Fisher Scientific	Loughborough, UK
GE Healthcare	Little Chalfont, UK
Labtech International	Uckfield, UK
Leica Biosystems	Nussloch, Germany
Life Tech (Applied Biosystems, Invitrogen and Gibco)	Carlsbad, CA, USA
Macherey-Nagel	Düren, Germany
MRC Holland	Amsterdam, Netherlands
Olympus	Tokyo, Japan
Promega	Madison, WI, USA
Qiagen	Venlo, Netherlands
Santa Cruz Biotechnologies	Dallas, TX, USA
SANYO	Osaka, Japan
Sigma-Aldrich	St. Louis, MO, USA
Syngene	Cambridge, UK
Takara	Mountain View, CA, USA
Thermo Scientific	Waltham, MA, USA
VWR International	Radnor, PA, USA

2.2.2 Laboratory Equipment

Equipment	**Company**
5415R Microcentrifuge	Eppendorf
2100 BioAnalyzer	Agilent Technologies
2720 Thermal Cycler	Applied Biosystems
Affymetrix GeneChip SNP 6.0 array	Affymetrix
Amersham ECL Gel Box	GE Life Sciences
Beckman Coulter CEQ 8800	Beckman Coulter
BX61 Fluorescent Microscope	Olympus
CO_2 Incubator	SANYO
Digital Heating Ceramic Plate	Velp Scientifica
FACS Canto II	Becton Dickinson
FACS Fusion Sorter	Becton Dickinson
Fluostar Omega Plate Reader	BMG Labtech
G:BOX Chemi XL1.4 Imaging System	Syngene
Gel Electrophoresis Tanks	Bio-Rad
GeneAmp PCR System 2700	Applied Biosystems
HemaVet	Erba Diagnostics
Hybrite	Vysis
IKA vortexer	Agilent Technologies
LSRFortessa X-20	Becton Dickinson
NanoDrop 1000	Thermo Scientific
Qubit 2.0 Fluorometer	Life Technologies
Rocking Platform	Bibby Scientific
SRT9 Roller Mixer	Stuart
Superior Haemocytometer	Marienfeld
Thermobrite Statspin	Abbott Molecular
ViiA 7 Real-Time PCR System	Applied Biosystems
ViiA7 Real-Time PCR System	Life Technologies

2.2.3 Software

Software	Company
CytoVision 7.1	Leica Biosystems
FACSDiva	Becton Dickinson
FlowJo	Flow Jo
GeneMarker V1.85	SoftGenetics
Genotyping Console V4.1.1.834	Affymetrix
ImageJ software (Version 1.48)	Freeware
New England Biolabs Cutter V2.0	New England Biolab
Primer 3Plus	General Public License
ViiA 7 Software 1.1	Life Technologies

2.2.4 Chemicals and Reagents

2.2.4.1 Chemicals and Reagents

Chemical / Reagent	Company
16% Formaldehyde solution	Thermo Scientific
1kb Ladder	Promega
Amersham ECL Gel	GE Life Sciences
Ampicillin	Sigma-Aldrich
Aqua dUTP 431nm (Red)	Abbot Molecular
Busulfan	Sigma-Aldrich
cDNA reagents	Promega
Chloramphenicol	Sigma-Aldrich
DAPI Solution	Vector
DMSO	Fisher Scientific
dNTP Master Mix	Enzo
Ethanol	Fisher Scientific
Ficoll-Paque PREMIUM	GE Healthcare
Gel Red	VWR International
Gold dUTP	Abbot Molecular
Green dUTP 496nm	Abbot Molecular
Human Cot-1 DNA	Invitrogen
Hybridisation Buffer	Cytocell

Hybridisation Buffer	Cytocell
Isopropanol	Fisher Scientific
Kanamycin	Sigma-Aldrich
Loading Dye (6x)	Promega
Methanol	Fisher Scientific
Mineral Oil	Beckman Coulter
PCR Markers	Promega
Phosphate buffered saline (PBS)	GIBCO
Phosphate buffered saline tablets	Sigma-Aldrich
Polybrene	Sigma-Aldrich
Polybrene	Sigma-Aldrich
Puromycin	Promega
PVDF membrane	Merck Millipore
Red dUTP 580nm	Abbot Molecular
Sodium Acetate	Sigma-Aldrich
Spectra Multicolor Broad Range Protein Ladder	Thermo Scientific
Taqman Universal Mastermix II	Applied Biosystems
Triton X-100	Sigma-Aldrich
Trypan blue	Bio-Rad
Ficoll-Paque PREMIUM	GE Healthcare
Restore™ Western Blot Stripping Buffer	Thermo Scientific

2.2.4.2 Experimental Kits

Experimental Kit	**Company**
Amersham Enhanced Chemiluminescence Prime Western Blotting Detection Reagent	GE Healthcare
Agilent RNA 6000 Nano Kit	Agilent
Anti-Mouse Ig, κ/Negative Control Compensation Particles Set	BD Bioscience
CD34 MicroBead Kit	Miltenyi Biotec
DNeasy Blood and Tissue Kit	Qiagen

EndoFree Plasmid Maxi Kit	Qiagen
GeneJET Gel Extraction Kit	Thermo Scientific
Illustra ProbeQuant G-50 Micro Column	GE Healthcare
Illustra ProbeQuant G-50 Micro Columns	GE Healthcare Life Sciences
Nick Translation DNA Labelling Kit	Enzo
NucleoBond Midi Kit	Macherey-Nagel
One Shot® Stbl3 Chemically Competent E. coli	Life Technologies
Pierce™ Coomassie (Bradford) Protein Assay Kit	Thermo Scientific
QIAPrep Spin Mini Kit	Qiagen
QIAshredder	Qiagen
RNeasy Micro Kit	Qiagen
RNeasy Mini Kit	Qiagen
SALSA MLPA P335 ALL-IKZF1 probemix	MRC Holland

2.2.4.3 Culture Media and Supplements

Culture Media and Supplements	**Company**
Foetal Bovine Serum	Gibco
Foetal Bovine Serum	Sigma-Aldrich
L-Glutamine	Sigma-Aldrich
Trypsin	Sigma-Aldrich
Cell Disassociation Buffer	Gibco
Dulbeco's Modified Eagle's Medium	Sigma-Aldrich
Iscove's Modified Dulbeco's Medium	Gibco
Minimum Essential Medium Eagle Alpha Modification Medium	Gibco / Sigma
Gelatin Solution	Sigma-Aldrich

2.2.4.4 Culture Media Cytokines

Cytokine	**Company**	**Product Code**
Recombinant human Flt-3 Ligand	Miltenyi	130-093-854
Recombinant human stem cell factor	Miltenyi	130-093-991
Recombinant human thrombopoietin	Miltenyi	130-094-011
Recombinant human IL-3	Miltenyi	130-093-908

Recombinant human IL-6	Miltenyi	130-095-365
Recombinant human IL-7	Life Technologies	PHC0075

2.2.5 Buffers and Media

2.2.5.1 Fluorescence in situ Hybridisation Buffers

Name	**Composition**
WASH 1 (1L)	20ml 20X SSC
	3ml NP40
WASH 2 (1L)	100ml 20X SSC
	1ml igepal-CA-630

2.2.5.2 Western Immunoblotting Buffers

2.2.5.2.1 Western Immunoblotting Transfer Buffers

Name	**Composition**
10x Running Buffer (1L)	30.3g Tris
	144g Glycine
	10g SDS
10x Transfer Buffer (1L)	144g Glycine
	30.3g Tris
Urea Buffer	7.92M Urea
	100mM NaH_2PO_4
	80mM Tris-HCl (pH 8)

2.2.5.2.2 Western Immunoblotting Direct Lysis Buffer

Reagents	**Volume**
2X Laemmli Sample Buffer:	700µl
β-Mercaptoethanol:	70µl
1X PBS	416µl
100X Na3VO4	14µl
7X protease inhibitor	200µl

2.2.5.3 Freezing Buffers

Name	Composition
Newcastle Freezing Solution	90% FBS
	10% DMSO
Cincinnati Freezing Solution 1	HES 500µl
	IMDM 300µl
	BSA (25%) 200µl
Cincinnati Freezing Solution 2	DMSO 100µl
	HES 500µl
	IMDM 300µl
	BSA (25%) 200µl

2.2.5.4 Bacterial Culture Buffers

Name	Composition
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

2.2.5.5 Cell Culture Media

Cells	Medium	Supplements	Cytokines (10ng/µl)
293T cells	Dulbecco's Modified Medium	10% FBS + 2mM L-glutamine	NA
CD34+ cells	Iscove's Modified DMEM	10% FBS + 2mM L-glutamine	SCF, TPO, FLT3-L, IL-3, IL-6, IL-7
CD34+ cells	Alpha-MEM	20% FBS + 2mM L-glutamine	SCF, FLT3-L, IL-7

MS-5 Cells, OP9 Cells	Alpha-MEM	20% FBS + 2mM L-glutamine	NA

2.2.5.6 Retroviral Transfection Buffers

Name	Composition
2M CaCl2 Solution (50ml)	14.701g CaCl2
	50ml H_2O
	0.2µm filter
HBSS x2 Solution (50ml)	43.822ml H_2O
	2.778ml 0.9M HEPES (pH 7.1) for a total conc of 50mM
	2.8ml 5M NaCl for a total conc of 280mM
	0.6ml 125mM Na2HPO4.7H_2O for a total conc of 75mM
	0.2µm filter

2.2.5.7 Xenograft Buffers

Name	Composition
Ammonium chloride red lysis buffer 10X	8.3g Ammonium Chloride
	1.0g Potassium Bicarbonate
	0.38g EDTA
	100ml dH2O

2.2.6 Oligonucleotide Sequences

2.2.6.1 Quantitative PCR Primers TaqMan

Gene	Exon Boundary	Assay Location	Amplicon Length	Life Tech ID
CEBPA	1 - 1	1538	77	Hs00269972_s1
CEBPB	1 - 1	838	75	hs00270923_s1
CEBPD	1 - 1	943	107	Hs00270931_s1
CEBPE	1 - 2	1030	121	Hs00357657_m1

CEBPG	2 - 2	3463	134	Hs01922818_s1
IKZF1	2 - 3	209	73	Hs00958474_m1
GAPDH	3 - 3	229	122	Hs99999905_m1
GUSB	11 - 12	1925	81	Hs99999908_m1
B2M	3 - 4	431	81	Hs00984230_m1
ABL1	3 - 4	549	91	Hs00245445_m1

2.2.6.2 Quantitative PCR Primers SYBR Green

Primer Name	Position	Forward Sequence	Reverse Sequence
ABL2 Set 1	Exon 2	TTTGAATGCCATGAAAAGGA	TCCATTCCCTGTTCTCCATC
ABL2 Set 2	Exon 2	ATCACTTTGCCAGCTGTGTG	AACCCTTGAATTTGTGGTTCC
ABL2 Set 1	Exon 3	CTTTGCATCGTCCCTATGGT	TGAGTGTGTTATCACCACTTGCT
ABL2 Set 2	Exon 3	AGCTTTGCATCGTCCCTATG	CTGAGTGTGTTATCACCACTTGC
ABL2 Set 1	Exon 4	TGCCAAGCAACTACATCACC	ACACGTCCCTCGTACCTGAG
ABL2 Set 2	Exon 4	GGTGCCAAGCAACTACATCA	GCCATCTGCAGTGGTATTGA
ABL2 Set 1	Exon 5	TTGGCAGAGCTTGTACACCA	GACGCCAACGTAAACCTCTC
ABL2 Set 2	Exon 5	TGGGCTGGTGACAACATTAC	GACGCCAACGTAAACCTCTC
ABL2 Set 1	Exon 7	TGTGTACTTTGGAGCCACCAT	AGTACTCCATTGCAGAAGAAATCTG
ABL2 Set 2	Exon 7	TTTGGAGCCACCATTTTACA	TTCTCTAAGTACTCCATTGCAGAAGA
ABL2 Set 1	Exon 9	TGGGGTATTGTTGTGGGAAA	CTCAGGCTGTTCCATTCGAT
ABL2 Set 2	Exon 11	TAGCTGAGGAGCTTGGGAGA	CTGGTGCTAAACTGGAAGCA
B2M	Intron 1-2	TCTAGGCGCCCGCTAAGTT	TCGCGTGCTGTTTCCTCC
RPLPO	Intron 2-3	ATAAACGGGCTCAGGCAAGTT	CGCGCTCTTTTAGAAGCCAG
TBP	Intron 5-6	TCTCTCTGACCATTGTAGCGGTT	CCGTGGTTCGTGGCTCTCT

2.2.7 Cloning Restriction Enzymes

Restriction Enzyme	Sequence	Promega Buffer	NEB Buffer	Incubation Temp	Heat Inactivation Temp
PvuII	CAG/CTG	B	3.1	37°C	65°C
EcoRI	G/AATTC	H	2.1	37°C	65°C
NcoI	C/CATGG	D	3.1	37°C	80°C
XhoI	C/TCGAG	D	CutSmart Buffer	37°C	65°C
BglII	A/GATCT	D	3.1	37°C	65°C
NotI	GC/GGCCGC	D	3.1	37°C	65°C

2.2.8 Mammalian Cell Lines

Cell Line	Origin
MS-5	Murine bone marrow cultures
OP-9	Murine bone marrow cultures
293T	Human embryonic kidney cells

2.2.9 Bacterial Strains

Name	Source
One Shot® Stbl3™ Chemically Competent E. coli	Life Technologies
E.coli	Bac Pac Resources
E.coli	Roswell Park

2.2.10 Antibodies

2.2.10.1 Flow Cytometry Antibodies

Marker	Fluorochrome	Clone	Laser		
			FACSCanto II	Fortessa X20	FACSAria II
CD34	Allophycocyanin (APC)	561	Red 663nm 660/20	Red 640nm 670/30	NA
CD19	Phycoerythrin (PE)	HIB19	Blue 488nm 585/42	Yellow/Green 561nm 586/15	Blue 488nm 575/26
CD19	Pacific Blue (PB)	HIB19	Violet 405nm 450/50	Violet 405nm 450/50	NA
CD19	Allophycocyanin (APC)	SJ25C1	Red 663nm 660/20	Red 640nm 670/30	NA
CD33	Phycoerythrin-Vio770 (PE-Vio770)	AC104.3E3	Blue 488nm 780/60	Yellow/Green 561nm 780/60	Blue 488nm 780/60
CD33	Pacific Blue (PB)	WM53	Violet 405nm 450/50	Violet 405nm 450/50	NA
CD33	Allophycocyanin (APC)	P67.6	Red 663nm 660/20	Red 640nm 670/30	NA
CD33	Phycoerythrin (PE)	P67.6	Blue 488nm 585/42	Yellow/Green 561nm 586/15	Blue 488nm 575/26
CD10	Phycoerythrin-Cyanine7 (PE-Cy7)	HI10a	Blue 488nm 780/60	Yellow/Green 561nm 780/60	Blue 488nm 780/60
CD10	Allophycocyanin-Cyanine7 (APC-Cy7)	HI10a	Red 663nm 780/60	Red 640nm 780/60	NA
CD11b	Pacific Blue (PB)	ICRF44	Violet 405nm 450/50	Violet 405nm 450/50	NA
hCD45	Peridinin Chlorophyll Protein Complex -Cyanine 5.5 (PerCP-Cy5.5)	HI30	Blue 488nm 670LP	Blue 488nm 710/50	Blue 488nm 710/50
hCD45	Phycoerythrin-Cyanine7 (PE-Cy7)	HI30	Blue 488nm 780/60	Yellow/Green 561nm 780/60	Blue 488nm 780/60

mCD45	Allophycocyanin-Cyanine7 (APC-Cy7)	30-F11	Red 663nm 780/60	Red 640nm 780/60	NA
Death Marker	Zombie Aqua	NA	Violet 405nm 525/50	Violet 405nm525/50	NA
7AAD	Far-Red	NA	Red 663nm 660/620	Red 663nm 670/14	NA
DAPI	NA	NA	Violet 405nm 450/50	UV 355nm 450/50	UV 355nm 450/50
EGFP	FITC	NA	Blue 488nm 530/30	Blue 488nm 530/30	Blue 488nm 520/20
Biotin (CD90)	NA	5E10	NA	NA	NA
Thy1	Phycoerythrin (PE) Sreptavidin	NA	Blue 488nm 585/42	Yellow/Green 561nm 586/15	Blue 488nm 575/26

Flow cytometry antibody markers, with corresponding emissions and the lasers used to detect them on the relevant flow cytometry analyser.

2.2.10.2 Flow Cytometry Isotype Controls

Marker	**Fluorochrome**	**Clone**
IgG1, κ Isotype Ctrl (FC) Antibody	APC	MOPC-173
IgG1, κ Isotype Ctrl (FC) Antibody	PE	MOPC-21
IgG1, κ Isotype Ctrl (FC) Antibody	APC-Cy7	MOPC-21
IgG1, κ Isotype Ctrl (FC) Antibody	PB	MOPC-21
Mouse IgG1	PE-Vio770	IS5-21F5

2.2.10.3 Primary Antibodies for Western Immunoblotting

Marker	**Species**	**Dilution**	**Company**
CEBPD	Rabbit	1:200	Santa Cruz
Ikaros	Goat	1:200	Santa Cruz
CEBPB	Rabbit	1:200	Santa Cruz
α-tubulin	Mouse	1:1000	Cell Signalling Technology
β-actin		1:2000	Abcam

2.2.10.4 Secondary Antibodies for Western Immunoblotting

Marker	**Conjugate**	**Dilution**	**Company**
Anti-rabbit	Horseradish Peroxidase (HRP)	1:2000	Santa Cruz

Anti-goat	Horseradish Peroxidase (HRP)	1:1000	Santa Cruz
Anti-mouse	Horseradish Peroxidase (HRP)	1:1000	Cell Signalling Technology

2.3 Methods

2.3.1 Genetic Characterisation Techniques

Genetic characterisation techniques were used to identify common and novel copy number alterations (CNAs) and chromosomal structural abnormalities in the *IGH-CEBP* cohort. *IGH-CEBP* translocations and CNAs were identified using fluorescence *in situ* hybridisation (FISH). Multiplex ligation-dependent probe amplification (MLPA) was used to investigate CNAs of a panel of preselected genes commonly abnormal in BCP-ALL, and SNP arrays were used to identify novel CNAs in the cohort.

2.3.1.1 Fluorescence *In Situ* Hybridisation (FISH)

Fluorescence in situ hybridisation (FISH) is a technique, which makes use of DNA segments bound to fluorescent reporter molecules to visualise gains, deletions, or chromosomal rearrangements through complementary DNA binding in targeted nuclei. The technique is highly customisable, allowing for use of multiple fluorochromes and probes of varying sizes. In this study, FISH was used to identify translocations, fusions, deletions, and gains of multiple target genes.

2.3.1.1.1 FISH Probes

Multiple probe sets were created to investigate CEBP gene rearrangements. Several probes were previously created by Dr. L.J. Russell (Akasaka *et al.*, 2007), over the course of this study several probes were altered to boost signal intensity, an IGH-CEBPD fusion probe was also created to confirm fusion events (Table 2.1). The DNA clones were selected using the UCSC (GRCh38/hg38) and Ensembl (Version 82) genome browsers. Clone sizes varied depending upon library coverage and gene location, larger bacterial

artificial chromosome (BAC) probes were between 150-350 kb in length and were ordered from Roswell Park (Buffalo, USA). Smaller fosmid clones, up to 40kb in length were ordered from Bac Pac Resources (Oakland, USA). Individual clone libraries were treated with specific antibiotics, RP-6, 21, 31, 91 and 92 libraries were grown on agar and in medium containing 50µg/ml Kanamycin, all other libraries were grown with 25µg/ml Chloramphenicol. Clones were ordered as agar stabs and grown on LB Agar plates with appropriate selection antibiotic at 37°C for 16 hours. Individual colonies were selected after incubating, and grown in a starter culture of 5mls of LB Broth and selection antibiotic for up to 8 hours in a shaking incubator at 37°C. 200ul of the starter culture was added to 200ml of LB Broth with the appropriate selection antibiotic. DNA was extracted using the NucleoBond Midi kit (Macherey-Nagel, Germany) following the manufacturer's instructions (http://www.mn-net.com/tabid/1479/default.aspx). The bacterial culture was decanted into 50ml falcon tubes and spun down at 6000 x g at 4°C for 15 minutes. The resulting supernatant was discarded and bacterial pellets were resuspended in a total of 16ml of Resuspension Buffer. A further 16ml of Lysis Buffer was added and the solution was inverted 10-12 times by hand. The sample was left to incubate at room temperature for 5 minutes. During this incubation step a separation column was equilibrated with 12ml of Equilibration Buffer by adding the buffer slowly to the rim of the column in a clockwise motion. After the incubation step 16ml of Neutralisation Buffer was added to the sample after which it was inverted five times, halting the lysis step. The sample was then pelleted down at 6000 x g at 4°C for 15 minutes. The resulting lysate was applied to the rim of the equilibrated separation column slowly, being careful to avoid uptake of any resulting flocculate. After the lysate had flowed through the column 5ml of Equilibration Buffer was added to the rim of the separation column after which the column filter was discarded. The remaining silica membrane was washed using 8ml of Wash Buffer. After the buffer had passed through the membrane the column was transferred to a 50ml falcon tube for plasmid collection. DNA was eluted by addition of 5ml of Elution Buffer. Elution buffer could be warmed to 37°C to increase DNA yield.

To precipitate the plasmid DNA 3.5ml of isopropanol was added to the elute. The sample was vortexed and left at room temperature for 2 minutes and then

centrifuged at 15,000 x *g* for 30 minutes. The supernatant was carefully removed and 3.5ml of 70% ethanol was added followed by a further centrifugation step at 15,000 x *g* for 5 minutes. The supernatant was carefully removed and the resulting pellet was left to air dry for 10-15 minutes at room temperature. Once the pellet was dry it was reconstituted with dH_2O. DNA was assessed using the NanoDrop 1000 (Thermo Scientific, USA) and stored at -20°C.

Probe Name	Clone	Start Probe (bp)	End Probe (bp)	Length (bp)	Location	Position to Gene	Fluorophores
CEBPA/G	RP11-475K23	33,379,992	33,528,442	148,451	19q13.11	Centromeric	Red
CEBPA/G	RP11-270I13	33,674,047	33,855,563	181,517	19q13.11	Centromeric	Red
CEBPA/G	RP11-547I3	33,778,074	33,941,111	163,038	19q13.11	Telomeric	Green
CEBPA/G	RP11-423J18	33,958,632	34,156,514	197,883	19q13.11	Telomeric	Green
CEBPB	RP4-710H13	48,578,865	48,658,543	79,778	20q13.13	Centromeric	Green
CEBPB	RP5-1185N5	48,658,544	48,772,032	113,588	20q13.13	Centromeric	Green
CEBPB	RP11-290F20	48,865,332	49,025,542	160,210	20q13.13	Telomeric	Red
CEBPB	RP5-894K16	49,025,543	49,157,049	131,606	20q13.13	Telomeric	Red
CEBPD	RP11-279A10	48,330,964	48511171	180,208	8q11.21	Centromeric	Red
CEBPD	RP11-626C19	48,437,683	48,624,740	187,058	8q11.21	Centromeric	Red
CEBPD	RP11-265N3	48,700,294	48,845,685	145,391	8q11.21	Telomeric	Green
CEBPD	RP11-17B11	48,815,291	48,977,937	162,646	8q11.21	Telomeric	Green
CEBPE	RP11-790N13	23,205,313	23,396,392	191,079	14q11.2	Centromeric	Green
CEBPE	RP11-298I3	23,452,107	23,467,632	15,525	14q11.2	Centromeric	Green
CEBPE	RP11-1075E4	23,895,114	24,068,603	173,490	14q11.2	Telomeric	Red
CEBPE	RP11-1080M7	23,641,443	23,806,016	164,574	14q11.2	Telomeric	Red
IGH	CTD-3034B12	106,081,078	106,124,038	42,960	14q32.33	Centromeric	Red
IGH	RP11-676G2	106,252,717	106,448,848	196,131	14q32.33	Centromeric	Red
ABL2	RP11-18E13	178,645,775	178,821,857	176,083	1q25.2	Centromeric	Red
ABL2	RP11-351I08	178,859,159	179,049,521	190,362	1q25.2	Centromeric	Red
ABL2	RP11-520H23	179,257,963	179,466,011	208,048	1q25.2	Telomeric	Green
ABL2	RP11-545A16	179,437,790	179,593,205	155,416	1q25.2	Telomeric	Green
CDKN2A/B	RP11-149I2	21,909,259	22,010,413	101,155	9p21.3	Covering	Aqua
IKZF1	W12-3001F15	50,381,496	50,422,338	40,842	7p12.2	Covering	Gold
PAX5	RP11-469D03	36,928,510	37,098,070	169,560	9p13.2	Covering	Gold
ABL2	G248P8248G11	179,079,924	179117361	37,438	1q25.2	Covering	Gold

Table 2.1. Composition of all FISH probes used in this study with clone names, sizes, locations and colours.

DNA was labelled with the Nick Translation DNA labelling kit (Enzo, UK) with fluorophores red (580nm), green (496nm), gold (525nm) or aqua (431nm)(Enzo, UK). 1µg of extracted DNA was diluted in a total of 25ul dH_2O. This was mixed with 5µl dNTP mix; 2.5µl dTTP mix; 2.5µl Fluorophore-dUTP; 5µl DNA Polymerase I; and 5µl DNase I. The solution was mixed gently and incubated at 15°C overnight in the dark. The reaction was halted by adding 3µl 0.5M EDTA. The sample was pipetted directly into an Illustra ProbeQuant G-50 Micro Column (GE Healthcare Life Sciences, UK) and centrifuged at 12,396 x g for one minute to remove any unbound nucleotides via gel filtration. The sample was precipitated by adding 3µg of Cot-1 DNA (Invitrogen) and then making up to 16µl with dH_2O. Following this 6µl of 3M sodium acetate (Sigma-Aldrich) and 160µl of ice-cold

100% ethanol were added, the sample was left for two hours at -80°C then spun for 30 minutes at 14,549 x g at -4°C. The supernatant was discarded and pellets air dried and re-suspended in 14µl hybridisation buffer (Cytocell, UK) and 6µl dH_2O.

2.3.1.1.2 General FISH Technique

Fixed patient nuclei were analysed by placing 2.5µl of the suspension onto the centre of 76mm x 26mm Superfrost glass microscope slides (Thermo Scientific). Slides were labelled and placed onto a heat block set at 60°C for 10 minutes to denature the DNA. Probes were mixed 1:1 with hybridisation solution unless stated, although this varied depending upon probe signal intensity. A total of 2µl probe mix was placed onto a 13mm diameter No. 1.5 glass coverslip (VWR International) and carefully placed onto the dried fixed cell nuclei. Air bubbles were removed by applying pressure to the coverslip which was sealed to the slide using Fixogum (Marabu). The slide was then placed into a HYBrite probe hybridisation platform (Abbott Molecular) which would heat the slide to 75°C for 5 minutes, to facilitate unravelling of nucleic material and improve access of the FISH probe, followed by incubation at 37°C for 16 hours.

After hybridisation the Fixogum was removed and the slides were soaked in 2xSSC solution until the coverslips detached. The slides were then placed into WASH1, warmed to 72°C in the water bath, for two minutes. Finally the slides were soaked in room temperature WASH 2 for 2 minutes, after which they were removed and drained on paper towels to remove excess solution. After this 7µl of DAPI solution (Vector Laboratories) was added to a 24mm x 50mm No. 1 glass cover slide (VWR International) which was placed onto a still wet sample slide. Any air bubbles were removed. Slides were analysed using a BX61 fluorescence microscope (Olympus) equipped with a 100x magnification, 1.30 aperture oil-immersion objective lens using the CytoVision 7.1 program.

2.3.1.1.3 **FISH Signal Interpretation**

The designed *CEBP* break apart probes showed several signal patterns, normal nuclei showed a signal pattern of 0 red, 0 green, and 2 fusion signals (0R0G2F) indicating that the red and green probes were in proximity to each other, creating

a yellow signal (Figure 2.1 Ai). Translocations involving one allele would separate the signals giving a signal pattern of 1R1G1F (Figure 2.1 Aiii).

For more complex cases, a fusion probe was created to confirm *IGH-CEBPD* rearrangements, this was achieved by labelling centromeric *IGH* clones in the red fluorophore and telomeric *CEBPD* clones in the green fluorophore. Gene rearrangement would be denoted by juxtaposition of the green *CEBPD* probe next to the red *IGH* probe giving one fusion signal and two separated signals (1R1G1F), while normal patient nuclei would show no translocation (2R2G0F) (Figure 2.1 B).

Cutoff slides were set up to determine the percentage of false positive results. Normal human peripheral blood samples were hybridised with the appropriate FISH probe on three independent slides. Slides were scored by both eye and automation to determine presence of abnormal nuclei. A mean value was calculated using the scores from the three blood samples to create a value above which abnormal populations would be considered as real.

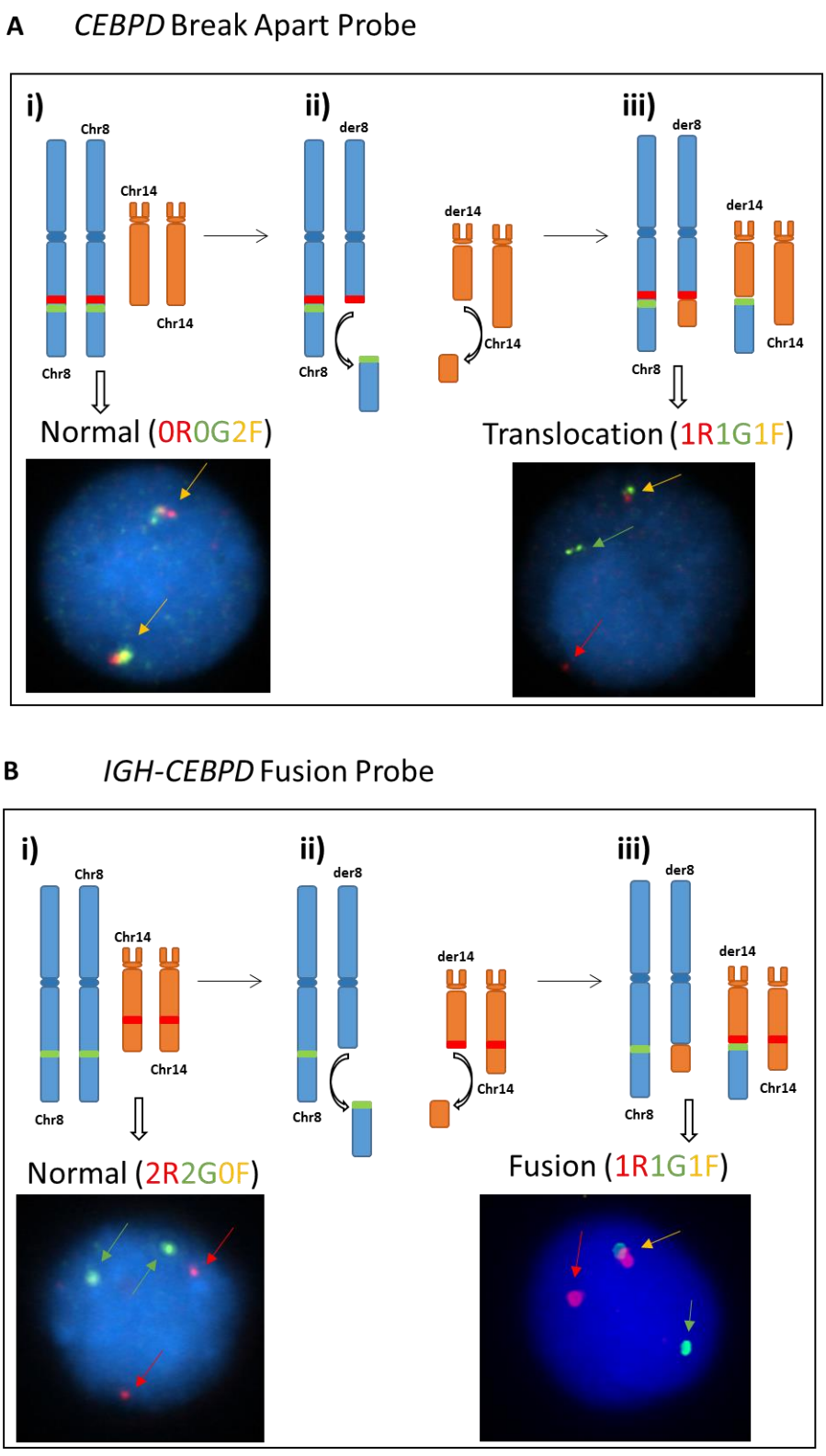


Figure 2.1. FISH probe designs and signal patterns.

A. The CEBPD break apart FISH probe was used to confirm CEBPD translocations. Centromeric DNA clones RP11-279A10 and RP11-626C19 labelled in red, telomeric DNA clones RP11-265N3 and RP11-17B11 (all at 8q11.21) labelled in green (Table 2.1). A.i. The pattern observed in normal patient nuclei; a signal pattern of 0R0G2F where the red and green signals are juxtaposed, this pattern indicates no translocation events. A.ii. The beginning of the translocation event, sections of chromosome 8 which contains the CEBPD gene and the green centromeric FISH probe, and chromosome 14 detach from their respective chromosomes. A.iii. The detached sections attach to the partner chromosomes, giving a

derived chromosomes 8 and 14, bringing the CEBPD gene into close proximity to the IGH enhancer and giving the FISH signal pattern of 1R1G1F. B. The IGH-CEBPD fusion probe was created to confirm the presence of the IGH-CEBPD translocation in patients with more complex translocation events. The telomeric clones from the CEBPD probe, RP11-265N3 and RP11-17B11 were combined with DNA clones CTD-3034B12 and RP11-676G2, within the IGH locus at 14q32.33, labelled in red (Table 2.1). B.i. Shows the pattern observed in a normal patient nucleus, with two probe sets, the green signals in the vicinity of the CEBPD gene on chromosome 8 and the red signals covering the IGH locus on chromosome 14, giving a signal pattern of 0R0G2F. B.ii. Shows the beginning of the translocation event, sections of chromosome 8 which contains the CEBPD gene and the red signal, and chromosome 14 detach from their respective chromosomes. B.iii. The detached sections attach to the partner chromosomes, giving a derived chromosomes 8 and 14. This translocation has brought CEBPD and the red signal into close proximity to the IGH locus and the green signal, creating a fusion signal with the signal pattern 1R1G1F.

2.3.1.2 Multiplex Ligation-dependent Probe Amplification (MLPA)

In the last three decades molecular techniques such as RT-PCR have also been used widely in screening for the detection of specific abnormalities. A more modern technique utilising RT-PCR is MLPA, a rapid multiplex PCR based technology allowing for the comparative quantification of multiple genetic regions (Schouten *et al.*, 2002). MLPA is very sensitive, detecting single nucleotide changes, with the advantage of simultaneous detection of up to 50 genomic DNA or RNA targets. The technique was created for use in high throughput genetic screening in a clinical setting, giving rapid, accurate results. A specific probe set is in routine use for the screening of BCP-ALL patients and investigation of CNA involving significant genes including (see below): *IKZF1*, *PAX5* and *CDKN2A/B* (Schwab *et al.*, 2010a) (Supplementary Table 7.1). The technique can identify single exon aberrations in genes, too small to be detected by FISH. Such aberrations have been shown to be important in prognosis (Moorman *et al.*, 2014). MLPA however cannot differentiate between biallelic deletions in mixed populations and monoallelic deletions, or reliably detect clonal populations of 20% or less, making it unsuitable for the analysis of aberrations expressed at a low level (Schwab *et al.*, 2010a). False positive results may be achieved when additional chromosomes are gained, making MLPA unsuitable for screening for ploidy changes.

2.3.1.2.1 **Multiplex Ligation-dependent Probe Amplification Protocol**

Patients with between 10-100ng DNA available were selected for MLPA analysis using the SALSA MLPA P335 kit (MRC Holland, The Netherlands).This kit includes probes for *IKZF1* (8 probes), *CDKN2A/B* (3 probes), *PAX5* (6 probes),

EBF1 (4 probes), *ETV6* (6 probes), *BTG1* (4 probes), *RB1* (5 probes), and the PAR1 region: *CRLF2, CSF2RA*, *IL-3RA* (1 probe each) (http://tinyurl.com/j5qn7ky). The protocol was performed as described by Schwab *et al* (Schwab *et al.*, 2010a), using the Beckman Coulter CEQ 8800 system and data were analysed using the CEQ 8800 analysis software and GeneMarker V1.85 analysis software (SoftGenetics).

2.3.1.3 Single Nucleotide Polymorphism Array

Single Nucleotide Polymorphism (SNP) arrays have replaced older methods such as Comparative Genomic Hybridisation (CGH), for investigation of the whole genome for copy number abnormalities and to detect loss of heterozygosity (LOH) (Raghavan *et al.*, 2005; Teh *et al.*, 2005). The Affymterix Genome-wide SNP 6.0 array provides 950,000 copy number variation markers and 744,000 probes, evenly spread throughout the genome, as well as 900,000 SNP markers that allow detection of LOH and uniparental disomies (copy number neutral LOH-CNLOH). Interestingly, CNLOH leads to inactivation of the *CEBPA* gene in a subset of AML patients, which is associated with an increased risk of leukaemia (Raghavan *et al.*, 2005; Teh *et al.*, 2005). SNP arrays provide an affordable method for novel gene target identification, as was demonstrated in 2007 with the discovery of the prognostic importance of *IKZF1* deletions in ALL (Mullighan *et al.*, 2007). SNP arrays however do have limits as the technology cannot detect balanced translocations or small abnormal populations.

2.3.1.3.1 **Single Nucleotide Polymorphism Array Protocol**

Patients with 750ng of available DNA were selected for SNP analysis. The samples were sent to Aros Biotechnology (Denmark) for analysis on the Affymetrix GeneChip SNP 6.0 array. Affymetrix genotyping console version 4.1.1.834 was used for data analysis.

The Median of the Absolute values of all Pairwise Differences (MAPD) was used to assess the quality of the array data. The MAPD value is the recommended method of assessing 'noise' in SNP copy number array data. The Log^2 gene expression value for a specific probe within a sample, divided by the reference gene expression value for the same probe. Log^2 values are taken for all probe

pairs, and a median value is assigned for the sample, this is the MAPD value. The MAPD value represents the typical variation between the reference SNP and the sample SNP values. The higher the MAPD value, the greater the deviation between reference and sample set, suggesting erroneous sample reads either due to sample quality or array issues.

2.3.2 General Laboratory Techniques

General techniques used throughout this study included the growth, storage, and extraction of bacterial cultures, for harvesting of desired plasmids. Protocols are also included for the DNA, RNA and protein extraction from mammalian cells, and their quantification.

2.3.2.1 Storage of Bacterial cultures

For long term storage of cultures, bacteria was grown overnight in LB broth, the following morning 800µl of the culture was mixed with 200µl of glycerol. The suspension was pipette mixed and placed in a microcentrifuge tube to be stored at -80°C.

2.3.2.2 DNA Extraction

2.3.2.2.1 Cell Lines and Patient Material

The DNeasy Blood & Tissue kit (Qiagen) was used to extract DNA from cell lines, xenograft cells, fixed cell patient material, and primary diagnostic viable cells. For fixed cells samples, cells were centrifuged at 14,000 x g for two minutes and the fix discarded. The resulting pellet was resuspended in sterile PBS and centrifuged again at 14,000 x g for two minutes and supernatant discarded, this wash step was repeated once more centrifuging for five minutes and discarding the supernatant. After this step samples from all sources were treated identically.

Pellet was resuspended in 200µl PBS and 20µl proteinase K. 200µl lysis buffer AL was added and samples were mixed thoroughly by vortexing and spinning down briefly to remove air bubbles. Samples were incubated on a hot block at 56°C for 10 minutes. After incubation 200µl, of 100% ethanol was added. The samples were vortexed and pipetted into DNeasy mini spin columns, the columns

were centrifuged at 5500 x g for one minute, flow through discarded and the column placed into a new 2ml collection tube. 500µl of wash buffer, Buffer AW1, was added and the column was centrifuged at 5500 x g for one minute. The flow through was discarded and the column placed into a new 2ml collection tube. 500 µl of Buffer AW2 was added and the column centrifuged at 16,800 x g for two minutes. Flow through was discarded and the column was spun for a further one minute at 16,800 x g to dry the DNeasy membrane. Columns were added to a pre-labelled 1.5ml Eppendorf for collection. 50µl of buffer AE was pipetted directly to the DNeasy membranes and left to incubate at room temperature for one minute. After incubation the columns were centrifuged at 5500 x g for one minute. To increase yield Buffer AE flow through could be passed through the DNeasy column for a second time or a second elution could be performed by pipetting fresh Buffer AE and leaving at -4°C overnight before centrifuging and collecting the DNA the following morning. Sample quality and quantity was analysed with the Nanodrop 1000 Spectrophotometer, with samples being stored at -20°C for future use.

2.3.2.2.2 Bacterial Plasmid Extraction

The QIAprep Spin Miniprep Kit (Qiagen) was used to extract plasmid DNA from bacterial cultures of 5ml or less. All centrifugation steps were performed at 18,000 x g in a minicentrifuge. The kit lyses bacterial cell walls and denatures genomic DNA and plasmid DNA through use of a sodium dodecyl sulphate and sodium hydroxide alkaline mixture. The addition of potassium acetate allows the smaller plasmid DNA to reanneal while the larger genomic DNA clumps together and is removed along with bacterial protein upon centrifugation with the selection column. Further spin steps remove salts and endonucleases.

A single colony growing on a LB agar selection plate was collected using a pipette tip and placed into a 15ml Falcon tube with 5ml of LB media, the colony was grown over night in a shaking incubator at 37°C. For long term storage of cultures, the following morning 800µl of the culture was mixed with 200µl of glycerol. The suspension was pipette mixed and placed in a microcentrifuge tube to be stored at -80°C. Cultures to be used immediately were pelleted by centrifuging at 6,800 x g for three minutes. Pellets were resuspended in 250µl Buffer P1 and transferred to individual 1.5ml microcentrifuge tubes. 250µl of Buffer P2 was

added and the samples were mixed thoroughly by inverting 4-6 times. The samples were incubated at room temperature for five minutes for the lysis step which was ended by adding 350µl Buffer N3 and inverting 4-6 times to halt the reaction. Samples were pelleted by centrifugation for 10 minutes. The resulting supernatant was added to QIAprep spin columns and centrifuged for one minute, the supernatant discarded. The QIAprep columns were washed with 500µl of Buffer PB and centrifuged for one minute, the flow through was discarded. The columns were then washed with 750µl of Buffer PE and centrifuged for one minute and flow through was discarded, a further one minute spin step was performed to dry the column membrane from any residual buffers. Columns were then switched to a 1.5ml microcentrifuge tube for collection of plasmid DNA. 50µl elution buffer EB or dH_2O applied directly to the membrane, incubated at room temperature for one minute and DNA was harvested by centrifugation for one minute. DNA concentration and quality was assessed using a Nanodrop 1000 Spectrophotometer, samples were stored at -20°C.

2.3.2.2.3 DNA Precipitation

DNA precipitation was performed to purify and concentrate DNA after extraction using the reagents in Table 2.2. All volumes used in this protocol were relative to the original amount of DNA lysate used. 1/10 volume of sodium acetate (final concentration of 0.3M) was added to the DNA lysate. If concentrating small amounts of DNA, 1µl of glycogen was added to the solution to improve visibility of the pellet. 2-2.5 volumes of ice cold 100% ethanol were added and mixed thoroughly. The solution was spun down at maximum speed in a microcentrifuge for 15 minutes. Supernatant was removed and discarded carefully, so as not to disturb the resulting pellets. 0.7-1ml of 70% ethanol was added to the pellets and mixed. The solution was spun down at maximum speed in a microcentrifuge for 5 minutes. Supernatant was discarded and the spin step repeated, and any residual ethanol was removed with a pipette. The DNA pellets were allowed to air dry for 10-30 minutes. When all the remaining ethanol had evaporated pellets were resuspended in H_20 or TE buffer.

Reagents
3M Sodium Acetate (pH 5.2)

DNA lysate
Ice cold 100% ethanol
Room temperature 70% ethanol
Glycogen 5mg/ml

Table 2.2. Reagents for DNA precipitation.

2.3.2.2.4 DNA Quantification Qubit Fluorometer

Qubit was used to analyse double stranded DNA using highly sensitive fluorescent based quantification assays which use dyes that only emit signal when bound to the specific target molecule.

Reagents were thawed in the dark and used when at room temperature. A master mix was created for each sample to be analysed plus an extra two reactions for the two Qubit standards. The master mix consisted of 199µl Qubit Buffer and 1µl Qubit Reagent for a total of 200µl of master mix. To set up the standards 10µl of the appropriate standard was added to 190µl of the master mix. Tubes were vortexed and spun down to remove any air bubbles, and incubated at room temperature in the dark for 2 minutes. The Qubit Fluorometer was switched on and the dsDNA assay was selected, standard one was inserted into the machine and measured, followed by standard two. Following this 199µl of the master mix and 1µl of sample DNA were mixed, vortexed and incubated in the dark at room temperature for 2 minutes. Samples were then analysed and values were recorded.

2.3.2.3 RNA Extraction

RNA extractions were performed using the RNeasy Micro (Qiagen) and Mini (Qiagen) kits. The kit used was dependent upon cell numbers: less than 5 x 10^5 cells extracted - Micro kit; above 5 x 10^5 cells - Mini kit.

2.3.2.3.1 QIAshredder

Collected cells were first homogenised by being passed through QIAshredder (Qiagen) columns. Protein was also collected from the resulting supernatant if cell numbers were over 5 x 10^5.

Up to 1 x 10^7 cells were resuspended in 350µl of RLT buffer and passed through QIAshredder columns (Qiagen), which employ a biopolymer-shredding system to homogenise cells and increase RNA yield. Cells were centrifuged at 20000 x g for two minutes and elute was kept for RNA and protein collection.

2.3.2.3.2 RNeasy Mini Kit

QIAshredder elute was collected in a 2ml collection tube to which an equal volume (350µl) of 70% ethanol was added. The suspension was pipette mixed and placed into an RNeasy column and centrifuged at 8000 x g for 30 seconds. Flow through was kept for protein isolation described in section 2.3.1.4.1. The RNeasy spin column was placed into a new 2ml collection tube and 700µl of Buffer RW1 was added to the sample. The column was centrifuged at 8000 x g for 30 seconds, flow through discarded and the collection tube dried on tissue paper to remove any remaining liquid. 500µl of Buffer RPE was added and the column was centrifuged at 8000 x g for 30 seconds, flow through was discarded and the 2ml collection tube dried on tissue paper. A second wash step was performed with 500µl Buffer RPE and the column was centrifuged at 8000 x g for two minutes, flow through was discarded. The samples were placed into clean 2ml collection tubes and centrifuged at 8000 x g for one minute to dry the membrane of any residual buffers. The columns were then placed into clean pre-labelled 1.5ml microcentrifuge tubes and RNA was eluted by adding dH_2O applied directly to the membrane, incubating at room temperature for one minute and centrifuging at 8000 x g for one minute, dH_2O amount varied depending upon cell number extracted with a range of 20-50µl. To increase yield dH_2O flow through could be put through the RNeasy column for a second time. RNA elute was analysed with the Nanodrop 1000 Spectrophotometer or the Bioanalyser. Samples were stored at -80°C.

2.3.2.3.3 RNeasy Micro Kit

QIAshredder elute was collected in a 2ml collection tube to which an equal volume (350µl) of 70% ethanol was added. The suspension was pipette mixed and placed into an RNeasy MiniElute spin column and centrifuged at 8000 x g for 30 seconds. Flow through was discarded and 350µl of Buffer RW1 were added

to the samples and then centrifuged 8000 x g for 30 seconds. Next 10µl of DNase I stock solution was added to 70µl of Buffer RDD, the solution was mixed by inversion then added directly to the RNeasy MiniElute spin column membrane and incubated at room temperature for 15 minutes. After the incubation step 350µl of Buffer RW1 were added to the samples and then 8000 x g for 30 seconds. Resulting elute and collection tubes were discarded. The columns were placed into a new 2ml collection tube and 500 µl of Buffer RPE were added to the spin columns, samples were centrifuged at 8000 x g for 30 seconds, flow through was discarded. Next, 500 µl of 80% ethanol were added, and columns were centrifuged at 8000 x g for two minutes, elute and collection tubes were discarded. Columns were placed into new 2ml collection tubes and lids were opened for a 5 minute dry spin at full centrifugal speed. Collection tubes were discarded and columns were placed into clean pre-labelled 1.5ml microcentrifuge tubes, RNA was eluted by adding 14µl dH_2O applied directly to the membrane and centrifuging at full speed for one minute. To increase yield dH_2O flow through could be passed through the RNeasy MiniElute spin column for a second time or a second elution could be performed by pipetting 14µl of fresh dH_2O and leaving at -4°C overnight before centrifuging and collecting the RNA the following morning as a second elute, although this approach would lead to some RNA degradation. RNA concentration and quality was assessed using a Nanodrop 1000 Spectrophotometer, samples were stored at -80°C.

2.3.2.3.4 RNA Analysis Agilent RNA 6000 Nano Kit

Several platforms require high quality RNA for analysis. RNA quality was measured on the Agilent 2100 bioanalyser to generate RNA integrity numbers (RIN). RIN scores are generated from the algorithmic analysis of the 28S and 18S rRNA ratio.

The priming station syringe clip was set to the stop position. All reagents were equilibrated to room temperature before use, and all dyes were kept in the dark when possible. RNA 6000 NanoLadder was heat denatured at 70^{o}C for 2 minutes before use. Aliquots were stored at -80^{o}C. RNA samples were aliquoted at a maximum of 50ng/µl in 3µl of solution. RNA was denatured at 70^{o}C for 2 minutes before use.

The RNA nano Chip gel was prepared by mixing 550µl of the gel into a spin filter and centrifuging at 1500 x g for 10 minutes. 65µl of the filtered gel was used and the rest was stored at 4^{o}C to be used within 4 weeks. 1µl of vortexed RNA dye was added to the gel. The solution was vortexed thoroughly and centrifuged at 13000 x g for 10 minutes. While the gel was spinning the RNA chip was placed into the priming station and 9µl of gel-dye mix was pipetted into the appropriate well displayed on the kit instructions. Placing the syringe plunger at the 1ml mark, the priming station was clipped into position and the plunger was pushed down until held by the metal clip. The plunger was left in this position for 30 seconds, and the clip was released, the plunger was left for 5 seconds and then moved back up to the 1ml mark on the syringe. The priming station was then unclipped and a further 9µl each was pipetted into two further marked wells. 5µl of RNA marker was added to all appropriate wells, 1µl of RNA ladder was added to the ladder well, 1µl of RNA sample was added per sample well, up to 12. 1µl of RNA marker was added to unused sample wells. The chip was then placed into an IKA vortexer and vortexed at 2400 rpm for 1 minute. The chip was ran on the Agilent 2100 Bioanalyzer instrument within 5 minutes of vortexing.

2.3.2.4 Protein Extraction

2.3.2.4.1 Newcastle Protocol

This protocol uses acetone to lower the dielectric constant of the cell lysate, lowering protein solubility and encouraging precipitation. The resulting protein is then solubilised and stored in urea buffer.

The first elute from RNeasy spin column protocol was mixed with two volumes (700µl) of acetone. The samples were capped and left to incubate on ice for one hour. After incubation samples were centrifuged at 20000 x g for 15 minutes. Acetone supernatant was discarded and samples were centrifuged 20000 x g for two minutes and any remaining solution was removed. Resulting pellets were resuspended in Urea Buffer at 100µl per 1 x 10^{7} cells. Samples were stored at -20°C.

2.3.2.4.2 Cincinnati Protocol

This protocol lyses cells through sonification, and protects the resulting protein by storage in Laemmli buffer, which consists of protease inhibitors, dye and β-Mercaptoethanol which serves to reduce intra and inter-molecular disulphide bonds within the protein, linearizing it.

Cells were pelleted and washed in 1ml of ice cold PBS and spun down at 10,400 x g for 2 minutes, supernatant was discarded and pellets were resuspended in 30ul Direct Lysis Buffer (Section 2.2.5.2.2) per 1 million cells, mixed thoroughly, and placed on ice. Cells were then sonicated and boiled at 96^{o}C for 5 minutes after which they were ready for use immediately or could be frozen down and stored at -80^{o}C.

2.3.2.5 cDNA Synthesis

The Promega M-MLV Reverse Transcriptase system was used to reverse transcribe RNA into cDNA.

1µg of extracted RNA was mixed with 1µl of Oligo (dT) primer (20µg) (Promega) to a total of 12.5µl with dH_2O. This solution was heated to 70°C for five minutes on a hot block. After the incubation step, the sample was pipette mixed with 12.5µl of the cDNA synthesis master mix (Table 2.3). The sample was incubated at 42°C for one hour, the reaction was inactivated by heating at 95°C for five minutes. The sample was diluted with 75µl of dH_2O for a final concentration of 1 µg of cDNA.

Reagents	**Volume**
5X Reaction Buffer	5µl
10mM dNTP mix	1.25µl
RNAsin Ribonuclease Inhibitor 40U/µl	0.625µl
M-MLV Reverse Transcriptase 200U/µl	1µl
dH_2Oµl	4.625µl

Table 2.3. cDNA synthesis master mix. All reagents from Promega.

2.3.3 General Cell Culture

General cell culture techniques include the thawing, counting, viable storage, and culture of viable mammalian cells.

2.3.3.1 Thawing Cells

Working quickly but gently, vials were thawed in a water bath set at 37°C, sterilised with ethanol and slowly re-suspended in 1 ml of the appropriate culture medium warmed to 37°C. The re-suspended cells were added to a 15ml falcon tube containing the appropriate culture medium to make a total of 10mls. Cells were centrifuged at 365 x g for 5 minutes, supernatant was discarded and cells were re-suspended and plated out for further growth.

2.3.3.2 Cell Counts Using Haemocytometer

Cell counts were performed using a haemocytometer, trypan blue was used to test viability. Cell suspensions were mixed with trypan blue at 1:1. 10ul of the solution was pipetted into the haemocytometer chamber and cells were counted using an inverted phase contrast microscope. Dead cells stained blue due to cytoplasmic rupture and were disregarded when counting. Four quadrants were counted, a quadrant has a volume of 0.1 mm3, the number of cells in four quadrants (at a 1:1 dilution) multiplied by 5000 gives the value of cells per ml.

2.3.3.3 Cell Culture of Suspension Cells

Free floating cells were counted using the haemocytometer to calculate the necessary dilutions. The cells and medium were transferred to a 50ml falcon tube and spun down at 365 x g for 5 minutes. The supernatant was aspirated and the cells were re-suspended in fresh medium. The cells were re-suspended at 5 x 10^5 cells per ml and added to a new culture flask and the passage number was noted.

2.3.3.4 Cell Culture of Adherent Cells

For the passaging of adherent cells, old medium was aspirated carefully so as not to disturb the cell layer, PBS at room temperature was added ensuring coverage of all cells to wash off any remaining FBS. PBS was aspirated and 1-

2mls of 1X trypsin was added to cover all cells, left at room temperature for 2 minutes and checked under an inverted phase microscope to observe detachment from the culture flask. Medium was added to the flask, to wash off remaining adhered cells. The solution was collected and added to a 50 ml falcon tube and spun down at 365 x g for 5 minutes. Supernatant was aspirated and cells re-suspended in fresh medium. T The cells were re-suspended at 5 x 10^5 cells per ml and added to a new culture flask with fresh medium. Passage number was noted.

2.3.3.5 Heat Inactivation of foetal bovine serum

Foetal bovine serum (FBS) was thawed at room temperature and mixed thoroughly. The FBS was then incubated at 56°C for 30 minutes, and was mixed 3-4 times during this incubation step. Heat inactivated FBS was then aliquoted into 50ml falcon tubes and stored at -20°C for future use.

2.3.3.6 Freezing Viable Cells – Newcastle and Cincinnati Protocols

Cells were frozen at up to 1x10^7 per ml. Selected cells were spun down at 365 x g for 5 minutes, supernatant was aspirated and the pellet was re-suspended with freezing solution (Table 2.2.5.3) in cryo vials and placed on ice until ready for storage at -80°C. In Cincinnati pelleted cells were re-suspended in 400µl of solution 1 combined with 100µl of 25% BSA and placed in a cryo vial. In a separate vial 400µl of solution 2 was combined with 100µl of 25% BSA solution and was slowly added dropwise to the solution 1, cell pellet mixture. Working rapidly, re-suspended cells were placed into a room temperature Mr. Frosty freezing container, and frozen down at -80°C. Vials were moved to liquid nitrogen for long term storage.

2.3.4 General Cloning Techniques

General cloning techniques were used to create plasmids through restriction digests, isolation of specific fragments through gel purification, and finally ligation. Desired plasmids were then expanded by transforming competent E. coli.

2.3.4.1 Collecting Plasmids from Whatman paper

Plasmids were shipped having been blotted onto Whatman. The plasmid was collected by soaking the paper in 100µl of TE buffer for 2 hours at 37°C. The TE buffer containing the plasmid was analysed on the Nanodrop 1000 Spectrophotometer and stored at -20°C for future use.

2.3.4.2 Plasmid Restriction Enzyme Digest

Restriction enzymes were used to integrate desired DNA sequences into plasmids of choice for downstream applications, such as retroviral particle production.

DNA sequences to be digested were pasted into the New England Biolabs Cutter V2.0 program to access optimal restriction enzymes. The appropriate restriction enzyme was selected and ordered. 0.2-1.5µg of template DNA was mixed with 2-10x excess of the restriction enzyme. A master mix was created (Table 2.4) up to the volume of 20µl and incubated at the enzymes optimum temperature for 1-4 hours. Restriction enzyme buffer and incubation temperature were both dependent upon the restriction enzyme used (Section 2.2.6.3). Results were analysed by gel electrophoresis and band size.

Component	**Volume**
dH_2O	16.5µl
Restriction Enzyme 10X Buffer	2µl
Acetylated BSA 10µg/µl	0.2µl
DNA 1µg/µl	1µl
Restriction Enzyme 10u/µl	0.3µl

Table 2.4. Restriction enzyme digest master mix.

2.3.4.3 Gel Digest

After performing successful restriction digests on target plasmids confirmed by gel electrophoresis, bands were analysed using the SYNGENE G:box imager and selected bands were excised using a clean scalpel. The cut gel was placed into a clean 1.5ml microcentrifuge tube and weighed. The GeneJET Gel Extraction Kit was used to dissolve the gel and access the DNA. Binding Buffer was added to the gel, 1µl per 1mg of gel weight. The mixture was incubated at

50-60°C for 10 minutes and inverted several times to facilitate the melting process. After ensuring that the gel had completely dissolved the mixture was vortexed and up to 800µl was added to a GeneJET Purification Column. The column was centrifuged at 13,000 x g for one minute. Flow through was discarded and the column was returned to the collection tube. 100µl Binding Buffer was added and the column was centrifuged at 13,000 x g for one minute, flow through was discarded. 700µl of Wash Buffer was added and the sample centrifuged at 13,000 x g for one minute, flow through was discarded. The empty column was centrifuged at 13,000 x g for one minute to remove any residual buffer. The column was then placed into a clean pre-labelled 1.5ml microcentrifuge tube and the DNA eluted with 50µl of Elution Buffer pipetted directly onto the column membrane after centrifugation for one minute. Elution volume can be altered to increase DNA concentration. DNA was analysed with the Nanodrop 1000 Spectrophotometer and stored at -20°C.

2.3.4.4 Thermosensitive Alkaline Phosphatase (TSAP)

TSAP was used to improve cloning efficiency by catalysing the removal of 5' phosphate groups and preventing the recircularization of plasmid DNA after restriction digests, facilitating insert DNA integration.

After a restriction digest where up to 1µg of DNA was used, 1µl of TSAP was added for digests using Promega Buffers A-L and MULTI-CORE 10X Buffer, and 2µl of TSAP for digests using Promega Buffer F. The samples were then incubated at 37°C for 15 minutes, followed by heat inactivation at 74°C for 15 minutes. The samples were then ready for use with T4 DNA ligase (Promega).

2.3.4.5 T4 DNA ligation

T4 DNA ligase catalyses the formation of phosphodiester bonds between 5' phosphate and 3' hydroxyl termini in DNA, joining blunt ends and repairing nicks in double stranded DNA.

After insert and vector DNA was isolated, and vector DNA was prepared with TSAP and the samples were ligated to create the desired plasmid.

The recommended ratio of vector to insert in the Promega T4 Ligase protocol varied between 1:1, 1:3 or 3:1. The calculation used to determine the amount of insert was as follows:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng of insert}$$

TSAP dephosphorylated vector DNA was mixed with the calculated amount of insert and mixed as shown in Table 2.5. Samples were incubated at either room temperature for three hours, 4°C overnight, or 15°C for 4-18 hours.

Component	**Volume**
Vector DNA	Up to 100ng
Insert DNA	Up to 17ng
Ligase 10X Buffer	1µl
T4 DNA Ligase	0.1-1u
Nuclease Free H_2O	Final volume of 10µl

Table 2.5. Composition of T4 DNA ligation master mix, per sample.

2.3.4.6 Transformation of competent bacteria

Whole and recombined plasmid constructs were amplified through transfection into Stbl3 *E. coli* cells, specialised bacteria designed to reduce the frequency of homologous recombination of long terminal repeats and provide a transformation efficiency greater than 1×10^8 transformants /µg plasmid DNA. Amplified plasmids were extracted and aliquots frozen down for future use.

500 µl of competent One Shot Stbl3 *E. coli* cells (Life Technologies) were thawed on ice, the bacteria were separated into aliquots of 250µl per reaction. Each aliquot was seeded with 1µl of the relevant DNA construct and one aliquot was seeded with 1µl pUC19 positive control DNA (10 pg/µl) and gently flick mixed, the reaction was then incubated on ice for 30 minutes. After this incubation period the sample was heat shocked by being placed at 42°C for 45 seconds before being incubated on ice again for a further two minutes. Following this 125µl SOC medium (Invitrogen) was then added per sample and incubated at 37°C in a shaking incubator for an hour. The samples were evenly spread onto pre-

prepared ampicillin LB agar selection plates, and left over night for up to 16 hours at 37°C. The following day plates were checked for colony formation, with only ampicillin resistant colonies growing on the selection plates indicating plasmid recombination was successful. Single colonies were selected, cultured, and stored as glycerol stocks and for growth for plasmid extraction using the Qiagen Mini prep kit as described in Section 2.3.1.1.

2.3.5 Isolation of CD34+ Cells

These techniques were performed for the isolation and expansion of human cord blood CD34+ cells.

2.3.5.1 Preparation of Cord Blood

Fresh cord blood was received in blood bags, which were sterilised with 70% ethanol before being opened in a tissue culture hood. The blood was split between several 50ml falcon tubes, never exceeding more than 40mls per tube. 7mls of sterile hydroxyethyl starch (HES) solution were added to each falcon tube, the solutions were gently mixed by hand and left at room temperature for 1 hour. After the incubation period the red cells pooled at the bottom of the tube, the top layer of the solution was collected when the red layer was less than one third of the total liquid. Supernatant from all tubes was collected and pooled into one 50ml falcon tube, the red cells were discarded. The collated solution was spun down at 300g for 10 minutes, supernatant was aspirated and cells were re-suspended in 20mls of FACS buffer. Cells were counted using either a haemocytometer or Hemavet 950 FS machine (Erba Diagnostics).

CD34+ cells were isolated following the Miltenyi Biotec CD34+ Micro Bead Kit following the LS column instructions with some amendments. (http://www.miltenyibiotec.com/~/media/Images/Products/Import/0001400/IM0001498.ashx).

The sample was centrifuged at 300×g for 10 minutes after which the supernatant was discarded. Up to 2 x 10^8 total cells were resuspended in 300µl of FACS buffer with 100µl of FcR Blocking Reagent (Miltenyi Biotec and 100µl of CD34 MicroBeads (Miltenyi Biotec). The solution was mixed well and left at 4°C for 30 minutes. After the incubation step cells were washed with 10mls of FACS buffer

and centrifuged at 300×g for 10 minutes, discarding the supernatant after the spin step. The sample was then resuspended in 500µl of FACS buffer ready for magnetic separation.

2.3.5.2 Magnetic Selection of CD34+ Cells

To prepare for the separation, the LS column was attached to the MACS Magnetic separator (Miltenyi Biotec), a 30µm Pre-Separation Filter (Miltenyi Biotec) was added to the column to avoid cell clumping and blockage, the column and separator were washed with 3ml of FACS buffer. The cell suspension was then slowly added to the column with a 50ml Falcon tube underneath the column to catch unbound cells. The column was then washed with 3ml of FACS buffer, this process was repeated twice for a total of three washes. The column was then removed from the MACS Magnetic separator placed into a 15ml Falcon tube, where the magnetised cells were washed out using 5mls of FACS buffer and a plunger. Cells were counted using a haemocytometer, a small sample was set aside for FACS analysis, the remainder were frozen down using the Cincinnati Freezing Protocol (Section 2.3.2.6). Isolated cells were stained with the anti-human CD34 fluorochrome and the 7AAD viability marker, purity was analysed via flow analysis.

2.3.6 Retroviral Protocols

Retroviral vectors were created to express *CEBPD* and IK6. Variants of the Murine Stem Cell Virus (MSCV)-internal ribosomal entry site (IRES) retroviral plasmids (Figure 2.2) were combined with the RD114 envelope plasmid, which specifically targets HSCs and reduces cell toxicity (Bell *et al.*, 2010), and the m57 gag/pol plasmid.

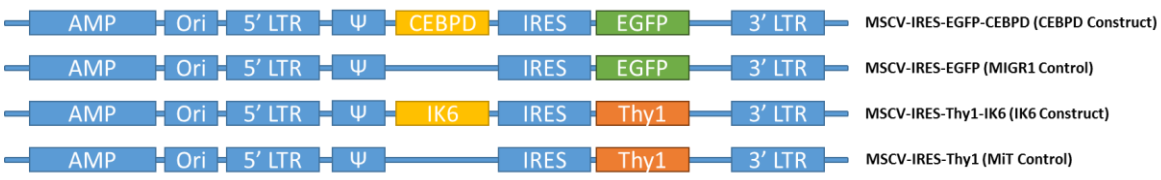


Figure 2.2. Retroviral Plasmids used during this study.

AMP = Ampicillin. Selection marker. Ori = Origin of amplification. Viral replication is initiated at this site. 5'/3'LTR = Long terminal repeats. Enhances transcriptional activation and prevents transcriptional repression, variable among viruses and

contain promoters, enhancers, and transcriptional termination. Ψ = Psi Packaging element. Regulates the packing of the retroviral genome in the viral capsid during replication. IRES = Internal ribosome entry site. Allows for the initiation of translocation in the middle of an mRNA sequence. Yellow boxes show target gene cDNA, green boxes represent the EGFP selection marker and orange boxes represent the Thy1 selection marker. MIGR1 denotes the MSCV-IRES-GFP empty vector plasmid, MiT denotes the MSCV-IRES-Thy1 empty vector plasmid (Coffin et al., 1997; Cherry et al., 2000; Buchschacher, 2001).

Retroviral particles were created using calcium phosphate co-precipitation and used to transduce CD34+ cells.

2.3.6.1 Calcium Phosphate Transfection

Calcium phosphate co-precipitation, introduced in 1973 (Graham and van der Eb, 1973) is a popular transfection method due to the low cost, ease of the technique, and efficiency with numerous culture and cell types. The technique functions by creating a precipitate of calcium, phosphate and DNA, which facilitates the binding of the precipitated DNA to the target cell surface, and entry into the cell via endocytosis. The protocol includes aeration of the phosphate buffer while adding the DNA-calcium chloride solution to ensure the resulting precipitate is fine, and not clumped, which hinders DNA adherence to the target cell surface. The protocol has several challenges including identifying and maintaining optimal pH for precipitate formation, causing cell toxicity due to buffer and salt concentrations, and a lower transfection efficiency in comparison to other chemical transfection methods.

2.3.6.1.1 **Calcium Phosphate Transfection Protocol**

2-3 million 293T cells were plated in a 10cm tissue culture treated plate 24 hours before transfection. Cells were transfected when ~70% confluent. All transfection reagents were thawed at room temperature and mixed thoroughly before use. Two tubes were prepared for the transfection, tube 1 containing DNA, water and $CaCl_2$ in that order. 1ml of HBSS x2 was added to tube 2.

Reagents	**µg per dish**
Viral envelope (RD114)	3
Viral gag/pol (M57)	10
Viral vector	12

Table 2.6. Insert DNA concentrations for retroviral particle synthesis per 10cm tissue culture dish.

In a tissue culture hood, while gently vortexing the HBSS x2 solution, the contents of tube 1 were added drop wise to tube 2, continuously vortexing. After combining, the mixture was left at room temperature for 30 minutes. After incubation the solution was once again vortexed and added drop wise to the 293T cells which were cultured in media appropriate for the cells to be transduced (Section 2.2.5.5). 10ul of 25mM chloroquine solution was also added. The plate was swirled to evenly distribute the virus. Plates were left for 24 hours before harvesting the first batch of virus.

2.3.6.2 Harvesting Viral Media

Viral medium was harvested from 293T cells 24 hours after transfection and replaced with fresh medium. Harvested media was syringe filtered (0.2µm), and was used fresh on cells undergoing transduction, remaining media was stored at -80°C. The second harvest was performed 12 hours after the first and followed the same protocol. The third and final harvest was performed 12 hours later.

2.3.6.3 CD34+ Cell Transduction

A 6 well tissue culture treated plate was coated with 2mls of retronectin, which was left either overnight at 4°C or for 1 hour at room temperature. Retronectin was collected and stored at -20°C to be re-used up to three times. The plate was washed with 1x PBS and blocked with 1x PBS + 2% BSA for 30 minutes at room temperature. The plate was washed with PBS and stored at 4°C with 2mls of PBS in coated wells to prevent them drying out. The plate can be stored for up to a week. Before use, PBS was aspirated and 3mls of viral media were added to coated wells, the plate was spun at 1000-2000 x g for 90 minutes. During this spin the cells for transduction were re-suspended in viral media with the appropriate cytokines and 4ug/ml of polybrene at a cell density of 125k cells per ml. The cells were centrifuged at 815 x g for 60 minutes. After spinning the cells were re-suspended in viral medium and cytokines and added to the retronectin coated wells. Cells were transduced twice more at 12 hour intervals. Medium was carefully aspirated and freshly harvested viral medium, cytokines and polybrene

were added. After a further 6-8 hours cells were harvested using disassociation buffer and plated out in normal medium with SCF, TPO, FLT3-L, all at 100ng/µl for up to 48 hours, to expand the primitive lymphoid population. Cells were plated out for specific experiments after 8 hours. No viral titrations were performed as the main aim of these experiments was the high expression of *CEBPD* and IK6.

Transduced CD34+ cells were cultured in cytokine rich medium to stimulate expansion of haematopoietic cells and progenitors. Cytokines included in this study were as follows:

SCF – A growth factor with a broad range of activities including the promotion of haematopoiesis. SCF is expressed as a membrane bound protein and can also be cleaved into a soluble form, both forms promote proliferation. It performs most efficiently combined with other cytokines: such as TPO, GM-CSF, G-CSF, M-CSF, IL-3, and IL-7, to induce proliferation.

TPO - Thrombopoeitin is a regulatory factor for megakaryocytopoiesis and thrombopoiesis. It functions by stimulating the growth and maturation of megakaryocytes, synergistically with other cytokines, to induce haematopoietic proliferation and differentiation.

FLT3-L – FLT3 Ligand binds and activates the FLT3 receptor. FLT3 is important in the development of dendritic cells. It synergistically stimulates the proliferation and differentiation of haematopoietic cells and haematopoietic progenitors.

IL-3 - A growth factor which promotes survival, differentiation and proliferation of multiple myeloid and erythroid lineages. It also directly activates monocytes, suggesting additional immunoregulatory roles. This cytokine is heavily dependent upon co-stimulation of other cytokines.

IL-6 – A pleiotropic cytokine, which regulating immune and inflammatory responses. It stimulates B-cell differentiation and synergizes with IL-3 in megakaryocyte development and platelet production.

IL-7 – A hematopoietic growth factor, which stimulates the differentiation of pluripotent hematopoietic cells into lymphoid progenitor and lymphoid lineage (B cells, T cells and NK cells) cells.

2.3.6.4 Co-culture of CD34+ Cells

CD34+ transduced cells were cultured with MS-5 stromal cells in α-MEM medium with 20% FBS and 1% pen/strep and SCF, Flt-3L, IL-7 cytokines at 10ng/ml; 6 well tissue culture plates and 10cm tissue culture dishes were used for co-culture. Wells were coated with a 50% gelatin solution and left for one minute at room temperature, to encourage stromal cell adherence to the plastic surface. Stromal cells were plated out in the gelatin coated wells and grown to confluence. CD34+ cells were seeded at 80,000 cells per ml. Cells were monitored and split once or twice a week depending upon confluence. Fresh medium was added every time cells were split. When splitting cultures, the stromal layer was gently washed with medium to loosen settled cells while avoiding detaching the stroma, to provide an accurate representation of cells when performing cell counts and to harvest sufficient cells for flow cytometry analysis or cell collection.

2.3.6.5 Harvesting of Co-cultured CD34+ Cells for Flow Sorting

To harvest co-cultured CD34+ cells, first the medium was removed and stored to be spun down. Next, the well and stromal layer was washed several time with PBS and added to the collected media. 1X trypsin was added, covering the stromal layer, and incubated at 37°Cfor 3 minutes to loosen the cell layer. After the incubation step the stromal layer was carefully removed from the well and added to a separate tube, the well was washed with more PBS to collect any remaining cells and added to the medium tube. The falcon containing the stromal layer was vortexed to encourage disassociation of any remaining attached CD34+ cells. The stromal and medium were filtered through 50µl mesh to remove the larger stromal cells, then collated and spun down at 365 x g for 5 minutes. The resulting pellet was counted and stained following the FACS staining protocol (Section 2.3.9.3).

2.3.7 Xenograft Techniques

Xenograft mouse models were used for the expansion of viable patient samples and transduced CD34+ cells.

2.3.7.1 General Monitoring of Mice

NOD scid gamma (NSG) mice were raised with up to six littermates in ventilated cages (IVCs) supplied by a sterile air supply. The animals were checked weekly by researchers and daily by trained technicians. The mice were weighed weekly to monitor health, and were sacrificed if a weight loss of more than 10% was observed for three consecutive days or if 20% weight loss was observed at any point, as disease burden or other factors had begun to affect quality of life.

2.3.7.2 Busulfan Conditioning of Immunodeficient Mice

Busulfan, a CML chemotherapeutic, was used to condition murine recipients by thinning bone marrow to improve engraftment of patient or transduced material, by removing competing cells from the environmental niche.

To create a 10X Busulfan stock, Busulfan powder was dissolved in DMSO at a concentration of 30mg/ml. After briefly vortexing, the solution was placed on a rocking platform for ~5 minutes until all Busulfan crystals had dissolved. The solution was kept at room temperature at all times as incubation on ice would lead to precipitation of the powder. The solution was passed through a 0.2µm syringe filter in a tissue culture hood to remove potential contaminants. The 10X stock was diluted to a 1X concentration using PBS prior to injection into mice by adding 150µl of 10X stock to 1350µl of PBS. Upon dilution to the 1X concentration the solution had to be used within a few minutes due to rapid precipitation of Buslfan in PBS. The 1X stock was injected intraperitoneally at 30mg/kg per NSG. Busulfan conditioning was performed 24 hours before intra-femoral injections which were performed by Dr. Helen Blair.

2.3.7.3 Flow Cytometry Analysis of Peripheral Blood from Xenografts

To harvest blood, mice were placed in a rotating tail injector where blood was collected by opening a tail vein. Around 50µl of blood was collected and stored in Microvette CD300 lithium heparin tubes. Blood was transferred to a 1.5ml Eppendorf where FACS antibodies were added, and the sample was incubated in the dark at room temperature for 20 minutes. After incubation, 1.2mls of 1X ammonium chloride red cell lysis solution was added to the sample and mixed. The Eppendorfs were then spun down at 300g for 5 minutes in a microcentrifuge

and washed twice with 1ml FACS buffer. Samples were then resuspended in 500ul of FACS buffer and analysed by flow cytometry.

2.3.7.4 Harvesting Xenograft Material

Organs, skull and bone marrow were dissected from sacrificed mice. The spleen was excised and weighed to judge potential engraftment. Kidneys and liver were also checked for engraftment by weight and the presence of white spots or discolouration on the surface of the organ. All organs were stored in sterile PBS after collection. Mouse heads were harvested for use by Dr. C. Halsey. Heads were detached, skin and eyes were removed and the samples were placed in formaldehyde solution for preservation and future analysis by Dr. Halsey's lab. Murine legs were collected for bone marrow material, both femurs and tibias were excised from flesh and flushed with sterile PBS to collect bone marrow. This was performed by removing the top and bottom of the stripped bones and inserting a 25 gauge needle attached to a 5ml syringe filled with PBS into the cut bone to flush the cells out. Flushed bone marrow was collected in 5ml bijou container prior to use.

For collection of engrafted cells, spleen/liver/kidneys were placed into 10cm sterile petri dishes and finely cut by scalpel to encourage release of cells. After mincing, the organ was washed with PBS and the resulting mix was passed through a 40µm cell strainer to remove larger clumps of tissue. These were forced through the strainer using the syringe top to increase cell yield. Bone marrow was also passed through a 40µm cell strainer to remove any bone fragments.

Harvested cells were centrifuged at 400 x g for 10 minutes and resuspended in PBS after which cell number was determined using a haemocytometer. Cells were frozen down using the Newcastle Protocol (Section 2.3.2.5). If required human cells were separated from murine cells by use of Ficoll-Paque PREMIUM solution. This solution works by density gradient centrifugation, separating out the higher density murine cells from human cells, specifically lymphocyte cells. 3ml of Ficoll-Paque was added to 15ml falcon tubes where 4ml of cell suspension was layered onto the solution using a glass Pasture pipette. This was done carefully to avoid mixing between the two solutions. The sample tubes were then centrifuged at 400 x g for 40 minutes at room temperature. After the spin step the

mononuclear human cells formed an isolated layer which was carefully harvested using a glass Pasture pipette. The cells were then washed with 10ml of cell culture media and centrifuged at 400 x g for 10 minutes. Cells were resuspended, density was determined using a haemocytometer and samples were frozen down using the Newcastle Protocol (Section 2.3.2.5) at the desired concentration.

2.3.8 Flow Cytometry

Flow cytometry was used to analyse cell surface marker expression on transduced cell populations to examine the impact of retroviral expression on cell lineage. The technique was also used to investigate engraftment of injected cells in xenograft samples.

Flow cytometry was performed using several different BD machines and fluorochrome panels. The machines used were FACSCanto II, Fortessa X20 and for cell sorting FACSAria II. All markers used and corresponding paired lasers are displayed in section 2.2.10.1.

2.3.8.1 Flow Cytometry Compensation Set Up

The use of multiple flow markers facilitates the necessity of compensation alignment to minimise the expression of false positives or negatives due to the overlap of fluorochrome emission spectra. Compensation set up was performed using Anti-Mouse Ig, κ/Negative Control Compensation Particles (BD Bioscience).

Two drops of Anti-Mouse Ig, κ beads and two drops of negative control beads were mixed in a Falcon 5ml round bottomed polystyrene test tube for each fluorochrome to be compensated, excluding CD33 PE-Vio770 which was incompatible with the Ig, κ beads due to the composition of the stain as a λ chain antibody. CD33 compensation was set up with CD34+ cells cultured for at least two weeks in standard liquid media (Section 2.3.2.3). ~ 1 x 10^6 of cells of a comparable morphology and marker expression of those to be tested were also stained with single antibodies, and were treated identically to the beads. An additional negative unstained cell control was prepared. Antibody concentrations for fluorochromes were ~0.20µg per 1 x 10^6 cells. Antibody stains were applied

and beads/cells were incubated in the dark for 15 minutes at room temperature or for 30 minutes at 4°C. Samples were washed with 1ml of FACS Buffer and centrifuged at 400 x g for 5 minutes. Supernatant was discarded and the wash step was repeated and supernatant was discarded. Samples were then prepared for compensation set up on the FACS machine of choice.

Negative and single fluorochrome cell preparations were used to adjust machine photomultiplier tube (PMT) voltages to ensure all desired cells were analysed. PMTs are optical detectors, which record the presence of fluorescence, converting photons to electrical signals as stained cells pass through the relevant lasers. Adjusting the voltage to the PMTs determines the intensity of the PMT signal. After PMT voltage had been adjusted a compensation wizard program provided with the FACS Diva suite was opened and prepared to capture the single antibody stains comprising the multicolour FACS panel. The compensation particles were analysed and data captured, gates were created to capture the positively stained Ig, κ beads, and the negative beads. This was repeated for each fluorochrome in the particular multicolour panel. CD33 compensation was performed using stained CD34+ cells as positive controls and unstained CD34+ cells as negative controls. After each stain was captured the FACS Diva program performed automatic compensation on the multicolour panel based on captured data, although compensation could also be adjusted at a later date if necessary. The experiment was saved, ready for use with experimental cells.

2.3.8.2 Analysis of Transduction Efficiency

Transduced cells were analysed using flow cytometry. EGFP expression representing MSCV-IRES-EGFP vector transduction and Thy1 expression representing MSCV-IRES-Thy1 vector transduction, which was tagged to the PE fluorochrome. Both markers were compared against positive and negative controls.

2.3.8.3 Analysis of Cell Surface Phenotype

In vitro cells were analysed weekly, up to 1 x 10^6 cells were isolated and aliquoted into individual pre-labelled 5ml polystyrene tubes. The cells were washed with 1ml of FACS Buffer and centrifuged at 400 x g for 5 minutes.

Supernatant was discarded and cells were stained with the relevant FACS panel (Supplementary Table 7.2, Supplementary Table 7.3) and 50µl of FACS Buffer. The cells were incubated in the dark for 15 minutes at room temperature or for 30 minutes at 4°C. Samples were washed with 1ml of FACS Buffer and centrifuged at 400 x g for 5 minutes. Supernatant was discarded and the wash step was repeated and supernatant was discarded. The samples were stored in the dark and used within 2 hours of preparation. Cells were run with a pre-prepared compensation set up (Section 2.3.8.1).

2.3.8.4 Thy1 Biotin-Streptavidin Labelling

Thy1 expressing cells were labelled with biotin conjugated CD90 antibody and stained with PE Streptavidin for visualisation using flow cytometry.

Cells were isolated and labelled with 0.5µl of biotin antibody in 20µl of FACS buffer per 200k-1 million cells, incubated for 30 minutes at 4°C, washed with FACS buffer, and pelleted by centrifuging at 1200g for 5 minutes, supernatant was discarded. Cells were resuspended with 0.5µl of PE Streptavidin in 20µl of FACS buffer and incubated for 30 minutes at 4°C, washed with FACS buffer, and pelleted by centrifuging at 1200g for 5 minutes. Supernatant was discarded and cells were analysed using flow cytometry.

2.3.8.5 Cell Sorting

Cells were treated identically to section 2.3.8.3, with the exception being that the amount of antibody used was raised to ~0.40µg per 1 x 10^6 cells. Cells were sorted on the FACSAria II. Single colour controls were analysed prior to sorting to optimise PMT voltage and compensation settings. Sorted cells were resuspended in RLT buffer and stored at -20°C before use.

2.3.8.5.1 Cell Fixation and DAPI Staining

Cells grown *in vitro* were analysed weekly, up to 1 x 10^6 cells were isolated and aliquoted into individual pre-labelled 5ml polystyrene tubes. Samples were washed with 1ml of FACS Buffer and centrifuged at 400 x g for 5 minutes, supernatant was discarded. Cells were then fixed using 2% Formaldehyde,

0.01% Triton X-100 solution. Cells were incubated at room temperature between 8-16 hours, cells were centrifuged at 400 x g for 5 minutes and supernatant was discarded. Samples were washed with 1ml of FACS Buffer and centrifuged at 400 x g for 5 minutes, supernatant was discarded. Samples were resuspended in 500µl of DAPI solution and were incubated in the dark at room temperature between 8-16 hours. Cells were centrifuged at 400 x g for 5 minutes, supernatant was discarded. The cells were then ready to be analysed on the flow cytometer. Cells were captured at a low flow rate for optimal results.

2.3.9 Quantitative Polymerase Chain Reaction

Quantitative PCR or Real Time PCR (qPCR) allows for the real time analysis of targeted products in a selected sample. qPCR can be used to test for relative expression levels of gene targets, gene copy number, or SNP detection and quantification.

The TaqMan platform uses a combination of probes, which bind to multiple points on exposed DNA/cDNA and specifically designed target primers. The probes contain two elements bound to the 5' and 3' end, at the 5' end is the reporter and the 3' end the quencher, when these two molecules are in close contact the quencher prevents the excitation of the reporter by absorbing light energy through fluorescence resonance energy transfer. The reporter can only produce signal when the probe is broken, separating the reporter and quencher. During a reaction, polymerase will begin elongation from a primer strand and upon contact with a probe will cleave the molecule separating the reporter from the quencher. The reporter can then fluoresce, free of the quencher, providing a permanent increase in fluorescence, which represents the doubling of the target sample. For gene expression profiling, a method termed relative quantification is used, gene expression in the selected sample is compared against the other samples in the assay. The comparison is performed by comparing the comparative threshold cycles (Ct) values. Ct denotes the PCR cycle number at which the fluorescence signal crosses the threshold barrier, or the point where significant and specific target amplification takes place.

The SYBR Green assay measures targeted product increase by binding of the reporter dye exclusively to double stranded DNA. The dye only fluoresces when

bound to double stranded DNA, and intensity increases with each cycle as more of the dye is bound. This requires high specificity when designing experiments as unspecific primer binding and primer dimers can create false positive results. The use of melting curves allows for the detection of multiple amplicons, suggesting unspecific binding in the assay. For copy number analysis with the SYBR green platform the absolute quantification method is used, where a standard curve is generated using known DNA dilutions, this can then be used to analyse unknown samples.

2.3.9.1 SYBR Green qPCR Primer Design

Gene sequences were identified using http://www.ncbi.nlm.nih.gov/ genome data base and primers were designed to cover potential transcript variant and exon-exon boundaries. Sequences were pasted into PrimerPlus3 to produce amplicons of 50-150bp, contain a G/C content between 30-80% and a melting temperature of 58-60^{o}C. The resulting primer sequences were then validated by BLAST and BLAT to determine specificity. Lysophilised primers were ordered from Sigma at a concentration of 100µmol, resuspended with dH2O and stored at -80^{o}C, working concentration at 10µmol was stored at -20^{o}C. Primer dimer formation was tested by investigating resulting melt curves, which indicate the presence of none specific product amplification. Upon the detection of multiple peaks, when investigating melt curves, different primer combinations were used when multiple primer sets for identical targets were available or different primer concentrations were tested. If both of these steps failed to produce one clean single peak of amplified material, new primers were designed.

2.3.9.2 SYBR Green qPCR Platform

SYBR Green master mix was created using reagents in Table 2.7 for each target gene, SYBR green contains ROX reference dye which reduces background noise by normalising non-PCR fluorescence variation. Samples were analysed in 384 well plates with 6ul of master mix pipetted per well and 4µl of 1ng patient DNA (4ng) added for a total of 10µl per well. Each well was repeated in triplicate. A standard curve is used to quantify unknown sample expression against a known set of standards, standard curves were generated for each primer target. A

standard master mix solution was used with dilutions of control DNA purchased from Promega. DNA dilutions used to create standard curves varied but the most common dilution combination was 100ng, 75ng, 50ng, 10ng, 7.5ng and 5ng. H_2O negative controls were created for each primer combination. When completed an adhesive cover was placed over the plate, which was then briefly centrifuged to remove air bubbles in individual wells. The plate was then placed into the ViiA 7 Real-Time PCR machine and analysed using the standard curve program settings.

Reagents	**Volume**
SYBR Green qPCR Master Mix	5.1µl
Forward Primer 10µM	0.2µl
Reverse Primer 10µM	0.2µl
H_2O	0.5µl

Table 2.7. Composition of SYBR Green master mix per well.

2.3.9.3 TaqMan Platform

TaqMan qPCR setup was highly similar to SYBR Green qPCR. Master Mix components differed (Table 2.8) and 5ng of cDNA was used for the sample template for a total of 20µl of solution per well. Experimental plates were created with triplicate wells per reaction and negative RNAse Free H_2O control wells per gene target. When placing into the ViiA 7 Real-Time PCR machine plates were analysed using the Delta CT program settings.

Reagents	**Volume**
Gene Expression qPCR Master Mix	10µl
Probe / Primer Mix	1µl
cDNA (1ng/µl)	5µl
RNAse Free H_2O	4µl

Table 2.8. Composition of TaqMan master mix per well.

2.3.10 Molecular characterisation of IGH-CEBP Patients

Only two patients had sufficient RNA of a high enough quality for further investigation. These two patients, 11739 and 23395, were analysed on the RNA-seq Illumina

HiSeq 2500 platform (requiring 200-400ng RNA with a RIN score >5) by Aros Applied Biotechnology (Denmark), and TaqMan qPCR platforms.

2.3.11 Protein Analysis

Protein analysis was performed to analyse the expression of proteins and protein isoforms in patient samples through use of Western Immunoblotting.

2.3.11.1 *Bradford* Assay – Coomassie Dye Based Protein Assay

Coomassie dye is used to quantify extracted protein lysate. The dye binds to the extracted protein resulting in a shift from the dyes standard absorbance maximum at 465nm (reddish brown) to 595nm (blue). The binding efficiency of the Coomassie dye ligands bound to protein is proportional to the number of positive charges on the protein. Generally the protein must be at least 3kDa in size to be bound by the dye.

Bradford assays were set up in 96 well plates. BSA provided with the assay was diluted into a series of standards according to Table 2.9 used to quantify protein samples, all standards were set up in triplicate. 250µl of Coomassie dye was mixed with 5µl of protein or standard sample. Protein samples were set up in duplicate due to limited availability. The plate was analysed within an hour of preparation, absorbance was measured at 595nm using the Fluostar Omega plate reader.

Tube	**Volume of Diluent**	**Volume and Source of BSA**	**Final BSA Concentration**
A	0	300µl of stock	2000µg/ml
B	125µl	375µl of stock	1500µg/ml
C	325µl	325µl of stock	1000µg/ml
D	175µl	175µl of vial B dilution	750µg/ml
E	325µl	325µl of vial C dilution	500µg/ml
F	325µl	325µl of vial E dilution	250µg/ml

G	325µl	325µl of vial F dilution	125µg/ml
H	400µl	325µl of vial G dilution	25µg/ml
I	400µl	0	0µg/ml

Table 2.9. Dilution scheme for Bradford Assay standards.

2.3.11.2 Western Immunoblotting

Western immunoblotting was performed to detect the presence and quantity of selected protein. This was achieved by first separating proteins by weight and 3-D structure by using polyacrylamide gel electrophoresis, upon separation proteins were transferred to a polyvinylidene fluoride (PVDF) membrane and probed by antibodies for the protein of interest.

2.3.11.2.1 Gel Electrophoresis

Amersham ECL Gel was briefly rinsed with dH2O and placed into an Amersham ECL Gel Box and then filled with 180ml 1x running buffer. The gel was run at 160V for 12 minutes, the gel comb was then removed and the wells filled with 6ml of 1x running buffer. Samples were then loaded into the gel with 10ul of Spectra Multicolor Broad Range Protein Ladder. Empty wells were filled with 30µl of urea buffer mixed with 5µl of loading buffer. The gel was run at 160V for 60 minutes.

2.3.11.2.2 Protein Transfer

The PVDF membrane was cut with clean scissors to match the size of gel, the membrane was activated by being soaked in 100% Methanol for one minute, after the incubation time the membrane was left submerged in 1x transfer buffer until needed.

Six pieces of filter paper matching the size of the gel and PVDF membrane were cut and soaked in 1x transfer buffer along with transfer sponges. Sponges were thoroughly squeezed and soaked in transfer buffer until all air bubbles had been removed.

After the gel had completed running it was detached from the plastic cassette and the stacking gel was removed. Keeping the gel moist, the PVDF membrane was carefully placed on the gel and sandwiched between the soaked filter paper and

sponges (Figure 2.3). This stack was placed into a transfer cassette and a protein gel tank, making sure that the membrane faced the positive electrode. The tank was filled with 1x transfer buffer. The tank was either placed in a polystyrene box filled with ice or an ice pack was placed in the tank to maintain a low temperature during protein transfer. The samples were transferred by running the power pack at 100V for 60 minutes.

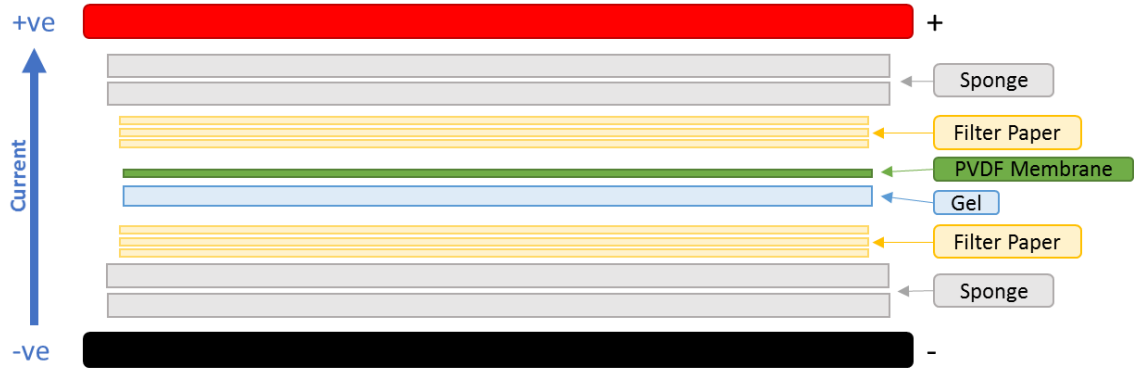


Figure 2.3. Western immunoblot set up for protein transfer to PVDF membrane adapted from abcam western blotting beginners guide.

2.3.11.2.3 Immunoblotting

After the run had finished, successful transfer was confirmed by viewing the transmission of the multi-coloured protein ladder to the PVDF membrane. The membrane was then blocked in 5% skimmed milk powder solution for at least 30 minutes at room temperature or overnight at 4°C on a rocking platform. The membrane was washed three times with TBST solution and stored in TBST to prevent it from drying out. The membrane was then cut appropriately using the protein ladder as a guide to separate the membrane containing the loading control protein from the target protein. Antibodies were diluted as displayed in Section 2.2.9.2. Membrane sections were placed in separate 50ml falcon tubes with 4mls of 50% 5% skimmed milk solution (2.5% after dilution) and 50% TBST solution with the respective diluted antibody. The falcons containing the membranes and antibody mix was incubated at room temperature for one hour on a roller mixer, or for up to 18 hours at 4°C. The membranes were then washed a further three times in TBST solution and the incubation process was repeated with the secondary antibody for one hour, followed by the same wash steps.

Finally to visualise the proteins the PVDF membrane was stained with Amersham Enhanced Chemiluminescence Prime Western Blotting Detection Reagent. This was performed by mixing 500ul of reagent A and B and adding the mixture to the membrane, and leaving at room temperature for one minute. After the incubation period excess solution was carefully removed by tipping off the membrane onto tissue paper. The membrane was then placed into a plastic folder, air bubbles were carefully removed and the blot was inserted into the G:Box for imaging and analysed using ImageJ software (Version 1.48).

2.3.11.2.4 Protein Stripping

Some membranes were stripped from their original antibodies due to the necessity of re-probing with different target antibodies, or the need to visualise a stronger antibody signal. The selected PVDF membrane was washed with TBST solution and immersed in Restore™ Western Blot Stripping Buffer (Thermo Scientific) for one hour at room temperature in a shaker. After stripping the membrane was washed a further three times with TBST solution and blocked in 5% milk solution before being incubated with the desired primary antibody, following the standard western immunoblotting protocol (Section 2.3.7.1.3).

Chapter 3 Genetic Characterisation of *IGH-CEBP* translocated B-cell precursor acute lymphoblastic leukaemia

3.1 Introduction

Understanding the genetic basis of ALL has led to the identification of novel subgroups. These findings have led to improved survival by stratification of these subgroups according to risk and the identification of novel therapeutic targets. Further, determining incidences of commonly occurring abnormalities has given scope for statistical analysis with added insights into patient outcome.

The development of new sequencing and array technologies has facilitated the screening of both known and novel genetic lesions by improving detection rates, lowering costs, or reducing the amount and quality of patient sample needed. These advances have led to the identification of the *BCR-ABL1*-like/Ph-like subgroup, (Den Boer *et al.*, 2009; Harvey *et al.*, 2010b). A group characterised by *JAK* pathway genetic lesions, including *CRLF2* and *JAK* translocations, and by kinase signalling lesions including *ABL1* and *ABL2*. This discovery has opened the possibility of the use of *JAK* inhibitors and TKIs for treatment, and has been tested successfully *in vitro* and *in vivo* (Roberts *et al.*, 2014a). This success for TKIs has now been demonstrated in patients carrying tyrosine pathway abnormalities, who are refractory to conventional chemotherapy (Eyre *et al.*, 2012; Weston *et al.*, 2013).

Refinement of risk stratification in low risk ALL subgroups has also been fruitful. Recently the use of MLPA, FISH and cytogenetic data, in combination with statistical analysis, identified subgroups with low risk BCP-ALL who may benefit from treatment deintensification (Moorman, 2012b). Such examples highlight the importance of further characterising both known and unknown subtypes of ALL.

3.2 Aims

The *IGH-CEBP* subgroup has been previously identified, but limited patient numbers have made full genetic characterisation and meaningful statistical analysis difficult to perform. In this chapter an *IGH-CEBP* cohort was created and genetically characterised by achieving the following aims:

- To define a cohort of *IGH-CEBP* positive patients by searching the CIMS database for patients with karyotypes showing translocations involving the chromosomal regions of *IGH* and the *CEBP* partner genes and confirm the translocations using custom designed FISH probes.
- To screen the cohort for commonly associated BCP-ALL genetic lesions.
- To screen the cohort for novel genetic lesions.
- To determine whether *IGH-CEBP* translocations were primary or secondary genetic insults in BCP-ALL development.
- To investigate common clinical characteristics of the cohort.

3.3 Results

3.3.1 ***Creating a Patient Cohort***

3.3.1.1 Database Search and FISH testing

The LRCG CIMS database was mined for karyotypes with cytogenetic breakpoints corresponding to known *IGH-CEBP* translocations (Supplementary Table 7.4). Patients with the appropriate karyotypes, with available fixed cells, were initially tested for the presence of an *IGH* translocation using the LSI *IGH* Dual Colour Break Apart Rearrangement Probe (Vysis). Eleven patients with verified *IGH* translocations were then tested with custom made *CEBP* FISH probes to confirm *IGH-CEBP* translocations. These patients were added to a previously established *IGH-CEBP* cohort of 19 patients created by Dr. L.J. Russell (Akasaka *et al.*, 2007) (Supplementary Table 7.5). Three patients, 25458, 25505 and 25541, were added to the cohort on cytogenetic data alone as no fixed cells were available for FISH analysis (Table 3.1), providing a total of 33 patients for study.

Patient Number	Age / Sex	WBC x 10^9L	Karyotype	IGH FISH %	CEBP FISH %	Interpretation
CEBPA (19q13.1)						
1798	10/M	5.9	46,XY,t(14;19)(q32;q13)[8]	93% [1R1G1F]	56% [1R1G1F]	Gene Rearrangement
4175	11/F	44.7	46,XX,t(14;19)(q32;q13)[4]	86% [1R1G1F]	59% [1R1G1F]	Gene Rearrangement
4198	28/F	6.6	46,XX,t(14;19)(q32;q13),add(15)(q26)[5]	81% [1R1G1F]	72% [1R1G1F]	Gene Rearrangement
4774	19/F	4.2	46,XX,t(14;19)(q32;q13)[9]	45% [0R1G2F]	73% [1R1G1F]	Gene Rearrangement
7143	44/F	5	57,XX,+X,+X,+4,+6,+10,+14,t(14;19)(q32;q13),ins(15)(q15),+17,+18,+21,+21,+mar[3]	82% [1R1G2F]	80% [1R1G1F]	Gene Rearrangement
7617	12/F	70.8	46,XX,dup(3)(q21q27),t(14;19)(q32;q13)[6]	97% [1R1G1F]	94% [1R1G1F]	Gene Rearrangement
24880	15/M	32.6	46,XY,t(14;19)(q32;q13)[3]	84% [1R1G1F]	77% [1R1G1F]	Gene Rearrangement
25505	52/M	1.3	47,XY,+X,t(14;19)(q32;q13)[6]/46,XY[2]	93% [1R1G1F]	FISH not performed (no fixed cells)	NA
25855	14/F	0.9	46,XX,t(14;19)(q32;q13)[5]	60% [1R1G1F]	60% [1R1G1F]	Gene Rearrangement
25952	55/M	9.87	46,XY,t(14;19)(q32;q13)[2]/46,XY,add(4)(p1),add(5)(q1),t(14;19)(q32;q13)[8]	91% [1R1G1F]	82% [1R1G1F]	Gene Rearrangement
CEBPG (19q13.1)						
11540	10/M	3.1	46,XY,t(14;19)(q32;q13)[18]	98% [1R1G3F]	92% [1R0G2F]	Other abnormal
CEBPB (20q13)						
3455	15/M	3	46,XY,inv(9)(p?13q?32),t(14;20)(q?32;q?13)[14]	83% [1R1G1F]	88% [1R1G1F]	Gene Rearrangement
5588	43/F	36.9	48,XX,+8,+mar,inc[cp13]	61% [1R1G1F]	70% [1R1G1F]	Gene Rearrangement
5632	13/F	103	47,XX,t(14;20)(q32;q13),add(21)(p1),+add(21)(p1)[9]	77% [1R1G1F]	85% [1R1G1F]	Gene Rearrangement
10859	34/F	80.6	46,XX,t(9;22)(q34;q11)[1]/ 46,s!,t(14;20)(q32;q13)[7]	93% [1R1G1F]	84% [1R1G1F]	Gene Rearrangement
11682	30/F	10.1	46,XX,-5,add(14)(q3?2),+mar[2]/ 46,XX[6]	76% [1R1G1F]	58% [1R1G1F]	Gene Rearrangement
11739	14/F	430	47,XX,t(5;14)(q175;q32;?3),add(9)(p?13),der(20)t(5;20)(q175;q13),+mar[19]/ 46,XX[1]	85% [1R1G1F]	34% [0R1G2F]	Other Abnormal
25458	31/M	1.2	46,XY,t(9;22)(q34;q11.2),t(14;20)(q32;q13)[7]/46,idem,add(19)(p13)[3]	66% [2R2G2F]	FISH not performed (no fixed cells)	NA
25686	59/F	1.5	46,XX,der(2)t(2;5)(q37;q12?3),add(10)(q23),t(14;20)(q32;q13)[5]/46,XX[6]	77% [1R1G1F]	52% [1R1G1F]	Gene Rearrangement
CEBPD (8q11)						
2734	5/M	2.8	47,XY,t(8;14)(q11;q32),+21c[5]/ 47,XY,+21c[23]	37% [1R1G3F]	54% [1R1G1F]	Gene Rearrangement
3622	9/M	6.4	46,XY,t(8;14)(q11;q32),?del(18)(p11.2p11.3),der(19)t(8;19)(?;p13)[9]	72% [1R1G1F]	73% [1R1G1F]	Gene Rearrangement
3759	15/F	4.6	46,XX,t(8;14)(q11;q32)[2]	76% [1R1G1F]	82% [1R1G1F]	Gene Rearrangement
6889	8/M	47.9	47,XY,+4,t(8;14)(q11;q32)[7]	97% [1R1G1F]	85% [1R1G1F]	Gene Rearrangement
19734	12/M	21.9	47,XY,der(14)t(8;14)(q11;q32),+21c[5]/47,XY,+21c[5]	42% [1R0G3F]	29% [0R0G3F]	Other Abnormal
19794	17/F	14.8	47,XX,der(14)t(8;14)(q11.2;q32),+21c[6]	78% [1R0G3F]	85% [0R1G2F]	Other Abnormal
20580	18/F	1.8	46,XX,t(8;14)(q11;q32)[3]/46,XX[20]	74% [1R1G1F]	76% [1R1G1F]	Gene Rearrangement
22355	6/F	20	46,XX,t(8;14)(q11;q32)[8]	100% [1R1G1F]	77% [1R1G1F]	Gene Rearrangement
23168	8/F	20.6	44,XX,der(3)(3qter->3p25::21q11->21q22.3::21q22.3->21pter)(-8,der(14)t(8;14)(q11;q32)[9]	80% [1R0G1F]	94% [0R1G1F]	Other Abnormal
23395	13/M	Unknown	47,XY,t(8;14)(q11;q32),+21c[10]	61% [1R1G1F]	27% [1R0G2F]	Other Abnormal
25541	8/M	24.8	47,XY,t(8;14)(q11;q32),+21c[10]	90% [1R1G1F]	FISH not performed (no fixed cells)	NA
CEBPE (14q11)						
7247	45/M	24.3	46,XY,der(2)t(1;2)(q1?;q37),inv(14)(q11q32),inc[3]/ 46,XY[4]	61% [1R1G1F]	57% [1R1G1F]	Gene Rearrangement
23567	2/F	2	46,XY,inv(14)(q11q32)[2]/46,XY[18]	30% [1R1G1F]	13% [1R1G1F]	Gene Rearrangement
27181	65/M	1.2	46,XX,inv(14)(q11q32)[7]	43% [1R1G1F]	61% [1R1G1F]	Gene rearrangement

Table 3.1. IGH-CEBP cohort showing age, sex, WBC and FISH results for IGH and corresponding CEBP partner gene.

3.3.2 ***Clinical and Demographic Features of the IGH-CEBP Cohort***

Over the course of the study, 14 additional patients were identified to possess *IGH-CEBP* translocations of which the majority displayed classic *IGH-CEBP* breakpoints. Six patients had variant breakpoints and are discussed further is section 3.3.3. Patient data from the 33 patients were statistically analysed using the Fisher exact test, used to derive significance from small sample numbers.

The final cohort consisted of 11 (34%) *CEBPD*, 10 (30%) *CEBPA*, 8 (24%) *CEBPB*, 3 (9%) *CEBPE* and 1 (3%) *CEBPG* patient (Table 3.1). The median age for the cohort was 15 years (range of 2 – 65 years). There were significant differences in age between *CEBP* subgroups, *IGH-CEBPD* patients comprised the majority of patients under the age of 10 years, while *CEBPA* and *CEBPB* patients were older (p=0.005) (Table 3.1) (Figure 3.1) (Supplementary Table 7.6). These findings remained significant both with (p=0.02) and without (p=0.04) the inclusion of DS patients in the *CEBPD* subgroup. There was a slight prevalence of female patients: 18 females to 15 males. The median WBC was low at 8.24 x 10^9/L, (range of 0.9 - 430 WBC x 10^9/L) (Supplementary Table 7.7). The highest median WBC was observed in the *CEBPB* subgroup at 23.5 x 10^9/L in comparison to *CEBPA* 6.25 x 10^9/L, *CEBPD* 17.40 x 10^9/L, and *CEBPE* 2 x 10^9/L (p=0.05) (Figure 3.2).

There was no consistent association with other established primary abnormalities, with the exception of DS. The abnormalities identified were t(9;22)(q34;q11) translocation (n=1, 10859), HeH (n=1, 7143), hypodiploidy (n=1, 23168), and DS (n=5, 2734, 19734, 19794, 23395 and 22541). All DS-ALL patients were in the *IGH-CEBPD* subgroup, supporting previous findings (Lundin *et al.*, 2009) (Table 3.1). Whole chromosomal gains were also rare and non-recurrent, with only three patients affected; 6889 (+4), 5588 (+8), and 25505 (+X).

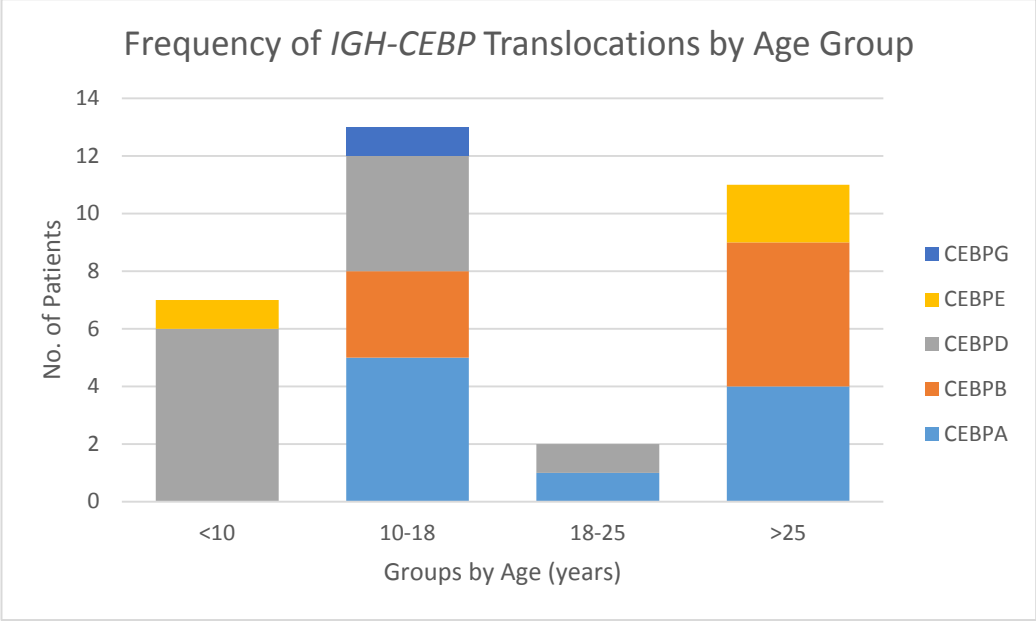


Figure 3.1. IGH-CEBP translocated patients divided by age group (P=0.005). Distribution shows CEBPD patients comprising the majority of the under 25 age group, while other subgroups showed greater variation.

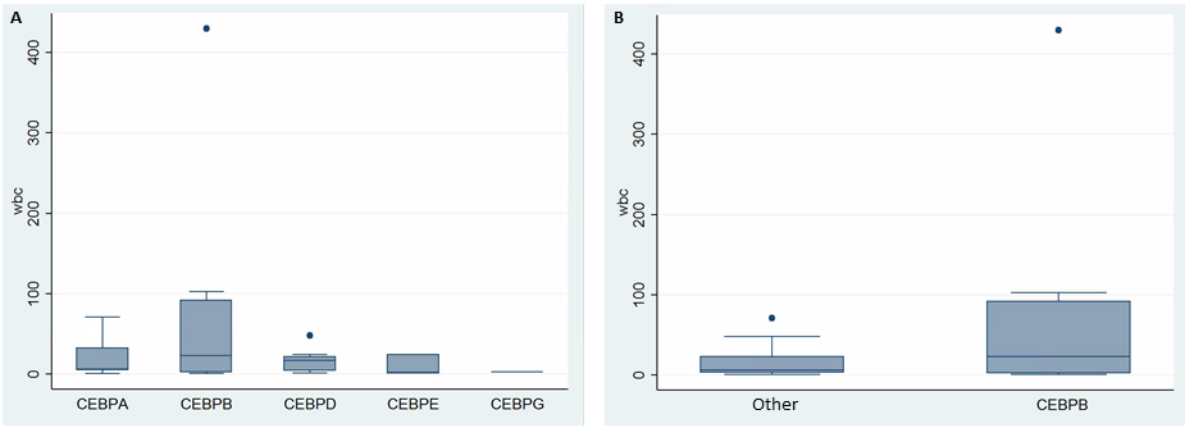


Figure 3.2. Box plots generated using patient WBC data, dots indicate outliers. CEBPB patients have a broader range of WBCs than other CEBP patients.

A. WBC separated by CEBP subgroup (P=0.5). B. WBC comparison between CEBPB subgroup against all other CEBPs (P=0.08).

Available clinical data were collected for all *IGH-CEBP* patients and analysed to identify potential trends within the cohort. Nine (27%) patients were deceased, these patients displayed no age or gender bias and included almost all *CEBP* genes. Mortality was highest in the *CEBPB* subgroup, with 4/8 (50%) patients deceased, followed by *CEBPE* with 1/3 (33%), *CEBPA* 2/10 (20%), and *CEBPD* 2/11 (18%). The most common cause of death was infection, recorded in 4/9

(44%) patients (Table 3.2). Five relapses occurred in the cohort (15%); three of which occurred *CEBPD* patients, which included one DS patient. Two of these patients subsequently died. The remaining relapses occurred in *CEBPA* and *CEBPB* translocated patients. All relapses occurred in the bone marrow, excluding patient 6889, whose site of relapse is unknown (Table 3.2). MRD data at day 28 were available for 8/33 (24%) patients; *CEBPD* (n=3), *CEBPA* (n=2), *CEBPB*, *CEBPE* and *CEBPG* (n=1 each). Although numbers were low, an interesting trend was identified in that all three MRD negative patients had *CEBPD* involvement, while the remaining five MRD positive patients comprised all other *CEBP* subgroups (p=0.09) (Table 3.2).

Patient ID	Age / Sex	WBC x 10^9L	Status	Cause of Death	Relapse	MRD Status
CEBPA (19q13.1)						
1798	10/M	5.9	Alive	NA	No	Unknown
4175	11/F	44.7	Alive	NA	No	Unknown
4198	28/F	6.6	Deceased	Infection / ALL	No	Unknown
4774	19/F	4.2	Alive	NA	No	Unknown
7143	44/F	5	Alive	NA	No	Unknown
7617	12/F	70.8	Alive	NA	No	Positive
24880	15/M	32.6	Alive	NA	No	Positive
25505	52/M	1.3	Alive	NA	Yes 1 - BM	Unknown
25855	14/F	0.9	Alive	NA	No	Unknown
25952	55/M	9.87	Deceased	Infection / ALL	No	Unknown
CEBPG (19q13.1)						
11540	10/M	3.1	Alive	NA	No	Positive
CEBPB (20q13)						
3455	15/M	3	Alive	NA	No	Unknown
5588	43/F	36.9	Deceased	Post BMT toxicity	No	Unknown
5632	13/F	103	Alive	NA	No	Unknown
10859	34/F	80.6	Deceased	ALL - Died in remission	No	Unknown
11682	30/F	10.1	Alive	NA	Yes 1 - BM	Unknown
11739	14/F	430	Deceased	ALL	No	Positive
25458	31/M	1.2	Alive	NA	No	Unknown
25686	59/F	1.5	Deceased	ALL	No	Unknown
CEBPD (8q11)						
2734	5/M	2.8	Deceased	ALL - 2nd Relapse	Yes 2 - BM	Unknown
3622	9/M	6.4	Alive	NA	No	Unknown
3759	15/F	4.6	Deceased	Infection / ALL	Yes 1 - BM	Unknown
6889	8/M	47.9	Alive	NA	Yes 1 - Unkown	Unknown
19734	12/M	21.9	Alive	NA	No	Unknown
19794	17/F	14.8	Alive	NA	No	Negative
20580	18/F	1.8	Alive	NA	No	Negative
22355	6/F	20	Alive	NA	No	Unknown
23168	8/F	20.6	Alive	NA	No	Negative
23395	13/M	Unknown	Alive	NA	No	Unknown
25541	8/M	24.8	Alive	NA	No	Unknown
CEBPE (14q11)						
7247	45/M	24.3	Alive	NA	No	Unknown
23567	2/F	2	Alive	NA	No	Positive
27181	65/M	1.2	Deceased	Infection / ALL	No	Unknown

Table 3.2. Relapsed and deceased patients in the IGH-CEBP cohort. Table shows cause of death and relapse status, with the relapse column showing the number of relapses, and site of relapse. Acute Lymphoblastic Leukaemia (ALL), Bone Marrow (BM), Bone Marrow Transplant (BMT), Female (F), Male (M).

3.3.3 *Other Abnormal Patients*

During screening, six patients were found to display more complex signal patterns requiring further investigation to determine *IGH-CEBP* involvement. These patients were classified as other abnormal (Table 3.3). Patient 11540 was not analysed further as *IGH-CEBP* involvement had been previously proven (Akasaka *et al.*, 2007). Patient 25458 was also omitted as no fixed cells were available for FISH.

Patient ID	Karyotype	FISH Probe	FISH Signal	Population
11739	47,XX,t(5;14)(q1?5;q32.?3),add(9)(p?13),der(20)t(5;20)(q1?5:q13),+mar[19]	*IGH*	1R1G1F	85%
		CEBPB	0R1G2F	34%
			1R1G1F	6%
			1R1G2F	5%
19734	47,XY,der(14)t(8;14)(q11;q32),+21c[5]/47,XY,+21c[5]	*IGH*	1R0G1F	42%
			1R1G1F	21%
		CEBPD	0R0G2F	67%
			0R0G3F	29%
		8 Alpha Satelite (green)	0R2G0F	80%
			0R3G0F	15%
19794	47,XX,der(14)t(8;14)(q11.2;q32),+21c[6]	*IGH*	1R0G1F	78%
		CEBPD	0R1G2F	85%
		8 Alpha Satelite (green)	0R2G0F	94%
		IGH-CEBPD	2R2G1F	34%
			1R2G1F	32%
			1R1G1F	30%
23168	44,XX,der(3)(3qter->3p25::21q11->21q22.3::21q22.3->21pter)c,-8,der(14)t(8;14)(q11;q32)[9]	*IGH*	1R0G1F	75%
		CEBPD	0R1G1F	94%
		14q Subtelomere (red)	1R0G0F	57%
			2R0G0F	41%
		IGH-CEBPD	1R1G1F	44%
			2R2G1F	25%
			2R1G1F	22%

Table 3.3. Table displaying patients classed as other abnormal due to more complex Karyotype and FISH data. Table shows multiple FISH tests performed.

3.3.3.1 Patient 11739

Patient 11739 displayed a complex karyotype with numerical and structural abnormalities. The main *CEBPB* FISH population (34%) in this sample showed a signal pattern of 0R1G2F with the *CEBPB* breakapart probe, the extra green

signal indicating a break within the centromeric probe (Table 3.3). Further FISH analysis of this patient was halted as DNA had been sent for WGS as part of a separate project.

Data gained from the WGS project showed a high level of complexity with a series of translocations between *IGH* and several regions on chromosomes 5 and 20 observed (*Figure 3.3*A).

Translocations of interest included *IGH-CEBPB*, where the observed breakpoint was more centromeric than the standard *IGH-CEBPB* translocation, causing a split within the green section of the *CEBPB* probe giving the 1G0R2F signal pattern (Table 3.3). However NGS data showed this translocation to be more complicated, with the *IGH* involved in several juxtapositions between *CEBPB* and a section of 5q14.3 (two genes of interest in this region are *MEF2C* a gene involved in myogenesis and *CETN3* important in microtubule organisation) essentially creating two *IGH* translocations in this patient. The presence of two individual *IGH* translocations supports the discrepant FISH analysis between the *IGH* probe, which was found to be translated in 85% of cell nuclei, and *CEBP* probe where translocations were found to comprise ~45% of cell nuclei. The remaining 40% of *IGH* translocations may represent the 5q14.3 translocation.

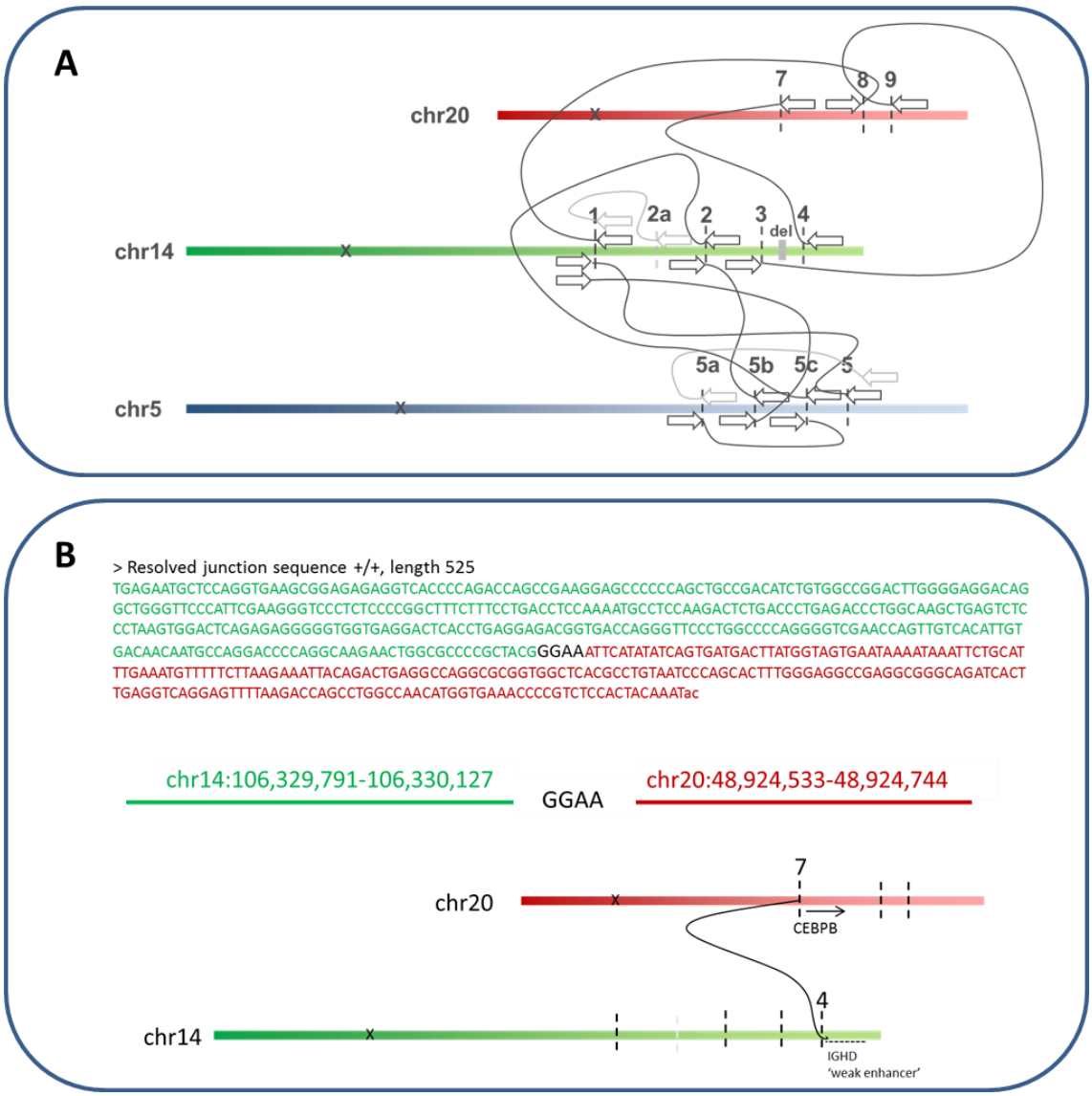


Figure 3.3. Mechanism of CEBPB deregulation in patient 11739.

Breakpoint identified by NGS of patient 11739 confirming the IGH-CEBPB breakpoint and the cytogenetic break within the CEBPB probe. A. Visualisation of NSG data which indicated a complex series of rearrangements between chromosome 14 and chromosomes 5 and 20. B. NSG data indicating the juxtaposition of the IGH and CEBPB genes. Green text indicates IGHD sequence while red text indicates CEBPD sequence. Figure created by Illumina.

3.3.3.2 Patient 19734

The t(8;14)(q11.2;q32)/*IGH-CEBPD* translocation was identified in patient 19734 by conventional cytogenetic analysis. *IGH* FISH identified two main signal patterns, 1R0G1F (34%) indicating a loss of the telomeric green signal, either through loss of a derived chromosome or natural deletions occurring through V(D)J rearrangements, and 1R1G1F (6%) indicating a split within the *IGH* locus (Table 3.3). The *CEBPD* probe however presented no visible translocation, with signal patterns of 0R0G2F (67%) and 0R0G3F (29%). Chromosomal copy

number was first investigated due to the gained fusion signal of *CEBPD*, a chromosome 8 alpha satellite centromeric probe (green) (Qbiogene) was used, displaying the predominant population with a normal signal pattern, 0R2G0F (80%). There was however a small population of 0R3G0F cells (15%), indicating a gain of chromosome 8 (Table 3.3).

Available SNP array data for this patient was analysed, showing two copies of chromosome 8 with a gain of the majority of the 8q arm (copy number 3). The breakpoint for this arm occurred at 8q11.21, encompassing the *CEBPD* gene (Figure 3.4 C). Other CNAs included a small region of the 14q arm which was lost (copy number 1). As this patient had both diagnostic and remission SNP data available rearrangements were compared and confirmed to be somatic.

The gain of this specific region of chromosome 8, linked with loss of 14q, the observations of a population of cells with three copies of chromosome 8 alpha satellite signals, and the loss of the *IGH* telomeric signal in a population of cells suggested the following;

During the evolution of this leukaemia, a copy of chromosome 8 was gained, from this population emerged a lineage of cells in which an *IGH-CEBPD* translocation occurred. This translocation involved a region more centromeric than the designed *CEBPD* break apart probe, resulting in the probe being translocated whole onto chromosome 14, this was supported by the SNP data. This translocation put a distance of 322kb between the *CEBPD* gene and the *IGH* promoter (Figure 3.4 A&C). This population of cells lost the derived chromosome 8, resulting in an altered copy number profile for this patient, where only a region of the chromosome 8q arm remained. Therefore the presence of the derived chromosome 14 shows the copy number of chromosome 8 as being at 2 with the exception of the 8q arm which remained on the derived chromosome 14, and a loss of 14q which was lost with the derived chromosome 8. This is supported by the small population of cells with three copies of the chromosome 8 alpha satellite, and the loss of the telomeric *IGH* break apart probe signal.

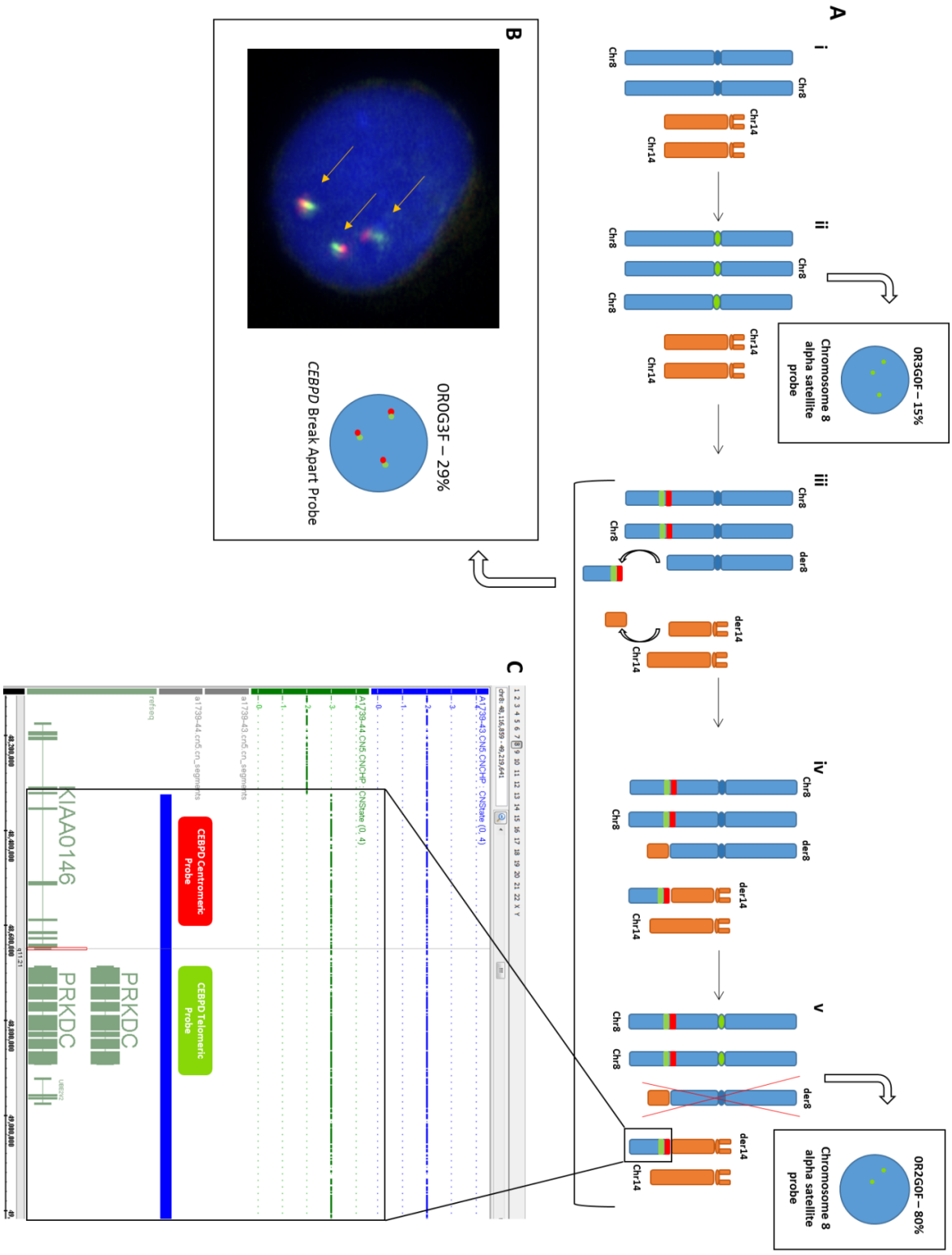


Figure 3.4. Mechanism of CEBPD deregulation in patient 19734.

A.i. Beginning with a normal nucleus with two copies of chromosome 8 and 14 respectively. Aii A small population of cells were found to have three copies of the chromosome 8 alpha satellite probe signal A.iii. A breakpoint on chromosome 14, and a centromeric breakpoint on chromosome 8 take place, failing to split the CEBPD signal, the 0R0G2F signal pattern is seen. A.iv. The chromosomal segments reattach to the respective derived chromosomes. A.v. During clonal evolution of the sample the derived chromosome 14 is lost and an extra chromosome 8 is acquired, giving the FISH signal pattern of 0R0G3F. B. Image of a 0R0G3F FISH pattern in a patient nucleus. C. Matched diagnostic (green) and remission (blue) SNP array data for patient 19734 showing gain of 8q, including the specific section of chromosomal material

encompassing the CEBPD gene translocated to the der (14), CEBPD FISH probe positions have been placed onto the image to display the centromeric break point of the patient.

3.3.3.3 Patient 19794

The t(8;14)(q11.2;q32)/*IGH-CEBPD* translocation was identified in patient 19794 by conventional cytogenetic analysis . *IGH* FISH showed the main pattern to be 1R0G1F, indicating a loss of the telomeric green signal, either through loss of derived chromosome or deletions occurring during V(D)J rearrangement (Table 3.3). The *CEBPD* probe showed the main clonal population to be 0R1G2F in 85% of cells. As a gain of the green FISH signal was observed the next step was to elucidate if this gain was chromosomal, confined to the gene, or a break within the green probe itself splitting the telomeric probe and giving an extra green signal (Table 3.3). Chromosome 8 copy number was first assessed using alpha satellite centromere 8 probe (green) (Qbiogene), which displayed 0R2G0F in 94% of cells, excluding gain of chromosome 8. Thus the possibility remained of an altered breakpoint. Rather than attempting to map the potential breakpoint, an *IGH-CEBPD* fusion probe was created to investigate the presence of the fusion gene (Figure 2.1). The signal pattern of this probe displayed three prevalent populations all with a fusion signal confirming the rearrangement (Table 3.3). The probe also showed additional red and green signals, suggesting potentially differing *IGH* and *CEBPD* breakpoints within the same patient.

3.3.3.4 Patient 23168

CEBPD FISH of patient 23168 showed a pattern of 0R1G1F in 94% of cells indicating a deleted centromeric signal (Table 3.3). The karyotype of this patient showed losses of chromosomes 8 and 14 with only the derived chromosome 14 present. To confirm the presence of the fusion, the *IGH-CEBPD* fusion probe was used (Figure 2.1) giving three prevalent populations confirming the translocation (Table 3.3). A 14q subtelomeric probe was used to check chromosome copy number confirming loss of chromosome 14 in 57% of blasts (Table 3.3). Figure 3.5 indicates the possible mechanism in this patient.

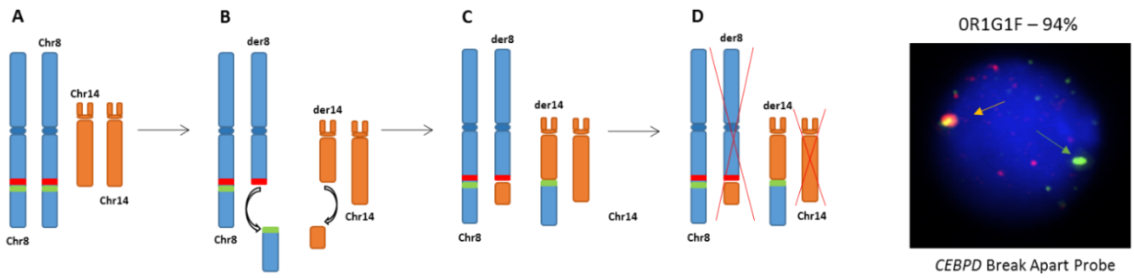


Figure 3.5. Potential mechanism leading to the translocation of CEBPD gene in patient 23168.

A. The pattern observed in normal patient nuclei, a signal pattern of 0R0G2F where the red and green FISH signals are expressed next to each other, with the CEBPD gene located between, this pattern indicates no translocation events. B. The start of the translocation event, sections of chromosome 8 which contains the CEBPD gene and the green centromeric FISH probe, and chromosome 14 detach from their respective chromosomes. C. The detached sections attach to the opposite chromosomes, giving a derived chromosomes 8 and 14, bringing the CEBPD gene into contact with the IGH locus and giving the FISH signal pattern of 1R1G1F. D. During clonal evolution of the sample the derived chromosome 8 and chromosome 14 are lost giving the FISH signal pattern of 0R1G1F, as shown.

3.3.4 ***MLPA Screening***

To identify commonly occurring genetic lesions, patients with sufficient DNA (n=28) were screened by MLPA with the SALSA MLPA P335 ALL *IKZF1* kit (Supplementary Table 7.1. Chromosomal copy number abnormalities based on karyotype were considered when analysing MLPA data.

CNAs between *CEBP* subgroups varied in both number of patients affected and number of alterations per patient. All patients in the *CEBPB* subgroup had CNA of the genes tested (6/6). Three out of seven *CEBPA* (43%), 3/11 *CEBPD* (23%) and 1/3 (33%) *CEBPE* patients also showed CNA of these genes (Figure 3.6) (Supplementary Table 7.8, Supplementary Table 7.9).

The *CEBPB* subgroup, showed 12 CNA within 6/6 patients, CEBPA with 12 CNA in 3/7 (43%) patients, followed by the *CEBPD* subgroup with 7 CNA in 3/11 (23%) patients, *CEBPE* subgroup had 1 CNA in 1/3 (33%) patients with *CEBPG* showing no CNA (Figure 3.7). Overall 15/28 (54%) patients displayed no CNA, four of which were DS patients (Figure 3.7). Incidence of CNA per patient was highest in *CEBPB* at 2 CNAs per patient, *CEBPA* at 1.7, *CEBPD* at 1.5 CNAs, *CEBPA* at 1.2 CNAs, and *CEBPE* 1 CNA per patient.

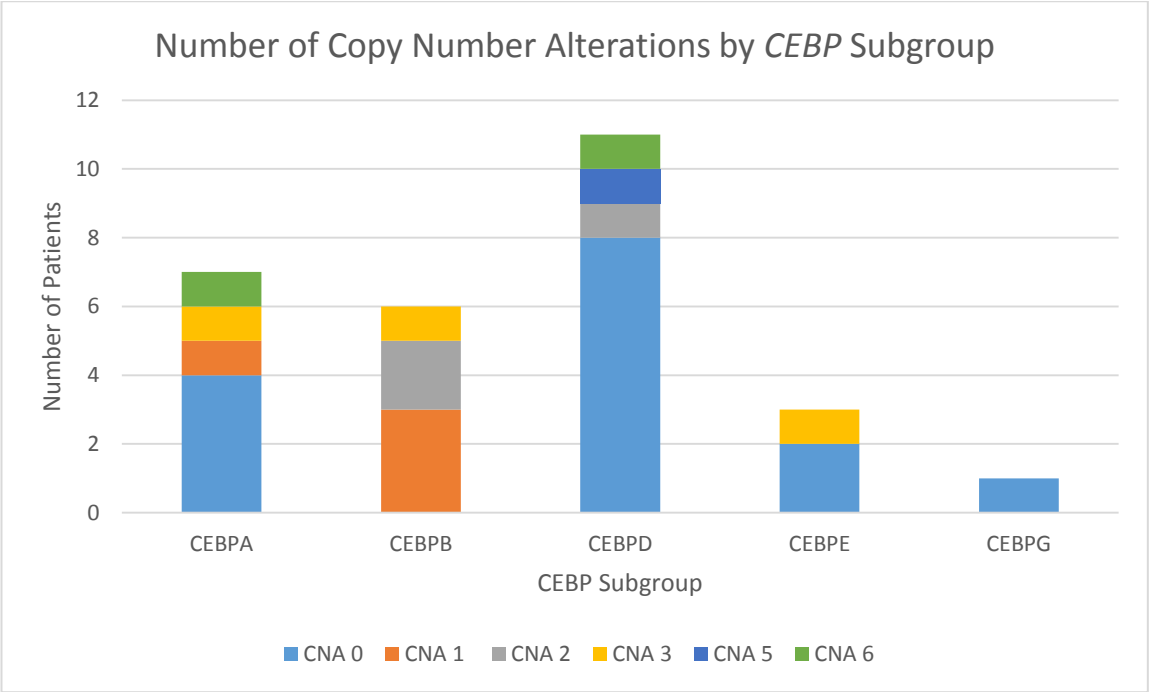


Figure 3.6. Graph showing incidence of copy number alterations by CEBP subgroup (P=0.05).

The CEBPB subgroup displayed CNAs in all patients, however the number of CNAs per patient was low, with the majority showing one or two aberrations. CEBPD conversely showed a low number of patients with CNAs, however patients affected had high numbers of aberrations, with two patients with 11 CNAs between them. CEBPA and CEBPD patients displayed a similar profile.

The most common CNA identified were whole gene deletions of *CDKN2A*/*CDKN2B* in 7/28 (25%) patients, and *IKZF1* deletions in 6/28 (21%), all of which were focal deletions of exons 4-7 resulting in the creation of the dominant negative IK6 protein isoform. These *IKZF1* deletions were found in only two subgroups, 4/6 (67%) in *CEBPB* patients and 2/6 (33%) in *CEBPD* patients (P=0.04) (Figure 3.7). CNA of *PAX5* occurred in 4/28 (14%) patients, two of these patients showed gains of the gene with exon 5 gained in patient 7247, and exons 7 and 8 gained in patient 3455. These gains are of interest as single exon gains have been predicted to lead to altered PAX5 protein activity (Familiades *et al.*, 2009; Schwab *et al.*, 2010b). The remaining two patients showed one whole gene deletion in patient 23168, and loss of exon 1 in 1798. Other CNA included whole gene deletions of *ETV6* in 3/28 (11%) of patients, and several other non-recurrent CNA (Figure 3.7).

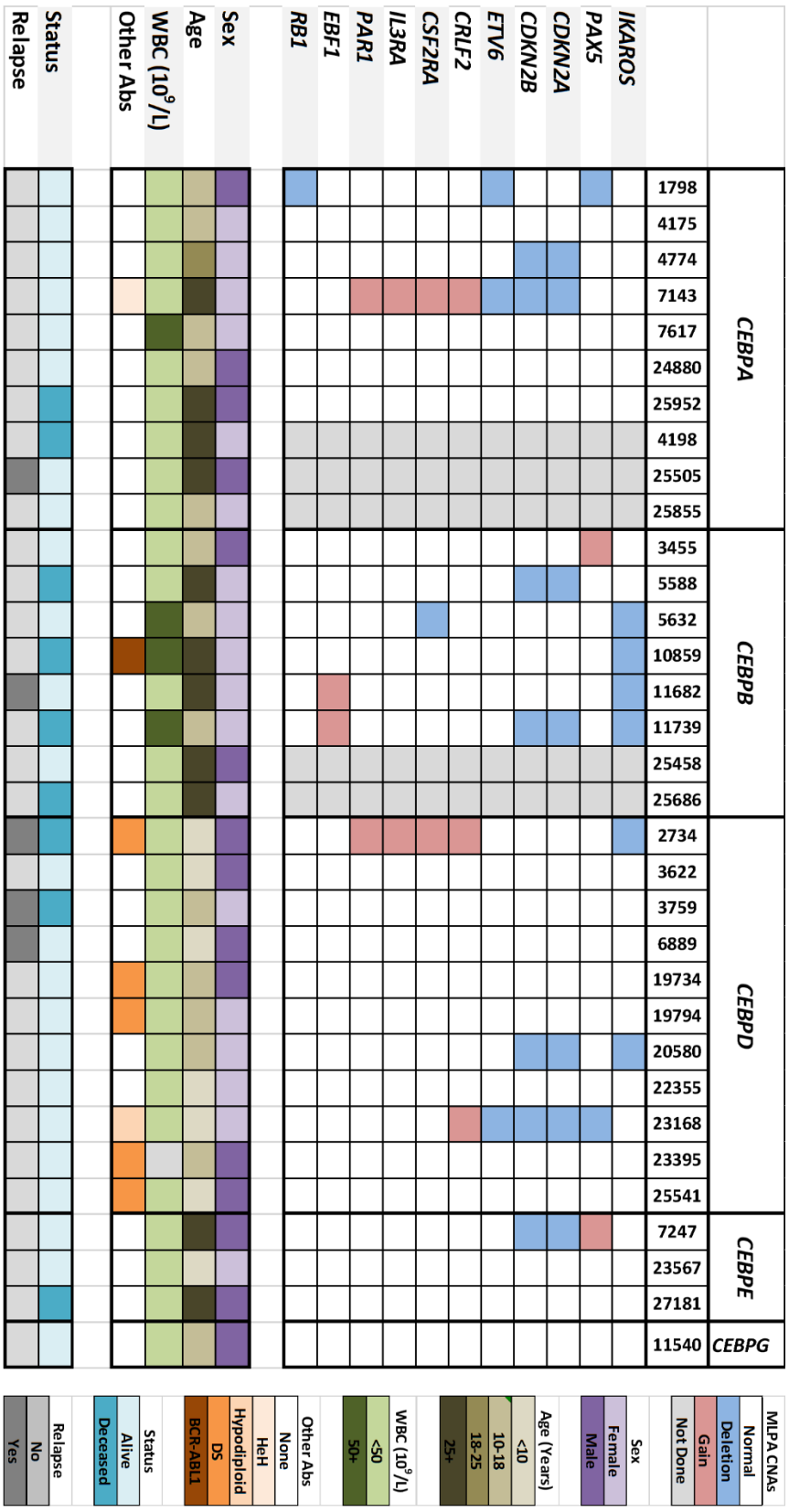


Figure 3.7. Collated heat map of all patient MLPA P335 ALL IKZF1 kit CNAs, sex, age, WBC, associating abnormalities, mortality and relapse data.

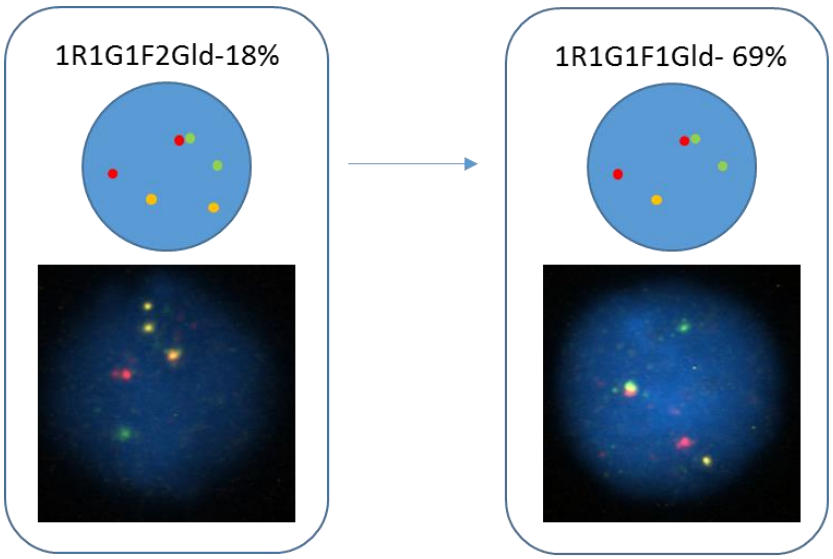
3.3.5 Clonal Evolution

Having screened for the most commonly occurring abnormalities in BCP-ALL, the next step of the project focused on identifying if *IGH-CEBP* translocations were a primary abnormality occurring early in disease progression, or an abnormality acquired later in leukaemic development.

Six patients with multiple CNA abnormalities that were trackable by FISH were chosen for analysis using a multiple colour FISH approach. These probes covered the relevant *CEBP* gene and the genes identified by MLPA. *IKZF1* and *PAX5* probes were labelled with spectrum gold, and *CDKN2A*/*B* with spectrum aqua, to complement the red and green *CEBP* break apart probes (Table 2.1).

3.3.5.1 Patient 10859

Patient 10859 displayed two main clonal populations, the predominant clone (69%) exhibited the translocation of the *CEBPB* gene and a heterozygous deletion of *IKZF1*, the second clone (18%) displayed only the *CEBPB* translocation suggesting the translocation arose before the *IKZF1* deletion (Figure 3.12). Table of all clones identified (Supplementary Table 7.10).



10859

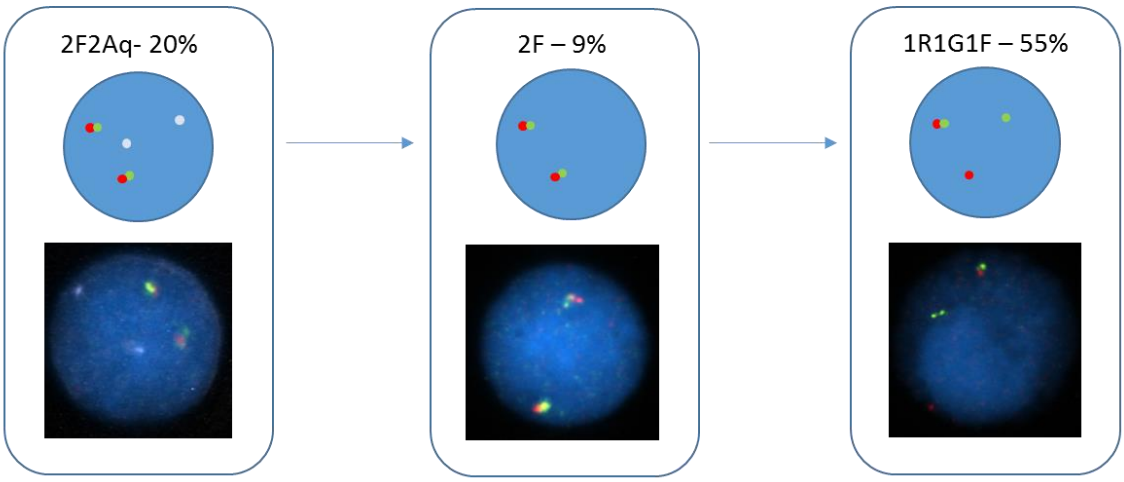
CEBPB ***IKZF1***

					FISH Interpretation	
Red	**Green**	**Fusion**	**Gold**	**Average %**	***CEBPB***	***IKZF1***
1	1	1	1	69	Translocation	Heteroygous Deletion
1	1	1	2	18	Translocation	Normal

Figure 3.8. Tracking the clonal evolution of patient 10859 through analysis of CEBPB and IKZF1, with additional images of nuclei showing FISH signal patterns.

3.3.5.2 Patient 4774

Patient 4774 displayed three predominant clones. The initiating clone (20%) was observed with no aberrations, the subsequent population displayed homozygous loss of the *CDKN2A/B* probe (9%). The terminal population (55%) displayed the homozygous loss of *CDKN2A/B* with a translocation of *CEBPA*. (Figure 3.13). These clones indicate that the *CDKN2A/B* deletion occurred prior to the *CEBPA* translocation in this patient. Table of all clones identified (Supplementary Table 7.11).



4774

CEBPA *CDKN2A/B*

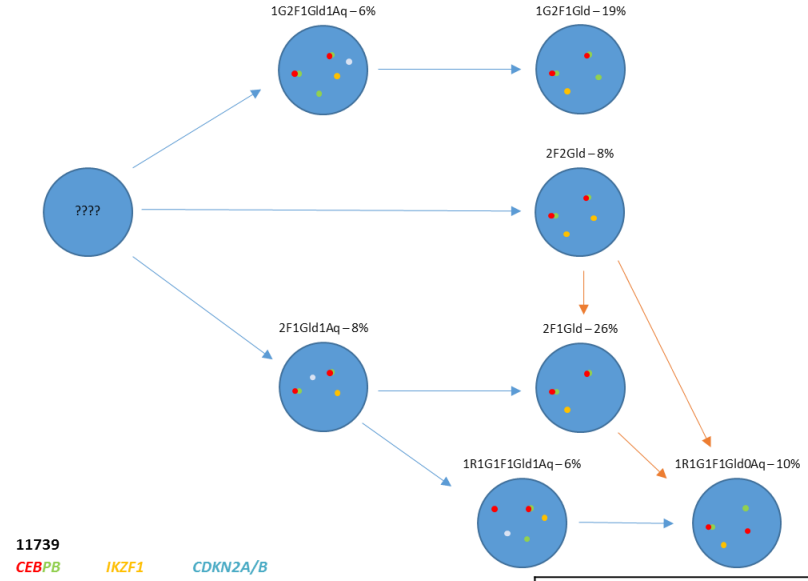
					FISH Interpretation	
Red	Green	Fusion	Aqua	Average %	*CEBPA*	*CDKN2A/B*
1	1	1	0	55	Translocation	Homozygous Deletion
0	0	2	2	20	Normal	Normal
0	0	2	0	9	Normal	Homozygous Deletion

Figure 3.9. Tracking the clonal evolution of patient 4774 through analysis of CEBPA and CDKN2A/B, with additional images of nuclei showing FISH signal patterns.

3.3.5.3 Patient 11739

Patient 11739 displayed seven predominant populations, divided between three potential clonal lineages. As mentioned previously (Section 3.3.3.1) the breakpoint of the *CEBPB* translocation in patient 11739 was more centromeric than standard *CEBPB* translocations, therefore all clonal populations with 0R1G2F are positive for a *CEBPB* translocation. The most prevalent clone (26%) exhibited no *CEBPB* translocation, a homozygous deletion of *CDKN2A*/*B* and a heterozygous deletion of *IKZF1*. The second most prevalent clone (19%) displayed the centromeric *CEBPB* translocation, a homozygous loss of *CDKN2A*/*B* and heterozygous loss of *IKZF1*. Other clones exhibited a mix of deletions and translocations (Figure 3.10). Interpretation of these data indicates that there were two lineages of cells with potentially independent *IGH-CEBPB* translocations, both the classic (1R1G1F) and the centromeric (0R1G2F) breakpoints with each showing homo and heterozygous deletions of *CDKN2A*/*B* and *IKZF1*. The remaining cells (~30%) did not show the *IGH-CEBPB* translocation, however these cells may contain the translocation involving *IGH* and chromosome 5. Interpretation of data indicates that deletions of both *CKDN2A*/*B* and *IKZF1* occurred as primary events in the same cells, prior to

CEBPB rearrangement. Table of all clones identified (Supplementary Table 7.12Supplementary Table 7.10).



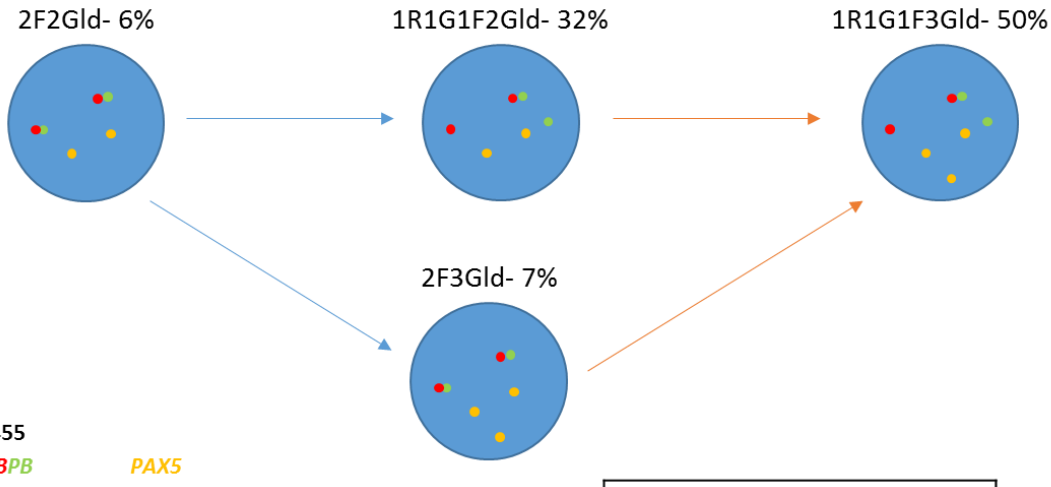
						FISH Interpretation		
Red	Green	Fusion	Gold	Aqua	Average %	*CEBPB*	*IKZF1*	*CDKN2A/B*
0	0	2	1	0	26	Normal	Heterozygous Deletion	Homozygous Deletion
0	1	2	1	0	19	Breakpoint Within Centromeric Probe	Heterozygous Deletion	Homozygous Deletion
1	1	1	1	0	10	Translocation	Heterozygous Deletion	Homozygous Deletion
0	0	2	1	1	8	Normal	Heterozygous Deletion	Heterozygous Deletion
0	0	2	2	0	8	Normal	Normal	Homozygous Deletion
1	1	1	1	1	6	Translocation	Heterozygous Deletion	Heterozygous Deletion
0	1	2	1	1	6	Breakpoint Within Centromeric Probe	Heterozygous Deletion	Heterozygous Deletion

Figure 3.10. Tracking the clonal evolution of patient 11739 through analysis of CEBPB, IKZF1 and CDKN2A/B. Blue arrows indicate known evolution, orange arrows indicate potential evolution.

3.3.5.4 Patient 3455

Patient 3455 displayed four predominant clones, the most commonly occurring (50%) exhibited a *CEBPB* translocation and gain of the *PAX5* gene. The other populations had *CEBPB* translocation only (32%), no translocation and a gain of *PAX5* (7%), and no aberrations (6%) (Figure 3.11). These results indicate two potential routes of clonal evolution, one with a gain of *PAX5* occurring prior to the translocation (7%) and the second suggesting that the translocation occurs before the gain of *PAX5* (32%), ultimately both would lead to the same final and predominant clone with both aberrations present (50%). It is more likely that the *IGH-CEBPB* translocation arose first followed by the gain of *PAX5* due to the higher incidence of the translocation blasts in comparison to the gained *PAX5*

blasts (32% vs 7% respectively). Table of all clones identified (Supplementary Table 7.13)

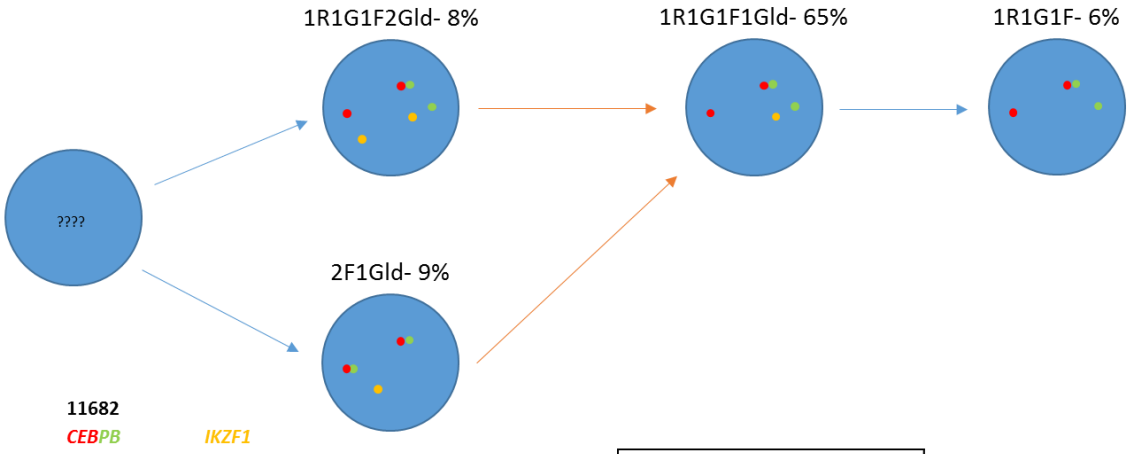


					FISH Interpretation	
Red	Green	Fusion	Gold	Average %	*CEBPB*	*PAX5*
1	1	1	3	50	Translocation	Gain
1	1	1	2	32	Translocation	Normal
0	0	2	3	7	Normal	Gain
0	0	2	2	6	Normal	Normal

Figure 3.11. Tracking the clonal evolution of patient 3455 through analysis of CEBPB and PAX5. Blue arrows indicate known evolution, orange arrows indicate potential evolution.

3.3.5.5 Patient 11682

Patient 11682 displayed four predominant clonal populations, the highest (65%) exhibited a translocation of *CEBPB* and a heterozygous deletion of *IKZF1*. The three remaining populations were a *IKZF1* deleted clone (9%), a *CEBPB* translocated clone (8%), and a *CEBPB* translocation with homozygous *IKZF1* deleted clone (6%) (Figure 3.16). The primary genetic lesion in the observed genes is difficult to ascertain in this sample, the early clones comprise a similar percentage of the population (8% and 9%), and both may have given rise to later clones. Table of all clones identified (Supplementary Table 7.14).

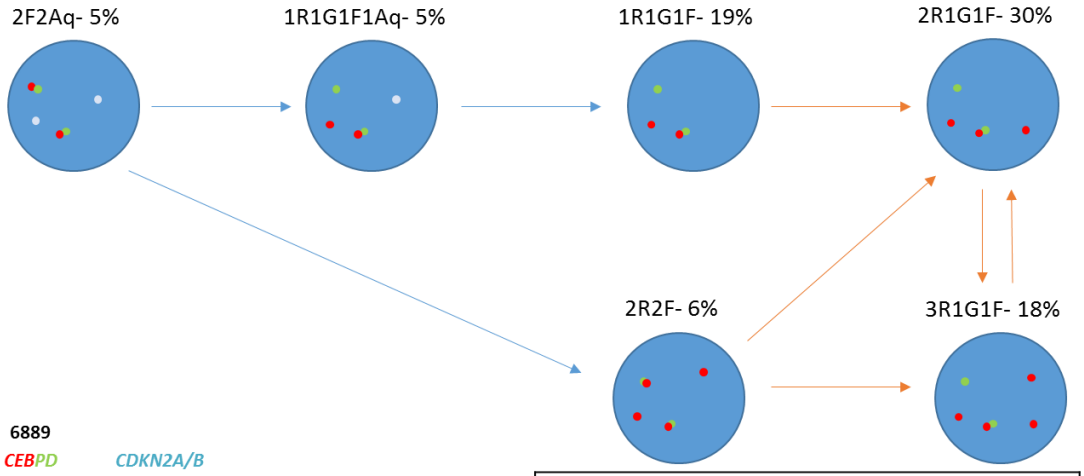


					FISH Interpretation	
Red	Green	Fusion	Gold	Average %	*CEBPD*	*IKZF1*
1	1	1	1	65	Translocation	Heteroygous Deletion
0	0	2	1	9	Normal	Heteroygous Deletion
1	1	1	2	8	Translocation	Normal
1	1	1	0	6	Translocation	Homozygous Deletion

Figure 3.12. Tracking the clonal evolution of patient 11682 through analysis of CEBPB and IKZF1. Blue arrows indicate known evolution, orange arrows indicate potential evolution.

3.3.5.6 Patient 6889

Patient 6889 displayed 6 clonal populations divided into two potential clonal lineages. The main lineage exhibiting a *CEBPD* translocation and heterozygous loss of *CDKN2A*/*B* (5%), which progresses into a population with homozygous loss of *CDKN2A*/*B* (19%) and further gain of the *CEBPD* centromeric signal (30%). The second potential lineage shows homozygous loss of *CDKN2A*/*B* and two additional copies of the *CEBPD* centromeric region (6%) which may culminate in a population with a *CEBPD* translocation in addition to the gained centromeric region (18%) (Figure 3.13). The complexity of this sample suggests several ways in which the more advanced clones may have arisen. Interestingly the additional *CEBPD* signals were not observed in the initial *CEBPD* FISH for this patient. Suggesting potentially more complex sequences of gains and losses in the region. Regardless of the *CEBPD* probe signal pattern, it is unclear which genetic lesion occurred first in this sample. Table of all clones identified (Supplementary Table 7.15).

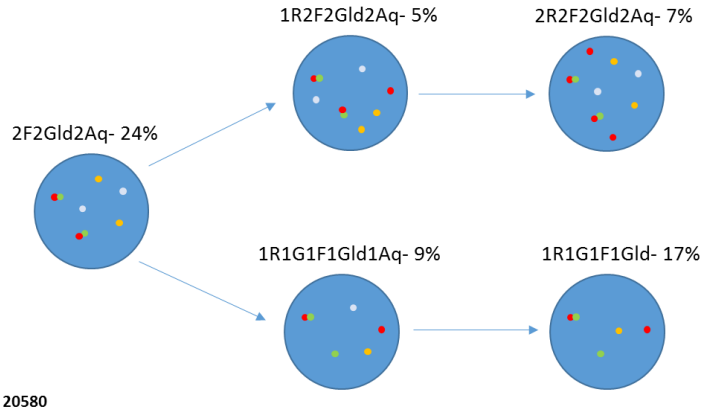


					FISH Interpretation	
Red	Green	Fusion	Aqua	Average %	*CEBPD*	*CDKN2A/B*
2	1	1	0	30	Translocation Gain of Centromeric Probe	Homozygous Deletion
1	1	1	0	19	Translocation	Homozygous Deletion
3	1	1	0	18	Translocation Gain of Centromeric Probe	Homozygous Deletion
2	0	2	0	6	Gain of Centromeric Probe	Homozygous Deletion
1	1	1	1	5	Translocation	Heteroygous Deletion
0	0	2	2	5	Normal	Normal

Figure 3.13. Tracking the clonal evolution of patient 6889 through analysis of CEBPD and CDKN2A/B. Blue arrows indicate known evolution, orange arrows indicate potential evolution.

3.3.5.7 Patient 20580

Patient 20580 displayed five predominant clonal populations divided into two potential clonal lineages. The populations suggest that this sample has two clonal lineages which developed independently from an ancestral clone (Figure 3.14). The more dominant lineage, displayed translocation of the *CEBPD* gene and heterozygous deletions of both *CDKN2A*/*B* and *IKZF1* (9%), which continues to develop an additional deletion of the *CDKN2A*/*B* gene (17%). The second less dominant lineage first exhibits one gain of the *CEBPD* centromeric region (5%), then gains a second addition of this region (7%), this could be suggestive of a separate centromeric *IGH-CEBPD* translocation breaking the centromeric probe apart and giving two signals, which then occurs at the second locus (Figure 2.1). The presence of the two lineages indicates two independent *CEBP* breakpoints in this patient. Table of all clones identified (Supplementary Table 7.16).



						FISH Interpretation		
Red	Green	Fusion	Gold	Aqua	Average%	*CEBPD*	*IKZF1*	*CDKN2A/B*
0	0	2	2	2	24	Normal	Normal	Normal
1	1	1	1	0	17	Translocation	Heteroygous Deletion	Heteroygous Deletion
1	1	1	1	1	9	Translocation	Heteroygous Deletion	Heteroygous Deletion
2	0	2	2	2	7	Gain of Centromeric Probe	Normal	Normal
1	0	2	2	2	5	Gain of Centromeric Probe	Normal	Normal

Figure 3.14. Tracking the clonal evolution of patient 20580 through analysis of CEBPD IKZF1, and CDKN2A/B.

*3.3.6 **SNP Screening***

Having investigated the common CNAs, the focus of the study shifted to search for novel genetic lesions. Fifteen patients in the cohort had sufficient diagnostic DNA (750ng) of high enough quality to be analysed on the SNP 6.0 platform. Of the fifteen, six also had a matched remission sample (Supplementary Table 7.17).

Samples were prepared (Section 2.3.1.3.1) and processed by Aros Applied Biotechnology, to be analysed on the Affymetrix Genome-wide Human SNP array 6.0. Samples with MAPD values higher than 0.35 were discarded. This included diagnostic samples 22355 and 6889, leaving 13 patients to analyse. CNAs in common regions of variation; 2p 11.2, 7p 14.1, 7q 34, 14q 11.2, 14q 32.33 and 22q 11.22 were discarded. CNAs found in diagnostic samples were compared against remission samples when available, CNAs also found to occur in remission samples were discarded as the focus of this study was to identify somatic lesions only. Patients without remission samples were analysed as CNAs had to occur in three or more patients to be investigated further. CNAs were cross checked between MLPA and SNP data, where a high level of correlation was found with all but two samples, 2734 and 6889, which did not allow for accurate comparison due to poor DNA quality (Supplementary Figure 7.1).

SNP analysis of the *IGH-CEBP* cohort identified several recurring CNAs, several were discarded due to being pseudogenes such as *REREP3* (three deletions within the cohort), functionally irrelevant such as *OR4N3P*, an olfactory receptor (three deletions within the cohort) with no functional links to cancer and no published work. Or CNAs were found in regions of naturally high variation such as *SCAPER* (five deletions in the cohort) (Supplementary Table 7.18). One gene of interest, the tyrosine kinase *ABL2* located on 1q25.2, did show promise both due to the function of the gene and due to the focal nature of the deletion observed (Figure 3.15). The log2 ratio data from the SNP arrays indicated a deletion within this gene. While the log2 ratio data showed more variation in *ABL2* deletion patients there was still an observable dip in this region (Figure 3.15B). Deletions involving this gene were found in three patients, 7143, 11682 and 10859.

When observing the array using the copy number state graph the first two patients displayed a focal exon 2 deletion, while patient 10859 displayed a larger deletion in the region covering exons 2-7. Patient 6889, although it had been discarded for a high MAPD score, was also found to contain a 2-7 exonic deletion matching patient 10859 (Figure 3.16A). While the log2 ratio data showed more variation in the *ABL2* deleted patients, particularly in patient 6889 which had a MAPD score, the observed array data warranted further investigation. A FISH probe covering the *ABL2* gene was designed, confirming the deletions in three of four patients (3.3.7).Locations of open reading frames were investigated to determine if the deletions observed in the SNP array data would lead to whole gene inactivation through frame shift or start codon loss. Transcript initiation for the *ABL2* gene begins at exon 1, as no deletions of exon 1 were observed it can be determined that the intragenic deletions of both exon 2 and exon 2-7 would be expressed. Multiple exons in the *ABL2* gene contained splice junctions, meaning that loss of exons could lead to frame shifts and loss of gene function. Exon 2 deletions would not result in frame shift mutations despite a splice junction between exons 1-2, as exons 2-3 also contained a splice junction which maintained the codon order (Figure 3.16Bi). However deletions of 2-7, resulting in fusion of exons 1-8 would result in a frame shift and a totally altered protein (Figure 3.16Bii).

To further investigate this novel deletion, a search was performed of other available patient data published by St. Judes Hospital in 2007, consisting of 242

paediatric ALL patients (Mullighan *et al.*, 2007), which revealed 38 patients with *ABL2* abnormalities mainly linked to whole chromosomal gains or losses. Non-focal whole gene deletions were observed (Supplementary Table 7.19).

To conclude, while numerical abnormalities of *ABL2* were not uncommon the *IGH-CEBP* cohort presented here, gains or deletions of this gene were absent in the St. Jude cohort. This observation highlights a possible association of *ABL2* deletions and *IGH-CEBP* translocations.

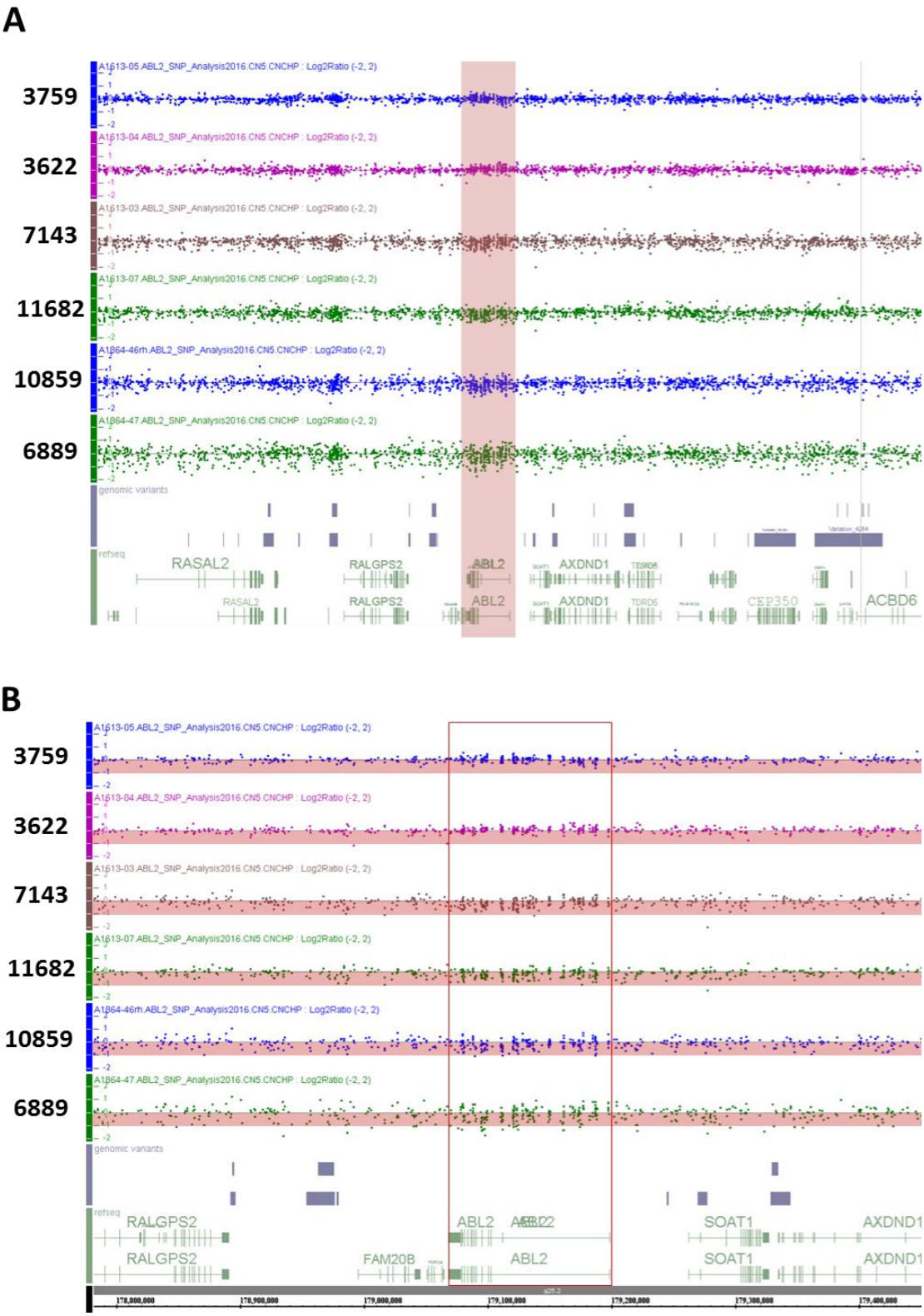


Figure 3.15 SNP identification of the ABL2 intragenic deletion.

A. The ABL2 gene (highlighted) shows no natural genetic variation, and the log2ratio for four patients showed a small dip, B. Log2 ratio of probes covering the ABL2 gene displays a dip over the covering the gene in four patients. Copy number from 0 to -1 is highlighted to aid in visualisation of probe location. Higher variation is observed in the ABL2 deletion positive patients but a greater loss is observed over the ABL2 gene.

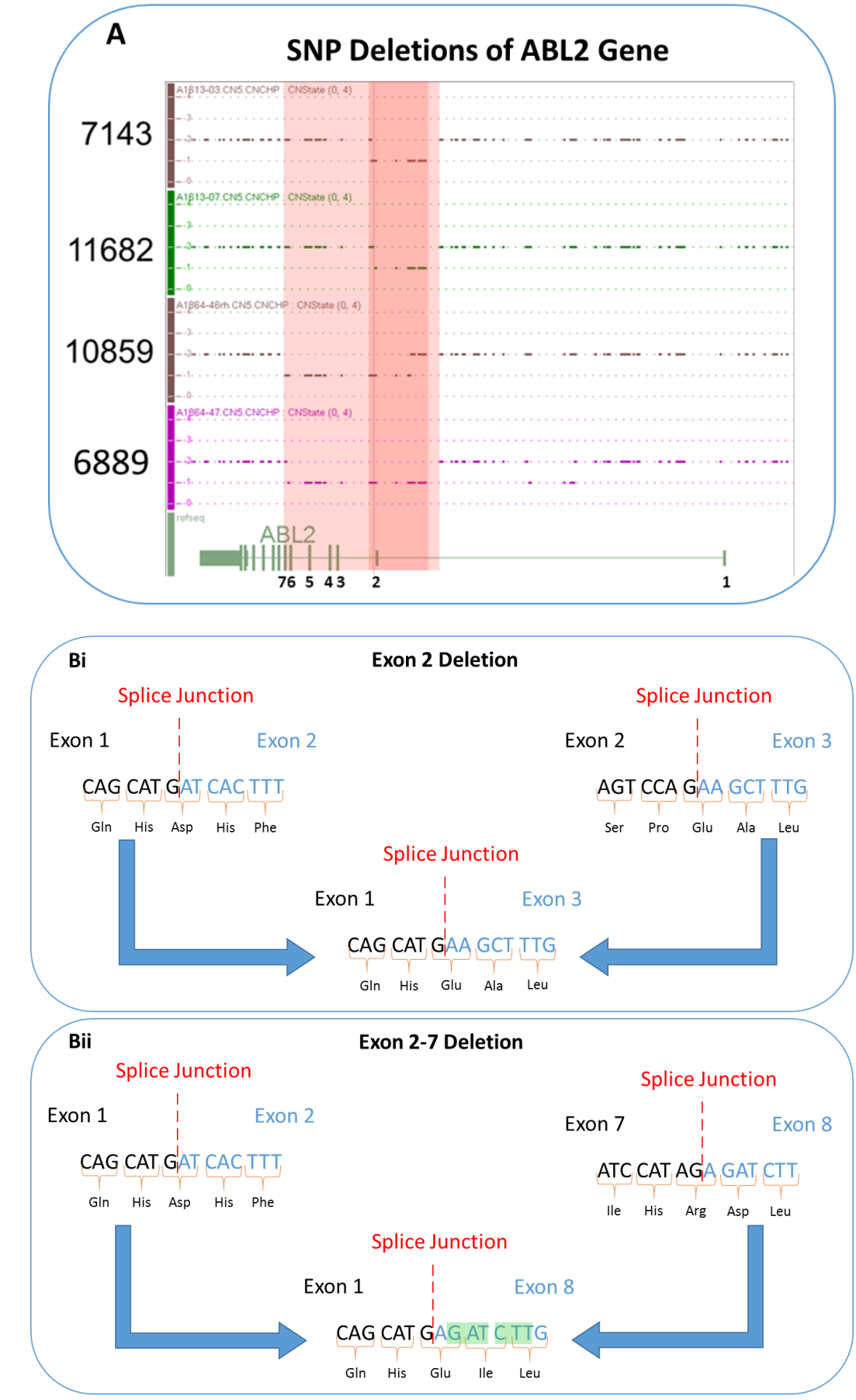


Figure 3.16. Observed deletions of the ABL2 gene.

A. SNP array data displaying the focal deletion of ABL2 found in four IGH-CEBP patients. Dark red shows the common region deleted in all four patients, covering exon 2 of the ABL2 gene, light red shows the extended deleted region of the ABL2 gene in patients 10859 and 6889 encompassing exons 3, 4, 5, 6, and 7.Bi. Potential consequence of the ABL2 exon 2 deletion in protein expression, with the open reading frame being unaffected in the fusion of the exon 1 and 3 fusion. Bii. Potential consequence of the ABL2 exon 2-7 deletion in protein expression, with the open reading frame being altered, leading to a frame shift in the gene. Green boxes indicating the original codons of the ABL2 gene prior to potential intragenic deletions.

3.3.7 ***ABL2 FISH Analysis***

Having observed *ABL2* loss in three high quality and one low quality SNP array, an *ABL2* FISH probe was designed to confirm the SNP findings. A search of the Ensembl and UCSC Genome Browser was performed to identify clones with the potential to identify both *ABL2* deletions. FISH probes are typically created from several bacterial artificial chromosome (BAC) clones 150-350Kb in size, situated alongside each other to bolster the signal strength, giving more robust results when analysing individual nuclei. However the common region of deletion covering exon 2 of the *ABL2* gene found in all four patients was a small target at 32,748bp, too small for a traditional BAC probe. Thus larger BAC clones were disregarded in favor of smaller DNA plasmids of up to 40kb named fosmids. The best located fosmid, G248P8248G11, covered exons 2-10 of the *ABL2* gene (Figure 3.17). Due to the small size of the fosmid the *ABL2* deletion probe was created using gold dNTP as this fluorochrome showed the highest signal intensity. The probe was created using standard methods (Section 2.3.1.1.1).

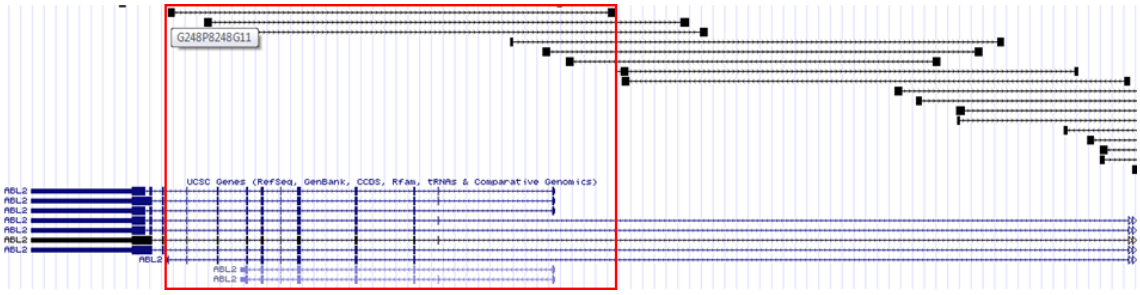


Figure 3.17 Location of the ABL2 fosmid clone G248P8248G11.

All four patient samples with observed SNP deletions were hybridized with the gold *ABL2* deletion probe and probe to the corresponding *IGH* partner gene. Patient 7143 failed to show any gold signal as the fixed cells available were of a

very poor quality. Although the *ABL2* signal was weak the remaining three patients were scored successfully and the presence of an *ABL2* deletion was confirmed in both exon 2 and exon 2-7 deleted patients (Supplementary Table 7.20, Supplementary Table 7.21, Supplementary Table 7.22). There were difficulties in analysis of *ABL2* FISH due to the weak signal of the fosmid probe, and failure of the probe to hybridise to samples with low quality patient samples.

ABL2 translocations had been previously reported in BCP-ALL (Roberts *et al.*, 2014a), a break apart *ABL2* probe was created to rule out the possibility of a translocation in the four patients. Patient 7143 once again failed, the remaining three patients exhibited no translocation of the *ABL2* gene (Table 2.1) (Data not shown).

3.3.8 ***ABL2 qPCR Screen***

Having confirmed the presence of the *ABL2* deletion in three patients, the possibility of further *ABL2* deletions was investigated. In addition to the 13 patients analysed by SNP platform, 13 patients had available fixed cells. Due to difficulties in analysis of the *ABL2* FISH, failure of hybridisation in lower quality fixed cell samples, and the large coverage of the fosmid clone, The alternative approach of copy number qPCR screening for the presence of this deletion was pursued.

The SYBR green qPCR platform was selected, due to the ability to design probes covering specific gene regions. Primers were designed using Primer3Plus, 7 exons within *ABL2* were selected for analysis and multiple primer pairs were evaluated for the best available primer combinations. Primer combinations and disassociation curves were analysed (Figure 3.18) (Supplementary Table 7.23).

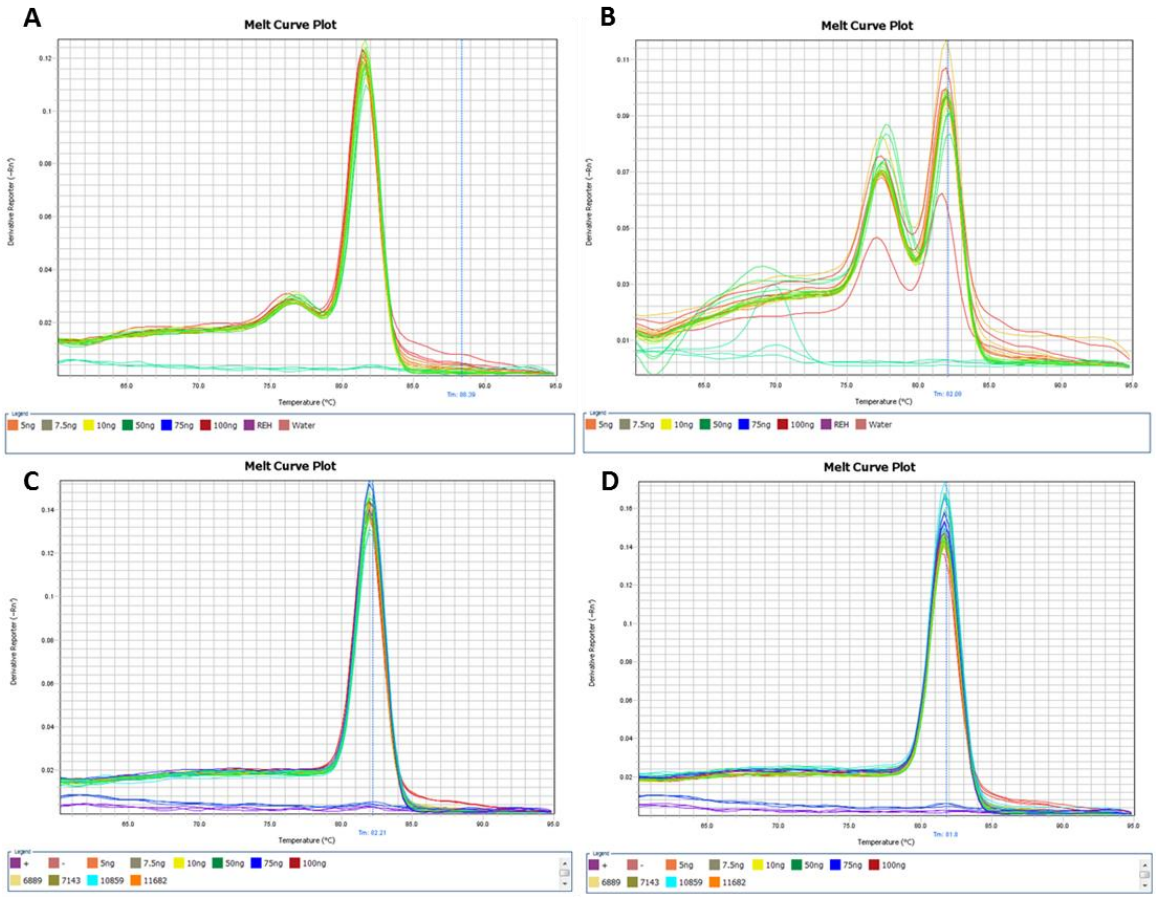


Figure 3.18. Optimisation of primer probes for SYBR Green qPCR for ABL2 copy number analysis.

A. Disassociation curve assessing ABL2 exon 2 primer pair combination 1, a clear 'shoulder' is observed for the run of the primer pair showing nonspecific amplification or primer dimerisation. B. Disassociation curve assessing ABL2 exon 2 primer pair combination 2, at least two separate peaks are observed showing nonspecific amplification. C & D. Disassociation curves assessing new ABL2 exon 2 primer combinations (C. ABL2exon2n1. D. ABL2exon2n2) and specific amplification of product.

Copy number was evaluated on the known deletion positive patients, discovered by SNP arrays, before being applied to the rest of the cohort. Patient 5632 was selected as a negative control for the deletion as SNP analysis displayed a normal *ABL2* copy number. Near haploid BCP-ALL patient 22795 was chosen as the positive control for the deletion as the patient's karyotype indicated a single copy of chromosome 1, the location of the *ABL2* gene. Initial qPCR plates had low R^2 values, indicating a poor fit of data points to the generated standard curve (Figure 3.19 A&B). Several further optimisations were attempted, including varying primer concentrations and combinations, DNA concentration for the creation of the standard curve and recalibration of pipettes. Two variations of *ABL2* exon 2 primers were tested together to investigate variation between primer

pairs, which improved efficiency (Figure 3.19 C&D). The experiment was performed twice, but was not successful. Copy number analysis of control and deleted patients failed to validate FISH and karyotype data (Figure 3.20), while patient 27795 displayed lower expression of the target exons than patient 5632, yet the fold change was not significant. Additionally, the three *ABL2* deletion patients showed high copy number of exon 2 in comparison to other exons, with no variation observed to suggest gene deletion. Further work was ruled out due to time constraints.

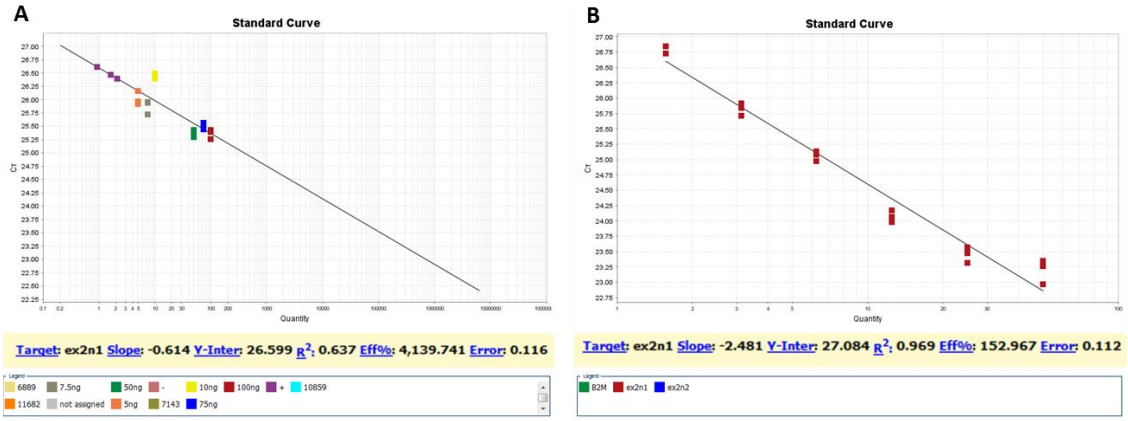


Figure 3.19. Standard curve data and optimisation for SYBR Green qPCR of ABL2 copy number analysis.

A. Standard curve of ABL2 exon 2n1 primer pair showing problems with the experiment as displayed by a low Slope value (-0.64) indicating poor efficiency, and a poor R^2 value indicating that the data is not plotting well to the standard curve. B. Standard curve of ABL2 exon 2n1 primer pair after use of new DNA and DNA dilution series 50ng, 25ng, 12.5ng, 6.25ng, 3.12ng and 1.56ng. Both Slope (-2.481) and R^2 value (0.969) have improved, although slope is not at the optimal -3.32 value indicating perfect efficiency resulting in 100% efficiency in doubling of template product during exponential amplification.

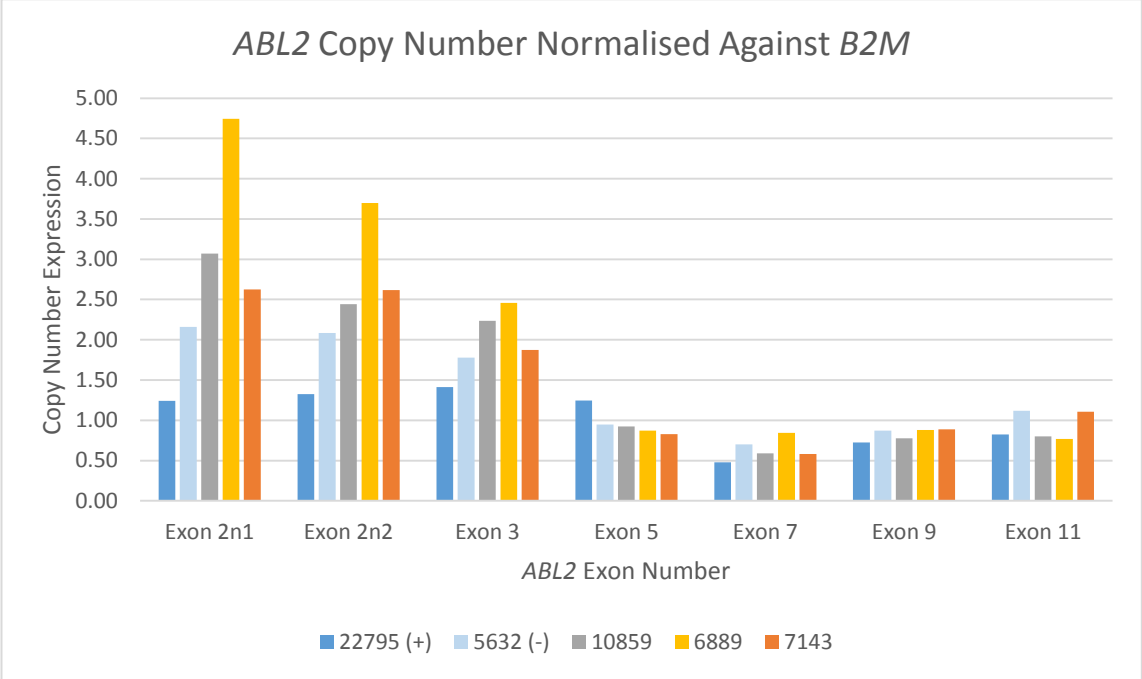


Figure 3.20. Graph displaying copy number data for ABL2 exons 2, 3, 5, 7, 9 and 11 in ABL2 deleted patients 10859, 6889 and 7143, and positive and negative control DNA normalised against B2M gene (N=2).

3.4 Discussion

Genetic characterisation of the *IGH-CEBP* cohort revealed several novel findings. The numbers of *IGH-CEBP* patients in this study was low, reflecting the rarity of the subgroup. Nevertheless this study comprises the largest *IGH-CEBP* cohort analysed to date which has provided the opportunity to perform statistical analysis on the subgroup as a whole, as well as on individual *IGH* partners; *CEBPA*, *CEBPB* and *CEBPD*.

For the majority of patients the cytogenetic breakpoint, indicating the relevant *IGH-CEBP* translocation, was easily confirmed using FISH. The exceptions were the four patients defined as 'other abnormal' who were all ultimately confirmed as *IGH-CEBP* (Section 3.3.3). Among these patients, complexity arose through variant breakpoints giving unusual signal patterns resembling extra probe signals and false fusions signals (patients 11739, 19734, 19794). In addition there was the loss of derived chromosomes following the translocation event, giving different probe patterns (patient 21368). However karyotype data was accurate and facilitated accurate signal interpretation of the FISH patterns. For this reason three patients (25505, 25458 and 25541) were able to be included in the cohort based on cytogenetic data alone.

Patient data were analysed and compared to previously published cohorts, these included a *IGH-CEBP* cohort from Akasaka *et al* (Akasaka *et al.*, 2007), two mixed *IGH* cohorts, one published by Russell *et al* (Russell *et al.*, 2014), and the second by Chapiro *et al* (Chapiro *et al.*, 2013). Two further studies focused on the analysis of the t(8;14)(q11;q32) breakpoint, which is indicative of the *IGH-CEBPD* translocation, the Lundin *et a*l study (Lundin *et al.*, 2009), and the Messinger *et al* study (Messinger *et al.*, 2012).

It is important to note that 17 of the patients from this cohort were included in the Russell *et al* study, and 19 (Supplementary Table 7.5) in the Akasaka study. Patients from this cohort were also included in the Lundin study as this was a review of all known t(8;14)(q11;q32) patients at the time.

The clinical data between all cohorts compared showed a high degree of correlation. Median ages were within the young adult age range of 15-30 years for all cohorts but the exclusive t(8;14)(q11;q32) subgroup which were younger (Table 3.4), supporting findings in this study suggesting *IGH-CEBPD* BCP-ALL initiates earlier. This could be due to the translocation requiring fewer oncogenic hits, or due to hits occurring *in utero*. WBC also showed a high degree of correlation with low expression in all studies, although the range differed (Table 3.4). Finally, incidence of DS was present in all cohorts with the exception of the Chapiro study which did not mention the subgroup. DS incidence was higher in the t(8;14)(q11;q32) cohorts, at 27% for Lundin, and 32% for Messinger, in comparison to this study, at 15%. This is logical as DS is almost totally unique to the *IGH-CEBPD* subgroup in *IGH-CEBP* BCP-ALL, with one exception found in the Akasaka cohort which contained one *IGH-CEBPE* DS patient (Table 3.4).

Study	**Cohort Size**	**Cohort**	**Median Age**	**Age Range**	**Median WBC 10^9/L**	**WBC Range 10^9/L**	**DS Incidence No. (%)**
Current Study	33	*IGH-CEBP*	15	2 - 65	9	1 - 430	5 (15%)
Akasaka ‡	27	*IGH-CEBP*	15	3 - 49	6	1 - 140	3 (11%)
Russell ‡	159	*IGH*	16	NA	11	NA	9 (16%)
Chapiro*	29	*IGH*	25	5 - 85	NA	NA	NA
Lundin ‡*	44	t(8;14)(q11;q32)	13	3 - 49	9	2 - 172	12 (27%)
Messinger	22	t(8;14)(q11;q32)	11	3 – 17	10	1 - 172	7 (32%)

Table 3.4. Comparison of a collection of studies including IGH-CEBP and potential IGH-CEBP patients.

*Median age is similar across most studies, as is median WBC. Incidence of DS appears slightly higher in the Lundin and Messinger cohorts which were comprised of entirely t(8;14)(q11;q32) translocated patients, which is indicative of the IGH-CEBPD translocation. All DS patients were found in the IGH-CEBPD subgroup with the exception of one patient in the Akasaka cohort, who was an IGH-CEBPE patient.‡ Cohort contains patients from this study. * Study is a review, with mix of new and previously published patients.*

Further in-depth analysis of the cohorts identified some interesting trends. The Akasaka *IGH-CEBP* cohort consisted of 27 patients (*CEBPD* n=10, *CEBPA* n=9, *CEBPE* n=4, *CEBPB* n=3, *CEBPG* n=1), and was the first report of multiple CEBP partners to be involved in *IGH* rearrangements. Comparisons between the Akasaka cohort and this cohort showed the median age for both studies was 15 years. However the ranges differed, with our study showing a larger spread (range, 2-65 years) in comparison (range, 3-49 years). This may have been influenced by the inclusion of older patients on the UKALL14 clinical trial (Supplementary Table 7.24). Both cohorts contained one patient each with t(9;22)(q34;q11) translocation.

The composition of the two cohorts differed slightly. This cohort identified an increased number of *IGH-CEBPB* patients, which comprised 24% of total patient numbers (n=8/33) in comparison to 11% (n=3/27) in the Akasaka cohort. Although *CEBPD* and *CEBPA* were the most prevalent *IGH* partner in both studies.

The Russell, *et al* cohort comprised 148 BCP-ALL patients and 11 T-ALL patients with *IGH* translocations, including *IGH-CEBP* (n=17). This study aimed to perform the biggest screen of the *IGH* subgroup to date, investigating clinical and genetic features in patients with both known and unknown *IGH* partners. Overall *IGH-CEBP* patients displayed both similarities and differences to the *IGH* group as a whole. Adding the new *IGH-CEBP* patients to the Russell cohort gave a total of 175 *IGH* positive patients and 3285 ALL patients, this gave an incidence of 19% *IGH-CEBP* patients in the *IGH* group and 1% incidence in ALL. A similar median age and WBC was observed in the Russell cohort, however relapse was lower (7%), in comparison to the *IGH-CEBP* subgroup (15%), indicating that the *CEBP* subgroup may be at higher risk than other *IGH* partners. This did not affect mortality rates which were similar at 22% in *IGH* and 27% in *IGH-CEBP*.

The Lundin study gathered data on previously published occurrences of t(8;14)(q11;q32), in BCP-ALL indicative of *IGH-CEBPD* rearrangement, and added three new patients to the study cohort. This report included the Akasaka cohort and several others, giving a total of 44 t(8;14)(q11;q32) patients, only the Akasaka cases having been confirmed as *IGH-CEBPD*.

Karyotype data for the Lundin cohort showed the t(8;14)(q11;q32) translocation as the sole chromosomal aberration in 13 (30%) of cases, occurring in 50% of DS patients and 22% of non-DS patients. Karyotype data also showed no additional acquired changes in any of the DS patients. In non-DS patients the t(9;22)(q34;q11) translocation was found in seven patients (16%). Gain of chromosome X (11%) and somatic gain of chromosome 21 (14%) were also observed, however one of the gains was observed in a HeH patient, individual gains of these chromosomes in non HeH patients occurred at 9% for both X and 21 gains. Overall the Lundin study highlighted the low level of associating abnormalities in the *IGH-CEBPD* subgroup, in particular that of the DS patients, and reported t(9;22)(q34;q11) as a recurring abnormality. My data did not observe this trend, with a single t(9;22)(q34;q11) occurring in a *IGH-CEBPB* patient.

The Messinger cohort analysed the treatment and survival data of 22 t(8;14)(q11;q32) BCP-ALL cases. The only t(9;22)(q34;q11) rearrangement was discovered in a non-DS patient. Interestingly, DS patients were found to have a better 5 year event free survival vs non-DS patients (100% vs. 50.1 ± 17.7%; p=0.04) using Children Oncology Group (COG) treatment protocols, which separated the group between standard-risk (7/22) and high-risk (15/22), although the criteria for the stratification of these patients was unclear with no age or WBC bias in the selection. Improvement in patient outcome was not due to treatment stratification as DS patients were split between standard (n=3) and high risk (n=4). Patients were treated on protocols spanning several decades, which may have impacted on survival, through refinement of treatment protocols. The general characteristics of the Messinger cohort were similar to both the Lundin and this cohort. Young adults with low WBC counts and a high incidence of DS patients. Although Messinger *et al* found a lower incidence of t(9;22)(q34;q11) rearranges in comparison to Lund, with only one patient. This work is interesting as it suggests that potentially *IGH-CEPD* patients may benefit from treatment stratification based on DS occurrence, which is not an expected outcome in DS

BCP-ALL as these patients have been shown to have worse outcome (Xavier and Taub, 2010). Although there needs to be a larger cohort of this subgroup to investigate this work further.

The Chapiro cohort consisted of 29 BCP-ALL *IGH* translocated patients, containing 11 *IGH-CEBP* patients, and several other known partners; *BCL2*, *ID4* and *EPOR,* and several unknown partners. This study found recurring gains of X (n=4/11) and non-constitutional gains of 21 (n=2/11). Both which had also been observed in the Lundin study, although neither trend was observed in this cohort.

While this *IGH-CEBP* cohort shows little variation from previous findings both in relation to other *IGH-CEBP* and other *IGH* partners, when the cohort is further subdivided into individual *CEBP* subgroups interesting trends are observed. There is a clear difference in age among *CEBP* subgroups (Figure 3.1). Patients with *CEBPB* translocations were significantly older, while patients with *CEBPD* involvement were predominantly children, a significant finding with and without the inclusion of DS patients in the *CEBPD* subgroup, suggesting that the young age of the *IGH-CEBPD* subgroup was not biased by DS patients (Xavier and Taub, 2010). These data are limited in consideration of patient survival but when added to the observation that *IGH-CEBPB* patients have the highest WBC, highest median number of CNAs, and the highest number of deceased patients, it raises an interesting point that this subgroup may benefit from more intensive treatment regimens. Although such suggestions should be interpreted with caution on such small patient numbers, and the natural correlation between worse outcome and age in BCP-ALL (Harrison, 2011).

A study of past literature is difficult as the majority of studies published on this subgroup contained patients in this study, with only Chapiro and Messinger having no overlap. Overall the representation of *IGH-CEBPs* and potential *IGH-CEBPs* is concurrent with this study, showing *IGH-CEBPs* as afflicting younger patients, showing low WBCs and being associated with DS. While previous observations of recurrent gains of +X and non-constitutional +21 was not found. Survival analysis is certainly an avenue of future research for this cohort as it grows larger, with Messinger *et al* highlighting some interesting trends in DS survival in their cohort.

This study is the only one to date to perform in-depth CNA analysis of the *IGH-CEBP* subgroup. CNAs were most common in the *CEBPB* subgroup (100% patients with CNAs) and least common in the *CEBPD* subgroup (27% patients with CNAs). This may be related to patient age, while 4/5 DS patients in the *CEBPD* subgroup exhibited no CNAs, which may suggest that trisomy 21 is a potent oncogenic event in itself, requiring fewer additional genetic lesions for BCP-ALL development, an observation partially supported by the low number of additional karyotype abnormalities in DS patients in the Messinger cohort. CNA comparisons between the Russell *IGH* cohort and this cohort showed several similarities; the dominant deletions in both cohorts were *CDKN2A*/*B* (occurring at 41% vs 25% respectively) and *IKZF1* (40% vs 21%), while *PAX5* abnormalities were the next most prevalent (25% vs 14%). Incidence of CNAs were consistently lower in the *IGH-CEBP* cohort. Other trends were observed, while *CDKN2A*/*B* deletions were evenly spread throughout the *IGH-CEBP* cohort, *IKZF1* deletions were found exclusively among the *CEBPB* and *CEBPD* subgroups, providing further evidence that the *CEBP* translocations may be subdivided by both clinical and genetic differences.

Tracking clonal evolution in selected *IGH-CEBP* patients showed heterogeneity between individual cases. Of the seven patients two patients, 10859 and 3455, displayed the *IGH-CEBP* translocation as the primary genetic event observed. A further four patients, 4774, 11739, 6889 and 20580, showed the emergence of the associating gene deletions prior to the *IGH-CEBP* translocation event. With *CDKN2A/B* being the gene deleted prior to the *IGH* translocation in all cases, and IKZF1 deletions observed both before and after translocation events. The order of events in patient 11682 remains unclear. Although populations with levels of signal patterns below the 5% cut off value suggested that associating deletions were the primary events, these findings require further investigation, either through increasing the numbers of nuclei scored by FISH or through use of alternative experimental platforms such as single cell qPCR to analyse individual cell transcript expression. Patients analysed were skewed towards the *CEBPB* subgroup as the majority of MLPA abnormalities discovered were in these patients. It is interesting however that the two patients showing the translocation as the primary event were *IGH-CEBPB* patients. The patients showing gene deletions as primary events were one *IGH-CEBPD* patient and one *IGH-CEBPB*

patient, although this data maybe misleading as other genetic aberrations may be occurring prior to the *IGH-CEBP* translocation. The heterogeneous nature of the *IGH-CEBP* translocations is supported by variation of expression between the patients. Median range of the incidence of *IGH-CEBP* positive clones was; *CEBPA* 73%, *CEBPB* 70%, *CEBPD* 77% and *CEBPE* 57%, with a range of 11-94% suggesting that *IGH-CEBP* translocations are early initiating lesions in some patients, and a later acquisition in others. Although no trends were observed, the percentage of the clonal population in *IGH-CEBPE* patients was lower than other *CEBP* subgroups, a trend highlighted when compared to other *IGH* partners as was observed in the Russell *et al* study (Figure 3.21). However as only three patients had this translocation, it is difficult to speculate on this point further.

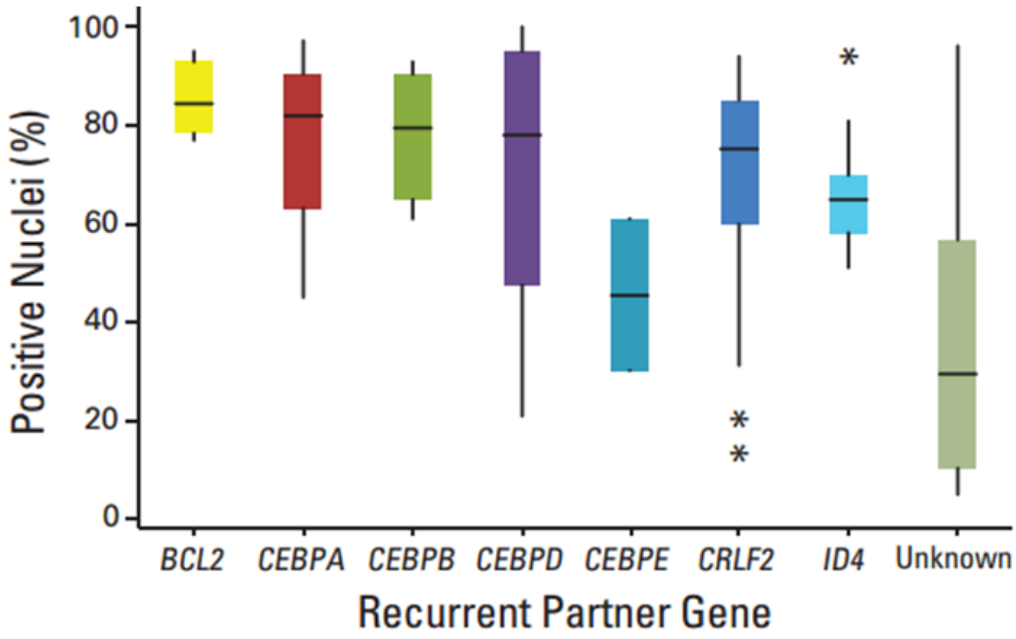


Figure 3.21. Box plot showing percentage of IGH translocated cells with different partner genes with CEBPE and unknown partner genes as the lowest occurring. Taken from (Russell et al., 2014).

Overall the *IGH-CEBP* cohort did not exhibit many recurring aberrations, those observed were mainly found in natural regions of high variability, were functionally irrelevant, or coded for pseudogenes.

However a novel recurrent finding was focal deletions within the *ABL2* gene. *ABL2*, a paralogue of *ABL1*, is responsible for transmembrane signaling and the activation of signaling pathways, which control multiple processes within the cell including cell proliferation and survival (Colicelli, 2010). Both genes have

individual and overlapping physiological roles, which have been tested in mice. Murine knock out of the genes showed differing consequences, with *ABL1* knock out mice displaying thymic and splenic atrophy and reduced B and T lymphocyte numbers among other effects, (Schwartzberg *et al.*, 1991; Tybulewicz *et al.*, 1991) while *ABL2* knock outs displayed dendrite destabilisation and neuronal defects (Miller *et al.*, 2004; Greuber and Pendergast, 2012). Although the aberrant expression of *ABL1* is well established in leukaemia, the role of *ABL2* is less well known. More information is available on solid tumors, where amplification of *ABL2* has been observed in breast cancer (Eirew *et al.*, 2015), hepatocelullar carcinomas (raw data NCI), and both amplifications and mutations observed in lung adenocarcinomas (Giordano, 2014). However the dominant route of de-regulation for *ABL2* across the majority of published datasets was amplification, with few mutations or deletions observed (Greuber *et al.*, 2013). Previous findings in leukaemia document translocations with the *ETV6* gene in rare cases of AML and T-ALL (De Braekeleer *et al.*, 2012), and more recently translocations between *PAG1-ABL2*, *RCSD1-ABL2* and *ZC3HAV1-ABL2* (in seven patients) in BCR-ABL1 like BCP-ALL have been reported, and result in cytokine-independent proliferation and activation of phosphorylated *STAT5* (Roberts *et al.*, 2014a). In all *ABL2* fusions the tyrosine kinase domain remained intact, suggesting a different downstream effect to the deletions observed in this cohort.

Combining these findings with those from St Jude 2007 cohort, suggests that focal deletions of the *ABL2* gene are a rare event in BCP-ALL. The common region of deletion in our patients covers the N-terminal cap region which functions in the stabilisation of the Src Homology 3 domain (SH3) and SH2 domain (Figure 3.22 A&B). The SH domains allow the *ABL2* protein to dock to phosphorylated tyrosine kinase residues. This function is important for signal transduction through the cell membrane, and activation or repression of multiple downstream targets. The SH3 domain has also been shown to bind the SH2 kinase linker which suppresses kinase activity of the protein. The N-terminal cap, also known as the myristoylated residue, stabilises this process by binding to the hydrophobic pocket within the C-lobe of the kinase domain, and maintains protein conformation in the inactive state (Figure 3.22 C). Loss of this cap would potentially lead to constitutive kinase activation due to inability of the SH2 and

SH3 domains to keep the kinase domain in a closed conformation (Figure 3.22 D) (Chen *et al.*, 2008b). Although only the 1b ABL2 isoform contains this cap, and therefore would be affected, while the 1a isoform is regulated by other methods. The importance of the N-terminal cap in control of *ABL* function is highlighted by type 3 TKIs GNF-2 and GNF-5 which target the myristoyl-binding pocket to maintain the closed protein conformation of the ABL1 protein (Zhang *et al.*, 2010). Patients 10859 and 6889 both showed larger deletions covering exons 2-7. These deletions remove part of the N-terminal cap, SH2, SH3 regions and part of the kinase domain and would result in disruption of the gene reading frame due to splice sites the affected exon boundaries. This would again leave the kinase domain uncovered due to lack of the SH regions to keep the ABL2 protein in a locked conformation, leading to constitutive activation, or more likely simply inactivated due to the frame shift.

Unfortunately attempts to perform a cohort wide screen of *ABL2* using qPCR were unsuccessful. Further analysis of the incidence and functional consequence of *ABL2* deletions would also be valuable to determine its specific role within the *IGH-CEBP* cohort. Developing the SYBR Green qPCR copy number assay would allow a cohort wide screen with little DNA needed. The functional consequence of the deletion would be an interesting avenue of research and could include using gene editing systems such as CRISPR to investigate the effects of *ABL2* exon 2, and exon 2-7 loss in CD34+ cells or pre-B cells, alone and in combination with other genetic lesions. Determining if these deletions lead to constitutive activation could allow a section of *IGH-CEBP* patients for treatment with class 1 TKIs, should activation prove to be resulting in real functional downstream consequences in the leukaemia. *ABL2* constitutive activation has been treated with the use of Disatinib in subsets of *BCR-ABL1*-like patients (Roberts *et al.*, 2014a). Potential methods of investigation would be investigating direct downstream targets using RNA sequencing (RNA-seq) for transcript expression or immunoblotting for protein expression. This would however require additional viable patient material. Investigation of downstream targets and regulatory molecules is another option. For example ABL kinases directly interact with the RIN1 gene at the SH2 and SH3 regions, this gene has been shown to function as an adaptor protein which enhances ABL catalytic activity (Cao *et al.*, 2008). Co-immunoprecipitation of RIN1 with the exon 2 deleted *ABL2* patients could be

performed to investigate protein-protein interaction. Interaction would be expected in wild type ABL2 and exon 2 deleted ABL2, but lost in the larger 2-7 exon deleted samples.

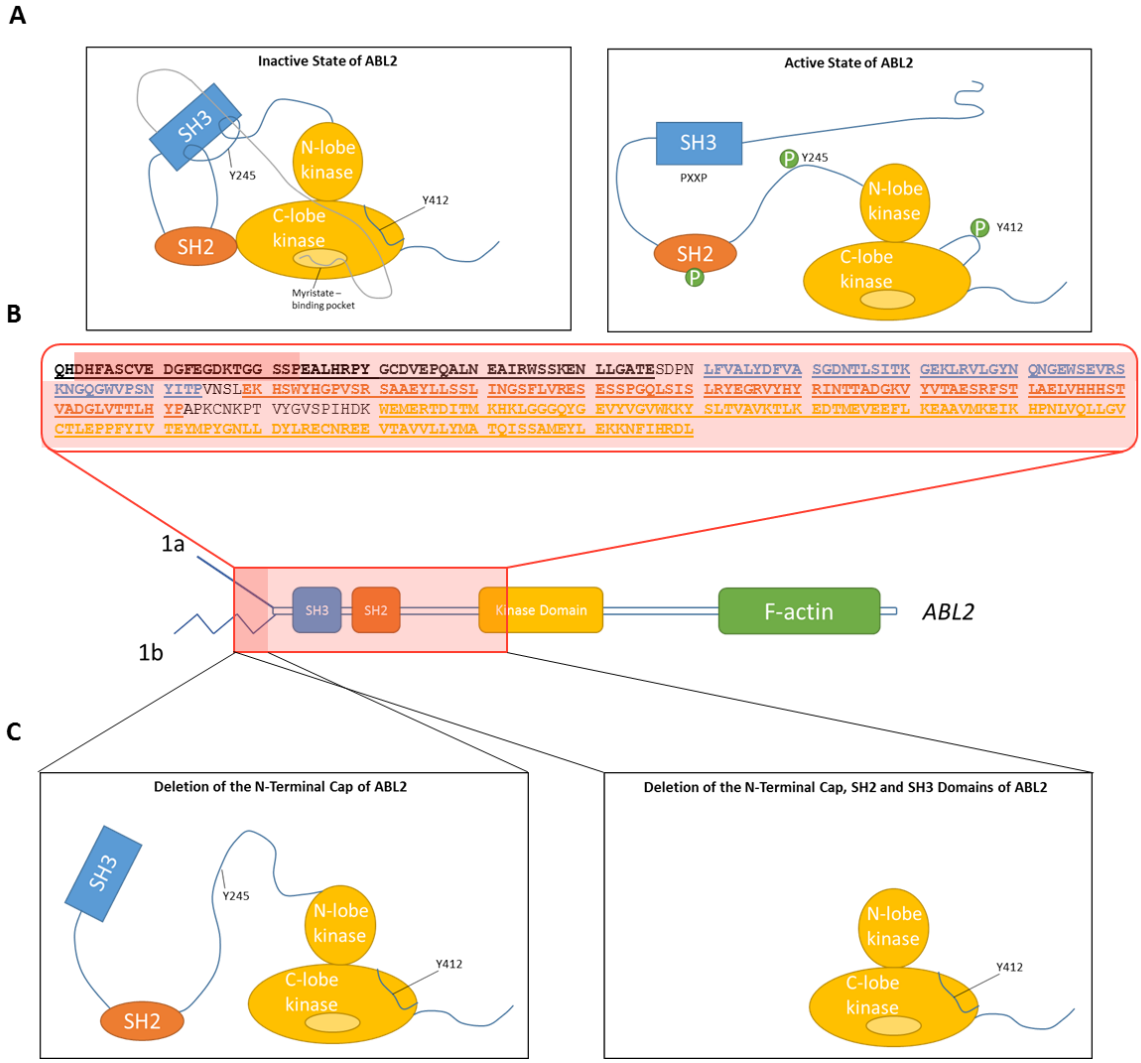


Figure 3.22. The location and potential consequences of the focal ABL2 deletion in IGH-CEBP patients.

A. Normal function of the ABL2 protein, displaying the function of the SH2 and SH3 regions, maintaining the protein in the inactive conformation by inhibiting contact with the kinase domain, and upon phosphorylation of tyrosine residues Y245 and Y412, the active state of the protein, with SH2 and SH3 releasing the linker chains of the protein, allowing interaction with the kinase domain region. B. Amino acid sequence of the deleted exons, covering part of the N-terminal cap (denoted by bold and underlined black text), SH3 region (denoted by bold and underlined blue text), SH2 region (denoted by bold and underlined orange text) and part of the kinase domain (denoted by bold and underlined yellow text). The highlighted sections in dark red display the region of the protein sequence lost due to exon 2 deletion, a part of the N-terminal cap. Highlighted sections in light red display the regions of the protein sequence lost due to exon 3-7 loss, covering the remaining N-terminal cap sequence, the SH2, SH3 regions and part of the Kinase domain. The same areas are displayed on the protein loci exhibited with alternative splicing isoforms 1a and 1b and with individual protein domains. C. Potential consequences of exon 2 and exon 2-7 deletions in the ABL2 protein. Exon 2 deletion resulting in partial deletion of the N-terminal cap would result in destabilization of the SH2 and SH3 domain due to failure to the linkage protein to bind to the Myristate binding pocket, leaving the protein in a constitutively active conformation. Exon 2-7 deletion would result in loss of the N-terminal cap and SH2, SH3 and partial kinase domain loss resulting in a constitutively active and potentially inactivated protein due to frame shift. Figures A and C adapted from (Greuber et al., 2013).

To conclude; of the five *IGH* partner genes, *CEBPB* seems to be linked to the development of a more aggressive disease phenotype, with higher WBC and reduced survival. With *CEBPB* patients showing at least one CNA by MLPA and *CEBPD* patients showing the lowest number of CNAs (Figure 3.11). Additionally *IKZF1* deletions appeared to be exclusive to the *CEBPA* and *CEBPB* subgroups. Comparing the cohort against those previously reported suggests that the *IGH-CEBP* subgroup is a more aggressive subset of *IGH* translocations in BCP-ALL. The *IGH-CEBP* translocations appear to be both primary and secondary oncogenic events. The discovery of the *ABL2* intragenic deletion appears to be unique to the *IGH-CEBP* cohort and may represent a new oncogenic aberration in BCP-ALL.

Chapter 4 Optimisation and Analysis of *in vivo* and *in vitro* Functional Work in the *IGH-CEBP* Cohort

4.1 Introduction

The use of disease models has become a powerful tool for the molecular characterisation of cancer. Disease models were first established using immortalised cell lines, which provided a virtually unlimited supply of homogeneous cells for analysis. Uses have included the identification of oncogenes, determining cytogenetic breakpoints (Drexler *et al.*, 1995), testing novel therapies (Carroll *et al.*, 1997), and investigating gene knock down and overexpression (Kennedy and Barabe, 2008). Weaknesses of cell line models include issues with pre-existing germline mutations interfering with induced genetic lesions potentially biasing data, and the discrepancy between *in vitro* and *in vivo* cell environments.

The adoption of transgenic mice for disease modelling is an alternative technique with several advantages over *in vitro* cultures. Similarities between human and murine genetics, physiology and anatomy, coupled with the relative ease of murine genetic modification and low costs of up keep, have resulted in a powerful tool for long term modelling in a favourable environment. A prime example are the NOD SCID Gamma (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ) (NSG) mice with combined genetic modifications resulting in wide scale immunodeficiency, which facilitates the expansion of foreign cells. NSG mice were the culmination of decades of research, beginning with the introduction of a mutation in the *Prkdc* gene, which was discovered to improve engraftment of primary material in mice due to hindrance of murine B and T cell development. However the action of NK cells remained a major problem in efficient expansion of foreign cells (Bosma *et al.*, 1983). Non-obese diabetic (NOD) mice were bred initially as an animal model for type 1 diabetes. As this murine strain also exhibited low levels of NK cells, it was crossed with the SCID strain, resulting in further improved engraftment of foreign cells (Shultz *et al.*, 1995). However, the activity of NK cells and components of innate immunity meant engraftment levels were still low. Finally a null mutation affecting the *IL2Rγ* chain was expressed in a murine strain,

inactivating receptor signalling of multiple interleukins vital for B, T and NK cell development and function. Crossing these mice with the previously developed NOD-SCID strains removed the functionality of the remaining NK cells, facilitating the use of mouse models in haematopoietic and leukaemic studies (Shultz *et al.*, 2005).

Such mice are used to expand primary patient material, investigate clonal evolution of patient samples and transduced cells, perform functional studies of oncogenes and tumour suppressor genes, and to test therapeutic agents (Bernardi *et al.*, 2002). Conditioning techniques, such as sub-lethal irradiation and the use of chemotherapeutics such as Busulfan (Robert-Richard *et al.*, 2006) have further improved engraftment of selected cells in mice by facilitating removal of the competing murine bone marrow microenvironment.

The use of primary human material in disease modelling conveys the obvious advantage of experimentation in a similar cellular environment to primary patient samples, particularly advantageous in leukaemic disease as human haematopoietic progenitors are relatively easy to isolate. Prime sources include adult bone marrow, peripheral blood (PB) and most commonly human cord blood, which is both abundant and enriched for haematopoietic progenitors such as CD34+ cells (Zhang *et al.*, 1997). The use of these cells in combination with retroviral particles allows for the expression of specific genes of interest, normally tagged to a selection marker, enabling visualisation of the induced genetic lesion. These approaches allow for long term tracking of expressed retroviral constructs and use in cell proliferation, self-renewal and differentiation assays. This system has been successfully used with multiple oncogenic targets including gene fusions such as *BCR-ABL1* (Chalandon *et al.*, 2002), *RUNX1-ETO* (Mulloy *et al.*, 2003), *KMT2A-AF9* (Barabe *et al.*, 2007), as well as single genes, such as *CEBPA* (Mulloy *et al.*, 2003). The linchpin for such work has been the development of increasingly sophisticated retroviruses which have become potent DNA delivery systems allowing for efficient and stable long term expression of target sequences.

Retroviruses are the most popular method of gene delivery and are divided into simple (murine leukaemia virus (MLV)) or complex (lentiviruses) classifications. The difference being additional regulatory and accessory genes present in lentiviruses, allowing for their integration into non-dividing cells.

A retrovirus consists of an inner viral core containing two copies of single stranded (ss) RNA, 7-12kb in length which code for the gag, pol and *env* genes. This core also contains three protein enzymes, which act at different stages of viral replication. The reverse transcriptase enzyme transcribes the single stranded RNA into a DNA strand. The integrase enzyme acts to integrate the transcribed viral DNA into the host genome and the protease enzyme cleaves transcribed viral polyproteins into functional protein segments (Figure 4.1).

During viral replication the gag gene is transcribed into a polyprotein, which is cleaved by the viral proteases to create the inner viral core consisting of the matrix protein and nucleocapsid. The *env* gene is transcribed into a glycopolyprotein (gp160), which is cleaved by host cell proteases to create surface envelope glycoprotein (gp120) and transmembrane protein (gp41). These two glycoproteins form the outer membrane of the retrovirus and facilitate the entry of the retroviral nucleocapsid into the host cell. The *pol* gene, transcribed as *gag-pol* due to a ribosomal frame shift during *gag* mRNA translocation, codes for the inner core enzymes; reverse transcriptase, integrase and protease (Freed, 2001; Barker and Planelles, 2003; Warnock *et al.*, 2011) (Figure 4.1).

Viral infection begins when the surface envelope g120 protein binds with the host cell cellular and core receptors. This binding alters the conformation of the viral outer envelope bringing the transmembrane envelope protein (gp41) into contact, and in turn fusion, with the host cell membrane leading to the delivery of the viral core into the host cell cytoplasm. Upon delivery to the host cell, the single stranded (ss) RNA is transcribed into double stranded (ds) DNA while still contained within the viral core (Freed, 2001; Fanales-Belasio *et al.*, 2010). For simple retroviruses the viral dsDNA can only enter the host cell nucleus upon breakdown of the nuclear membrane during mitosis (Warnock *et al.*, 2011). Lentiviruses have an additional set of genes, which allow for integration of the viral dsDNA outside of cell replication; a beneficial characteristic of lentiviral vectors.

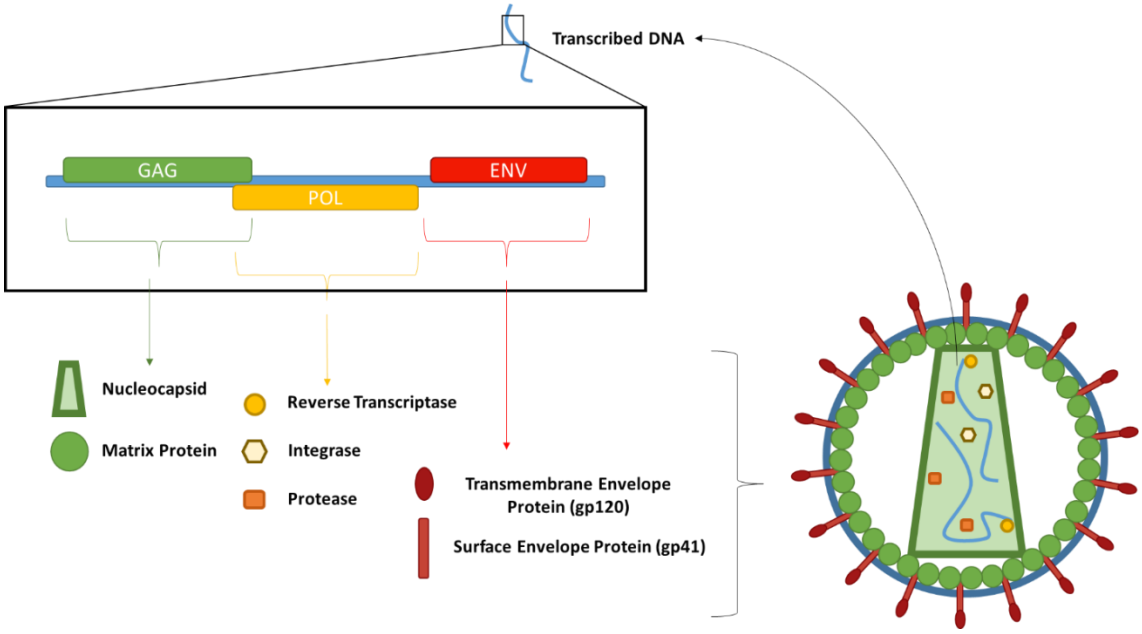


Figure 4.1. The composition of a retrovirus with a representation of the gag, pol and env genes and their respective products.

Integration is mediated by the integrase enzyme, which digests the 3' termini of both DNA strands and cleaves the cellular DNA, allowing for ligation of the viral DNA. Site integration, while initially thought to be random, appears to be at least partially targeted with different preferences depending upon the retrovirus. Preferences include integration into coiled or uncoiled chromatin (Daniel and Smith, 2008), integration within genes (Daniel and Smith, 2008), seclusion from transcriptional start sites, near 5' transcriptional units, in proximity to DNase I-hypersensitive sites, or CpG Islands (Bolognesi, 1993; Dropulic, 2011). The level of complexity is increased when considering other host cell factors. Generally however it is accepted that lentiviral vectors produce lower oncogenic integrations in comparison to retroviral vectors (Bolognesi, 1993; Dropulic, 2011).

Once integrated, viral DNA is then expressed by the natural cellular machinery , using viral promoters in the 5' long terminal repeat (LTR) region (Buchschacher, 2001), and transported into the cellular cytoplasm, where new viral protein components are assembled in the rough endoplasmic reticulum. These are assembled together with viral RNA synthesised in the nucleus at the cell membrane, where the virus exits the cell to begin the process again.

A popular lineage of retroviral vectors stems from the MLV virus, which is classed as a group VI gamma retrovirus. Group VI viruses contain positive strand RNA within the viral nucleocapsid, which is reverse transcribed into DNA, ready for

insertion into the host genome. These retroviruses also contain a conserved RNA structural element, named the core encapsidation signal, which is crucial for genome packaging during viral assembly (D'Souza *et al.*, 2004).

The MSCV retroviral vector was developed from the MLV virus due to difficulties in expressing viral vectors in embryonic and haematopoietic stem cells. The first difficulty was due to repression by *trans*-acting regulatory factors expressed by the host cell upon infection. This repressive action was supplemented by host cell *cis*-factors, which effectively prevented expression of the retroviral vectors (Laker *et al.*, 1998). Three important modifications were made to the MLV LTR site; a high affinity binding site for the *SP1* transcription factor was introduced, the binding site for the *ECF-1* regulatory transcription factor was disrupted and the negative regulatory element was removed, eliminating a potent repressor of LTR mediated transcription (Laker *et al.*, 1998). Other modifications include the addition of the GFP marker (Cherry *et al.*, 2000) and the use of the IRES, which allows for the expression of multiple genes from one promoter, ensuring expression of the selected marker in tandem with the target insert (Coffin *et al.*, 1997).

To further facilitate viral integration into haematopoietic cells, specific gag/pol and env sequences have been isolated and optimised. The RD114 envelope is derived from the feline leukaemia virus, another gamma retrovirus. This envelope provides greater particle stability, lower host cell toxicity, and the expression of the neutral amino acid transporter receptor, which is widely expressed in HSCs (Bell *et al.*, 2010). Gag/pol sequences have also been adapted for use, the m57 gag/pol fragment was isolated from the MLV, giving separate functional expression of pol and gag proteins (Enssle *et al.*, 1996).

4.2 Aims

Using the knowledge acquired from the *IGH-CEBP* patient cohort, the focus of this chapter was to investigate the consequences of *CEBP* overexpression in a haematopoietic setting. *CEBPD*, the most commonly occurring *IGH-CEBP* partner, and IK6, a commonly occurring genetic lesion within the cohort, were selected to be cloned into retroviral vectors and expressed in CD34+ cells. To investigate the effects of *CEBPD* overexpression on haematopoietic

development, using *in vitro* and *in vivo* modelling. The development and initial optimisation of this model was performed in collaboration with Dr. James Mulloy's group at the Cincinnati Children's Hospital, and was based on a previous study developed by the group which created lymphoid lineage cells (Mulloy *et al.*, 2003). One pilot and two additional experiments, Experiments 1 and 2, were performed in Cincinnati, after which the technique was brought to Newcastle to be repeated in our laboratory. After several optimisation steps, transduction of CD34+ cells was performed successfully and Experiments 3 and 4 were performed.

The aims of this section of the project were as follows:

- To clone *CEBPD* and IK6 cDNA into appropriate retroviral vectors.
- To create retroviral particles and successfully overexpress *CEBPD* in CD34+ cells.
- To create retroviral particles and successfully express IK6 in CD34+ cells.
- To express CEBPD and IK6 retroviral vectors together in CD34+ cells.
- To optimise the CD34+ cell transduction protocol within our laboratory in Newcastle.
- To expand transduced cells *in vitro* and *in vivo*.
- To track lineage progression of transduced cells and monitor for the emergence of a CD19+ lymphoid lineage.
- To sort transduced cells by lineage for comparative gene expression array analysis.

4.3 Method Development and Results

4.3.1 Generating Retroviral Particles

To express target genes, retroviral vectors were first created by inserting the desired target cDNA sequence into the relevant MSCV-IRES plasmids (Section 2.3.6).

4.3.2 Confirming CEBPD cDNA presence and sequence

A full coding cDNA transcript for the *CEBPD* gene was identified (NCBI genome browser) (Supplementary Sequence 7.1) and purchased (Fermentas, Germany). The construct was received as a bacterial stab, which was expanded using standard bacterial culture techniques (Section 2.3.2.2.2). The *CEBPD* construct was extracted from bacterial pellets (Section 2.3.2.2.2) and was excised from the pCR4-TOPO plasmid (Supplementary Figure 7.2) by performing a restriction digest using the EcoRI site (Section 2.3.4.2). The digested DNA was run on a 1% agar gel to confirm the presence of the 879bp *CEBPD* insert. Three pCR4-TOPO-CEBPD colonies were grown and extracted, DNA from colony 1 was discarded as the sample displayed multiple bands indicating nonspecific DNA digestion (Figure 4.2). *CEBPD* colonies 2 and 3 exhibited bands between the 750bp and 1000bp DNA ladder indicating the presence of the *CEBPD* transcript (Figure 4.2). Undigested DNA from *CEBPD* colonies 2 and 3 was sent for sequencing (DBS Genomics, England), using the -21M13 and M13 primers (Supplementary Table 7.25), a common primer sequence cloned into the pCR4-TOPO construct, to confirm the presence of the *CEBPD* sequence. Both colonies were found to contain the full *CEBPD* cDNA sequence and were suitable for downstream applications.

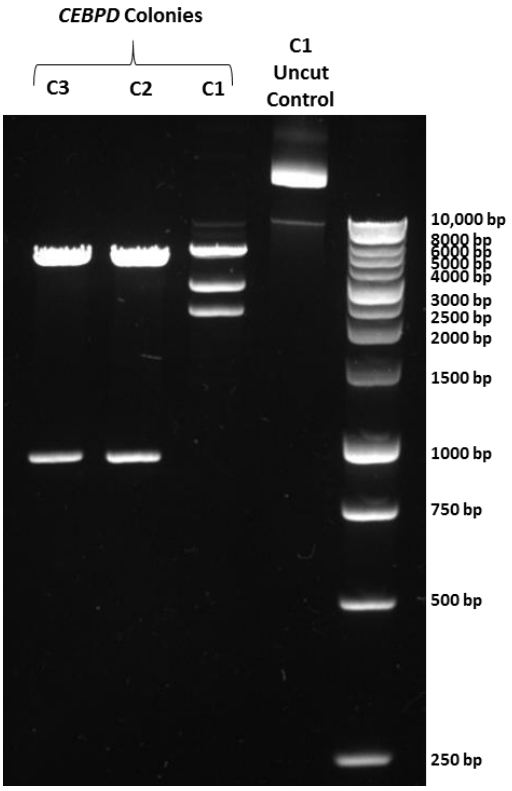


Figure 4.2. Excised pCR4-TOPO-CEBPD colony DNA run on 1% agarose gel, with uncut DNA control.

Colony 1 shows multiple DNA bands indicating unspecific DNA digestion by the EcoRI restriction enzyme. Colonies 2 and 3 show two specific bands, the ~6000bp band indicating the pCR4-TOPO construct and the ~1000bp band indicating the CEBPD sequence.

To isolate the *CEBPD* sequence for downstream applications, pCR4-TOPO-CEBPD colonies 2 and 3 were re-digested with the EcoRI restriction enzyme and samples were run on a 1% agarose gel. Bands at ~1000bp were excised and DNA was extracted using a gel extraction kit (Section 2.3.4.3) to collect the excised *CEBPD* cDNA transcript. The transcript was stored at -20^{o}C for future use with selected retroviral vectors.

4.3.2.1 Retroviral Vectors MIGR1, MIVR1 and MiT isolation and preparation.

Retroviral vectors used in this study were kindly donated by Dr. James Mulloy and Dr. Eric Clambey. All vectors were variations of the MSCV retroviral vector. MIGR1 retroviral vector contained the EGFP marker (MSCV-IRES-EGFP) (Supplementary Sequence 7.2), the MIVR1 vector contained the venus marker (MSCV-IRES-Venus) (Supplementary Sequence 7.3) and the MiT vector contained the Thy1 marker (MSCV-IRES-Thy1) (Section 2.3.6).

Retroviral vector plasmids, MIGR1, MIVR1 and MiT, were harvested from Whatman paper (Section 2.3.4.1) and inserted into Stbl3 *E.coli* (Section 2.3.4.6) to be expanded for glycerol and DNA stocks. Plasmids were grown using standard bacterial culture methods with the chloramphenicol antibiotic (Section 2.3.2). Three individual colonies were selected per plasmid, grown up and DNA extracted (Section 2.3.2.2). PvuII restriction digests were performed to confirm the presence of the vectors in individual colonies (Section2.3.4.2). The PvuII enzyme was selected as it creates three bands for the vectors at sizes ~3700bp, ~2360bp and ~430bp. The PvuII digest worked successfully on all colonies (Figure 4.3).

Four different bands were observed; ~375bp, ~1750bp, ~2500bp, and ~4500bp (Figure 4.3). MiT and MIGR1 vectors expressed the expected fragment sizes at ~375bp, ~2500bp, and ~4500bp (Figure 4.3). Bands for the MIVR1 vector were smaller than expected, with fragments at ~1750bp and no fragment at ~4500bp visible. However bands were consistent across all colonies, discounting the possibility of a random loss of genetic material. All colonies were sent for sequencing (DBS genomics, England), confirming vector sequences. Colonies were expanded for DNA and glycerol stocks.

While initially all three vectors were considered for use, ultimately only two were needed. The MIGR1 and MiT vectors were selected for further downstream applications, MIGR1, due to ease of visualisation of the EGFP protein, and MiT, due to the variability of the Thy1 marker, which could be biotin labelled with a variety of fluorochromes and also used for magnetic bead selection.

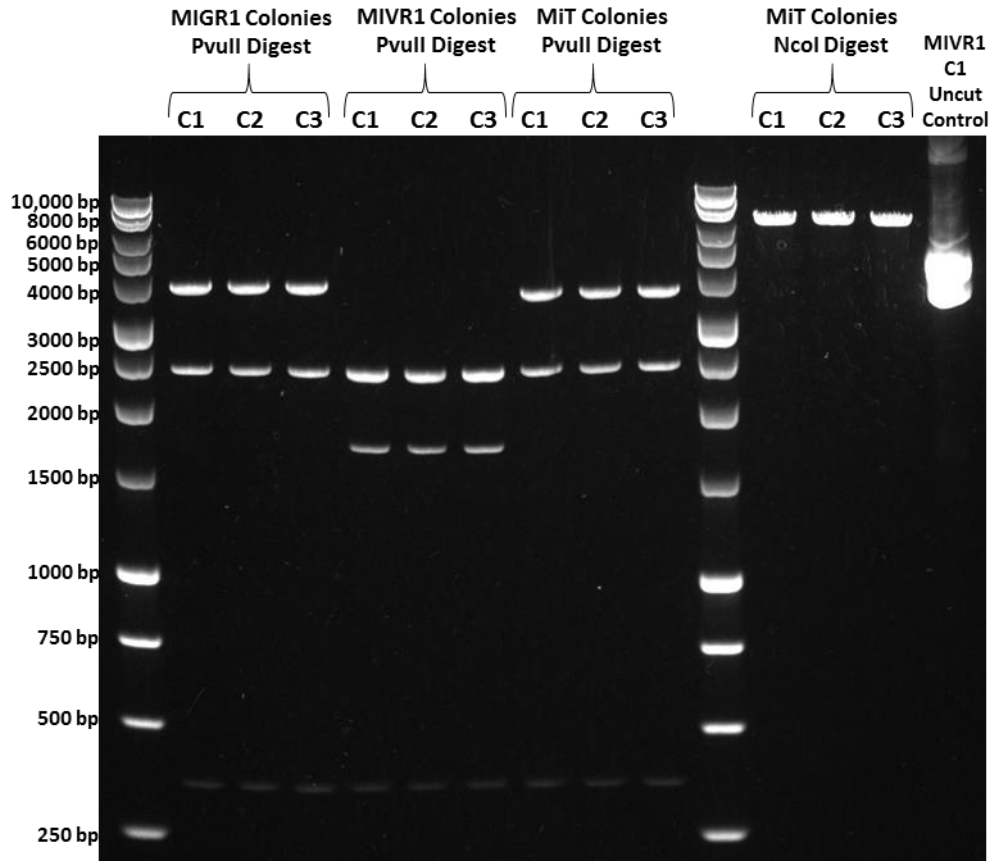


Figure 4.3. Restriction digest of retroviral vector MIGR1, MIVR1 and MiT colonies using the PvuII enzyme, with MIVR1 C1 uncut control run on a 1% gel.

All colonies show consistent DNA bands. Retroviral plasmids for the MiT and MIGR1 vectors show expected bands at ~375bp, ~2500bp, and ~4500bp. MIVR1 plasmid shows presence of a smaller band ~1750bp and no fragment at ~4500bp.

4.3.2.2 Cloning of the CEBPD sequence into the MIGR1 retroviral vector

With the *CEBPD* transcript isolated and the retroviral vectors extracted, the next stage focused on cloning the *CEBPD* transcript into the MIGR1 retroviral construct.

The MIGR1 vector was digested using the EcoRI restriction enzyme, creating a staggered cut end to the double stranded DNA, leaving an overhang of nucleotides. This overhang readily adheres to a complementary overhang to allow for more efficient integration of the selected DNA insert. To prevent re-annealing of the cut plasmid, the sample was treated with TSAP (Section 2.3.4.4), and run on a 1% agarose gel to confirm digestion of the plasmid. The uncut coiled MIGR1 plasmid was carried through the gel at a greater speed than the linearized

plasmid, resulting in different positions of the two samples in the agarose gel (Figure 4.4).

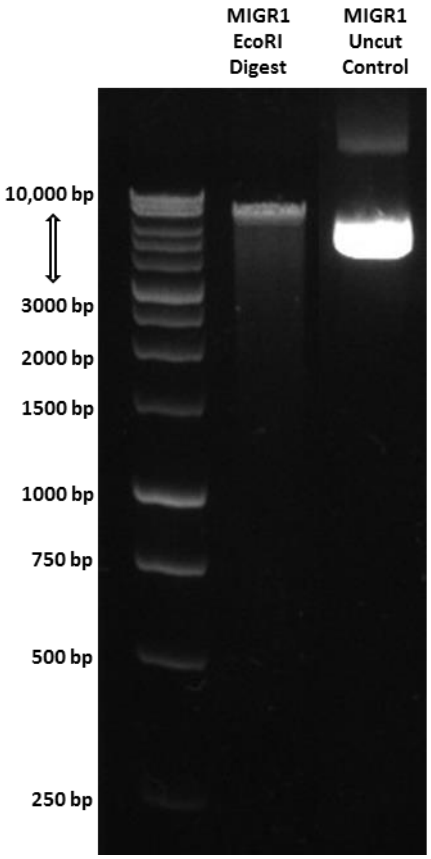


Figure 4.4. EcoRI digest of the MIGR1 retroviral vector with an uncut MIGR1 control, run on a 1% agarose gel.

While the DNA ran on the gel is identical, there is a clear disparity in band position, due to linearised plasmid traveling through the agarose gel at a slower pace in comparison to the coiled uncut plasmid. The CEBPD *cDNA construct was added to the vector and ligated. The resulting DNA construct was expanded in Stbl3* E.coli*. Custom primers were designed to span from within the* CEBPD *insert across the ligation point into the MIGR1 construct, with successful integrations creating PCR products ~250bp in size. Cloned colonies were analysed on a 1% agar gel, two colonies displayed PCR bands for CEBPD-MIGR1 clones confirmed via sequencing.*

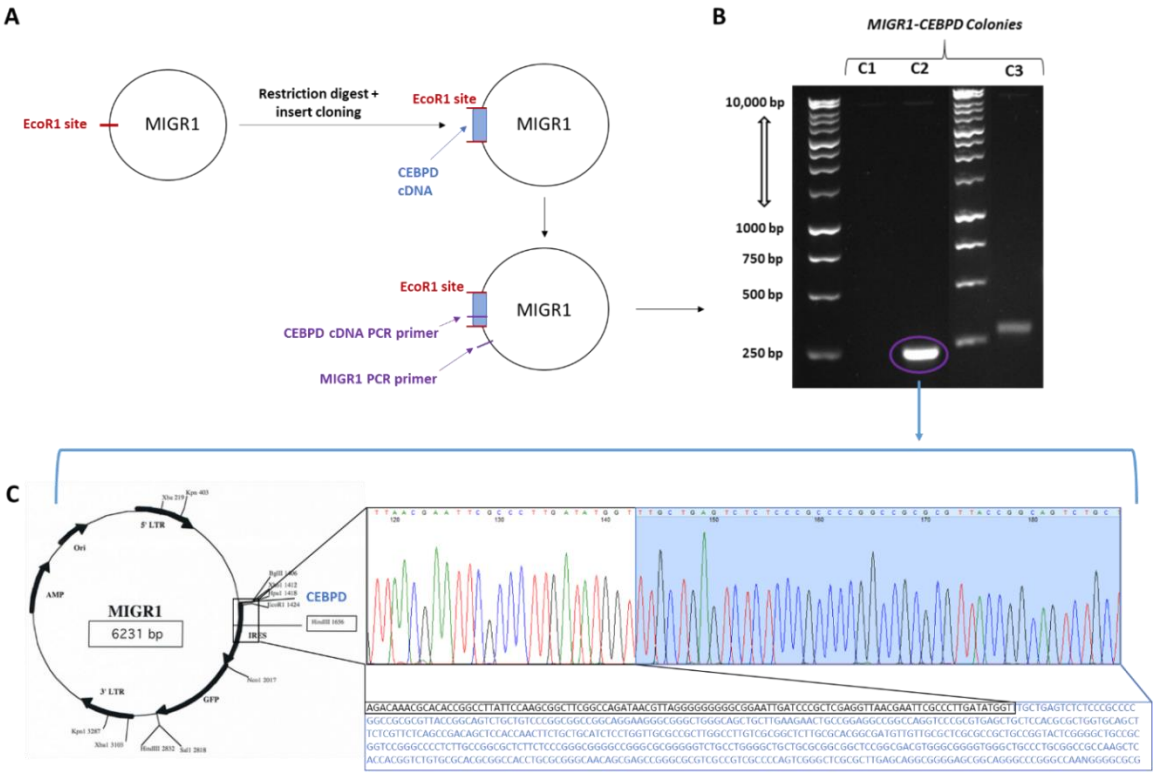


Figure 4.5. Cloning of the CEBPD cDNA insert into the MIGR1 vector and subsequent confirmation of the product.

A. Strategy for identifying CEBPD cDNA integration into the MIGR1 vector using PCR primers. B. PCR products of cloned MIGR1-CEBPD colonies run on a 1% agarose gel. Image was created from two separate gel images. Colonies 2 and 3 showing PCR products at ~250bp for primers spanning from the MIGR1 sequence to the CEBPD sequence, indicating successful integration of the CEBPD cDNA construct. C. A chromatogram view of CEBPD cDNA sequence present in MIGR1-CEBPD colony 2, with the blue segment and letters indicating the presence of the CEBPD cDNA and the black segment and letters representing the MIGR1 vector sequence.

4.3.2.3 Preparation of MiT-IK6 retroviral vector.

The *IKZF1* dominant negative isoform, IK6, was selected as the associating abnormality to be expressed in combination with *CEBPD* in CD34+ cells.

The IK6 transcript was kindly donated by Dr. Eric Clambey both individually and cloned into the MiT vector. All DNA was received on Whatman filter paper. The plasmids were harvested (Section 2.3.4.1), transformed in Stbl3 *E.coli* (Section 2.3.4.6) and expanded for glycerol stocks and DNA. Five colonies of the MiT-IK6 construct were extracted to confirm the presence of the IK6 insert. The 869bp IK6 cDNA (Supplementary Sequence 7.4) was excised from the MiT by performing a double digest of the construct using BglII and NotI restriction enzymes (Section 2.3.4.2). The digested DNA was analysed on a 1% gel, where the IK6 insert was observed at ~900kb, with a second band produced at ~6.1kb for the MiT vector

(Figure 4.6). Colonies were expanded and stored as glycerol stocks and DNA extracted for future use.

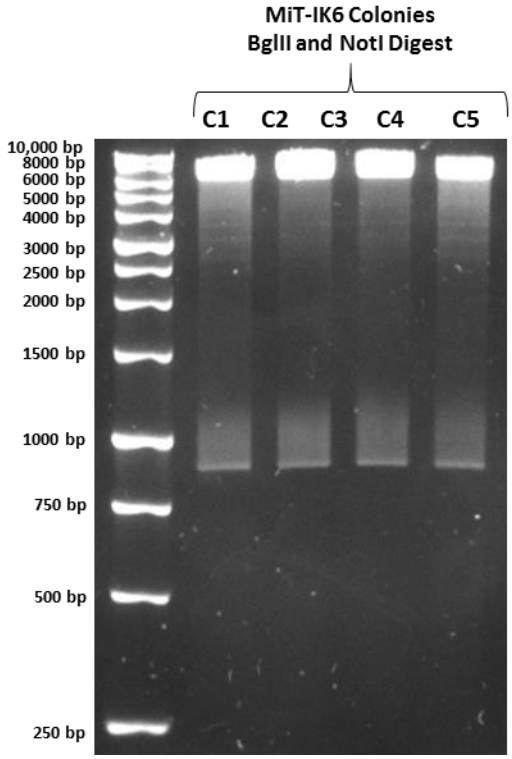


Figure 4.6. Restriction digest of MiT-IK6 construct colonies, run on a 1% gel. All colonies indicate the presence of the IK6 cDNA sequence at ~1000bp, with the MiT vector visible at ~6000 bp.

4.3.2.4 Creation of Retroviral Particles

Empty vector controls for MIGR1 and MiT were added to the prepared MIGR1-CEBPD and MiT-IK6 retroviral constructs (Section 2.3.6), ready for retroviral particle production. The M57 gag/pol plasmid and the RD114 envelope plasmid (Supplementary Sequence 4.1), kindly supplied by Dr. James Mulloy, were used with the selected constructs to create functioning retroviral particles in 293T cells (Section 2.3.6.1). Media was harvested over three days (Section 2.3.6.2) and used immediately for the transduction of CD34+ cells or stored at -80^{o}C for future use.

Retroviral function was tested by harvesting protein lysate from 293T cells used in the creation of the respective retroviral particles (Section 2.3.2.4). The presence of the *CEBPD* protein and IK6 protein isoform were determined using the Western Blot technique (Section 2.3.11.2). Empty vectors for MIGR1 and MiT and non-transduced 293T cells served as controls (Figure 4.7). 293T cells used

in the production of retroviral particles showed clear expression of the respective protein. CEBPD, at ~28kDa, was highly expressed in comparison to both controls (Figure 4.7 A). Only IK6 293T cells showed expression of the IK6 isoform present at ~37kDa (Figure 4.7 B).

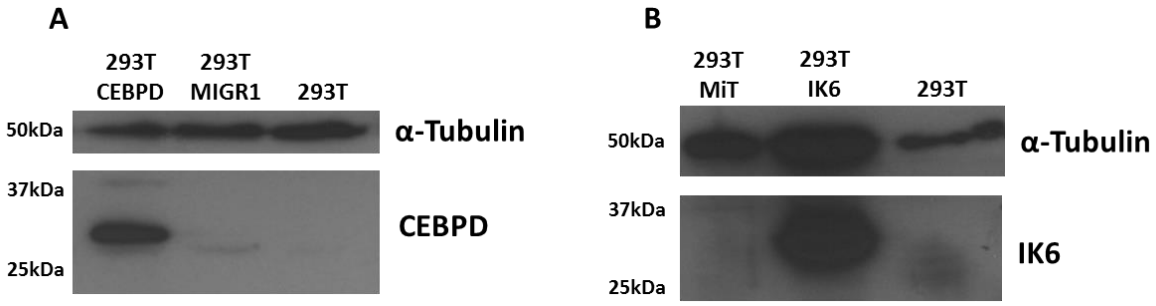


Figure 4.7. Western immunoblots of 293T cells used in the production of retroviral vectors.

A. Western immunoblot showing CEBPD expression in 293T cells used for CEBPD retroviral production, 293T cells used for MIGR1 empty vector retroviral production and parental 293T cells, with α-Tubulin loading control. B. Western immunoblot showing IKZF1 antibody expression in 293T cells used for IK6 retroviral production, 293T cells used for MiT empty vector retroviral production and wild type 293T cells, with α-Tubulin loading control.

4.3.3 Transduction of CD34+ Cells in Cincinnati and Newcastle

Having successfully prepared the MSCV-IRES plasmids, the next stage of the protocol was to create the relevant retroviral particles and to transduce haematopoietic progenitors for modelling of *CEBPD* overexpression. To perform this task, I worked for six months with Dr. James Mulloy's group at the Cincinnati Children's Hospital with the aim of achieving the following targets:

- To isolate CD34+ cells from cord blood.
- To transduce the isolated CD34+ cells with retroviral particles expressing *CEBPD* and IK6.
- To inject the transduced CD34+ cells into mice to create xenograft (*in vivo*) models.
- To culture transduced CD34+ cells *in vitro* with and without a stromal co-culture.
- To use flow cytometry to analyse cell surface marker expression to track lineage differentiation.
- To identify *CEBPD* expressing CD19+ and CD33+ cells arising *in vivo* or *in vitro* and to isolate them by FACS sorting.

Having achieved the above aims and acquired the skills to perform CD34+ cell isolation and transduction, I returned to our laboratory in Newcastle to repeat the transduction experiments, which I had performed in Cincinnati. However, fresh cords were not available in Newcastle, resulting in the purchase of isolated CD34+ cells (Allcells, England). The aims for work performed in Newcastle were:

- To repeat the transduction of CD34+ cells with *CEBPD* in Newcastle.
- To repeat the aims and experiments set in Cincinnati.
- To perform molecular analysis on flow sorted cells with the aim of identifying the function of *CEBPD* in lineage differentiation in haematopoietic cells.

While the above aims were met and two further experiments were performed *in vivo* and *in vitro* (Experiment 3 and Experiment 4), there were several challenges which had to be overcome before CD34+ transduction could be performed successfully in Newcastle. Several early attempts to transduce CD34+ cells failed and optimisation steps were performed prior to transduction. The issues which arose and subsequent optimisation steps performed are detailed below.

4.3.4 Optimisation of CD34+ Cell Retroviral Transduction Protocol in Newcastle

The protocol selected for the transduction of CD34+ cells with the prepared retroviral vectors was based upon a protocol from the Mulloy laboratory (Mulloy *et al.*, 2003). This protocol, using a combination of SCF, FLT-3L and IL-7 cytokines co-cultured with MS-5 stromal cells, had been developed to culture *RUNX1-ETO* transduced CD34+ cells for an extended period of time, which resulted in an expansion of both lymphoid and myeloid cell lineages.

When initially performing this protocol in Cincinnati, CD34+ cells were co-cultured with both MS-5 and HS-5 stromal cell lines, to investigate efficacy of the co-cultures in expansion of lymphoid cells. The MS-5 cell line was selected for future use as results from both stromal co-cultures showed little variation (data not shown) and MS-5 stroma had been used by the Mulloy group.

Several difficulties were experienced during the recreation of the transduction experiments in Newcastle, those difficulties are discussed below, after which all experimental data from Cincinnati and Newcastle is presented together.

4.3.4.1 Protocol Optimisation: Retroviral Solution pH

The first challenge encountered during the attempted transduction of CD34+ cells in Newcastle was optimisation of the pH of the calcium phosphate solution. A fine precipitate should be created to improve uptake by recipient 293T cells. After combination, cultures are viewed using a light microscope to search for crystalline formations in the medium, indicating the formation of a calcium-phosphate-DNA co-precipitate. This co-precipitate facilitates binding to the cell surface, resulting in improved endocytic integration. For successful calcium phosphate co-precipitation, the pH value of the final solution must be between pH 7.0 – 7.2, optimally at pH 7.1. Solution pH had to be readjusted after filtration.

4.3.4.2 Protocol Optimisation: Foetal Bovine Serum Batch Tests

After optimisation of pH, the second challenge encountered in Newcastle was failure of CD34+ cells to exhibit transduction after culturing with viral supernatant. The CD34+ cells showed low GFP expression for both CEBPD (5%) and MIGR1 (7%) (Figure 4.8 A). To determine the source of the failure, 293T cells were used to produce new retroviral particles, then analysed by FACS to determine the GFP levels in the cells. After confirming GFP expression in CEBPD (71%) and MIGR1 (88%) 293T cells (Figure 4.8 B), other variables were tested.

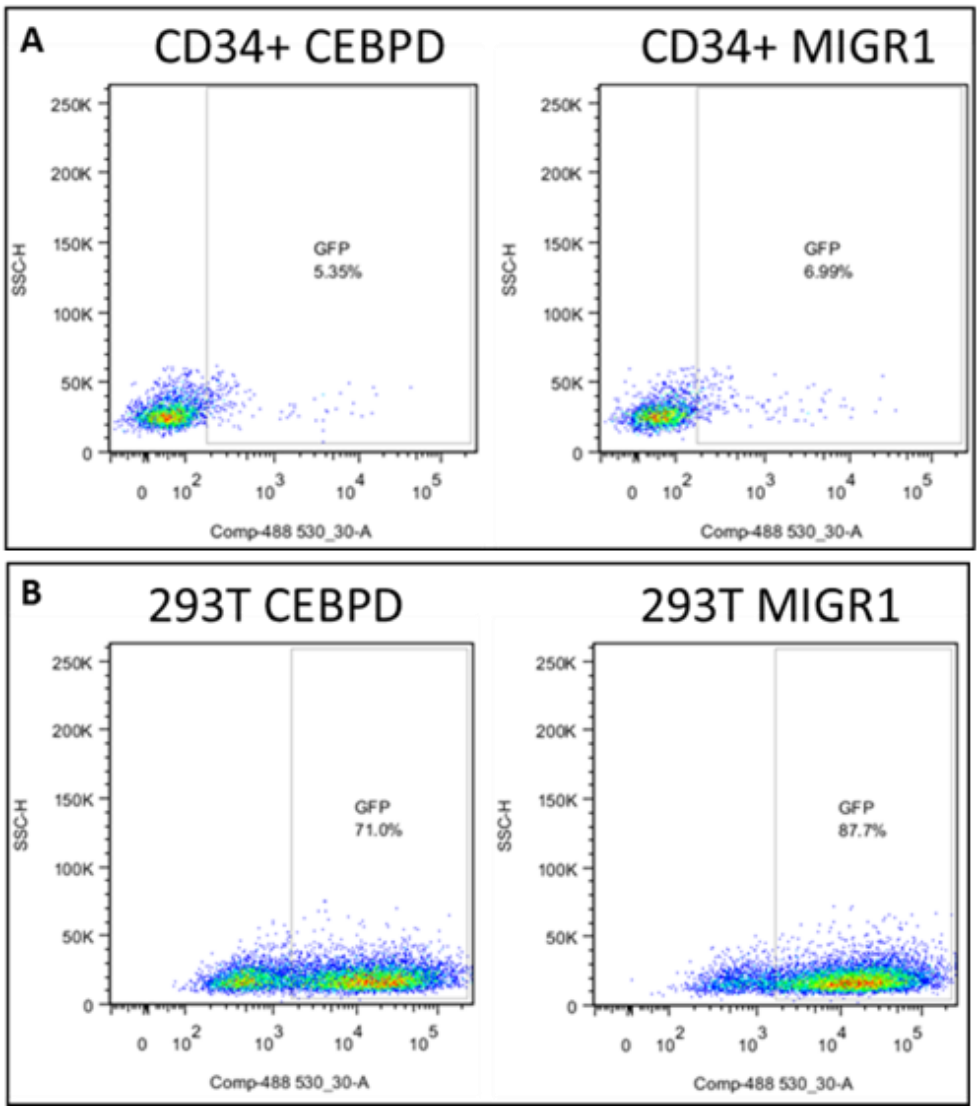


Figure 4.8. FACS analysis of GFP expression measured by FITC expression.

A. CD34+ cells from the first transduction experiment in Newcastle, showing low GFP expression in CEBPD and MIGR1 infected cells. B. 293T cells used in a subsequent culture to produce CEBPD and MIGR1 retroviral particles show strong expression of GFP, indicating retroviral particles were produced, but did not integrate into CD34+ cells.

CD34+ cells were cultured for 24 and 48 hours in KTF 100ng/ml cytokine medium, to determine whether CD34+ cells cultured from frozen aliquots required a longer incubation period in the cytokine rich environment than fresh cells. No proliferation was observed (data not shown). Cells were re-plated in standard KTF36 10ng/ml cytokine medium and monitored over 12 days (Figure 4.9 A). As no growth was seen, different sources of FBS were used to determine whether endotoxin levels were responsible for failure of cell expansion. High performance FBS, with endotoxin level ≤ 5 EU/ml, was ordered (Gibco, USA) and 8.6 x 10^4 cells were re-plated and grown in the medium containing the new FBS. CD34+

cells were plated out and counted over 23 days, at which point cell growth was observed (Figure 4.9 B).

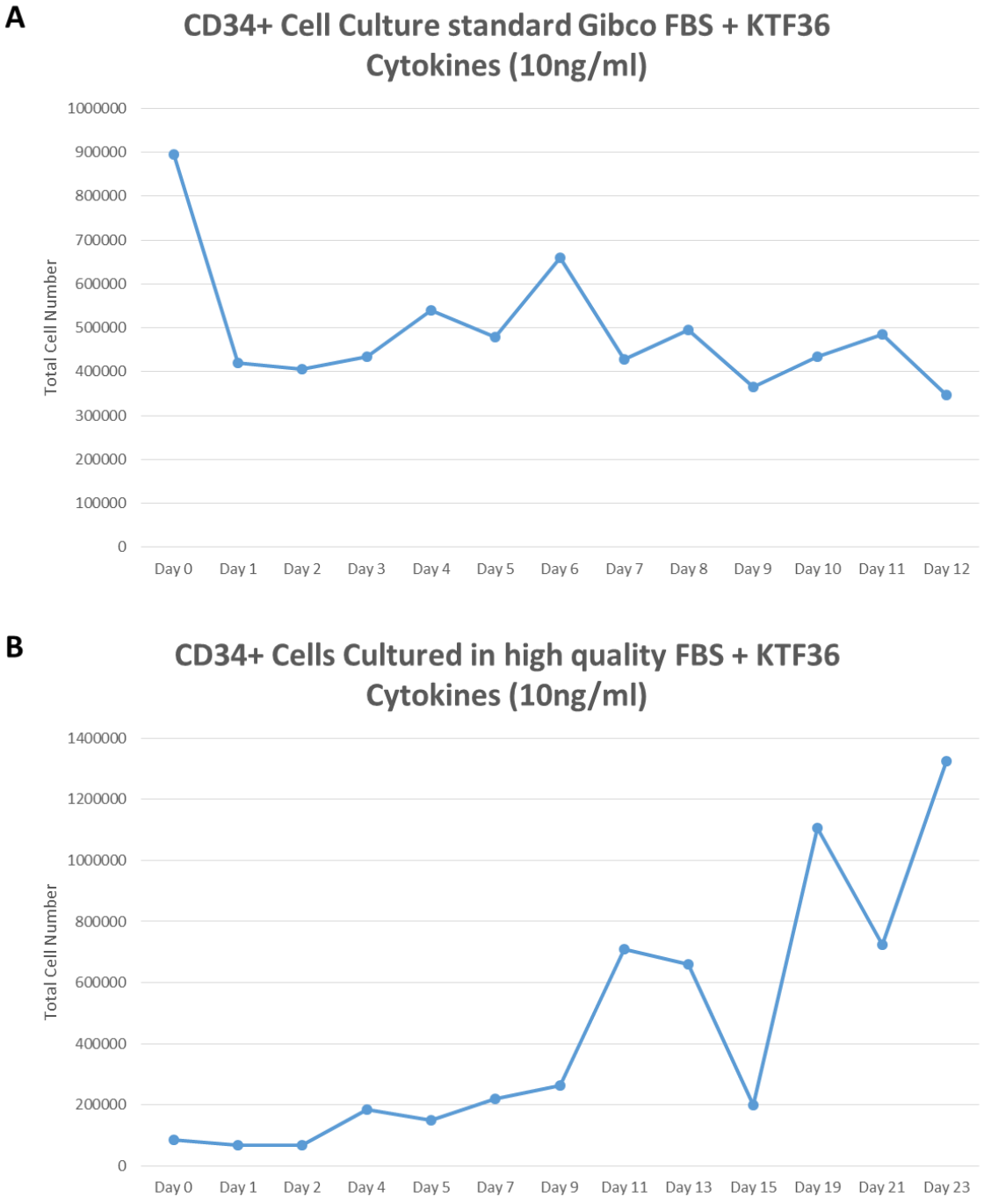


Figure 4.9. Absolute cell counts of CD34+ cells depicting cellular proliferation.

A .CD34+ cells grown in culture medium with standard FBS with SCF, TPO, Flt3-L, IL-3 and IL-6 at 10ng/ml over 12 days. B. CD34+ cells grown in high quality FBS with SCF, TPO, Flt3-L, IL-3 and IL-6 at 10ng/ml over 23 days, showing improved CD34+ cell proliferation.

High quality FBS (FBS1) and two standard FBS sources, non-USA origin FBS batch F7524-024M3398 from Sigma (FBS2), and USA origin FBS batch F2442-13E120 from Sigma (FBS 3), were analysed to determine if only high quality, low endotoxin FBS was suitable for expansion of CD34+ cells. Heat inactivation of FBS was also tested (Section 2.3.3.5), as this was the standard protocol in the Mulloy lab. Cells were plated out at 4.3 x 10^5 cells per well and monitored over 20 days. Growth was observed in all batches of FBS. Heat inactivation appeared to improve CD34+ proliferation in all three FBS batches (Figure 4.10). Heat inactivated FBS2 exhibited high growth, and showed the greatest variation between heat inactivated and non-heat inactivated FBS (Figure 4.10 B). FBS3 exhibited high growth for both heat inactivated and non-heat inactivated FBS (Figure 4.10 C&D). A peak in cell proliferation was observed in all cultures nearing the end of the experiment. This peak declined in FBS1 and FBS3 (Figure 4.10 A&C), but continued in FBS2 heat inactivated cultures (Figure 4.10 B), suggesting that the cells had reached the end of their proliferative capacity in cultures FBS1 and FBS3.

Using these assays, it was determined that the batch of FBS initially used for CD34+ culture was unsuitable and that F3 was the best choice for long term culture of the cells, due to the highest mean cell counts (Figure 4.10 E).

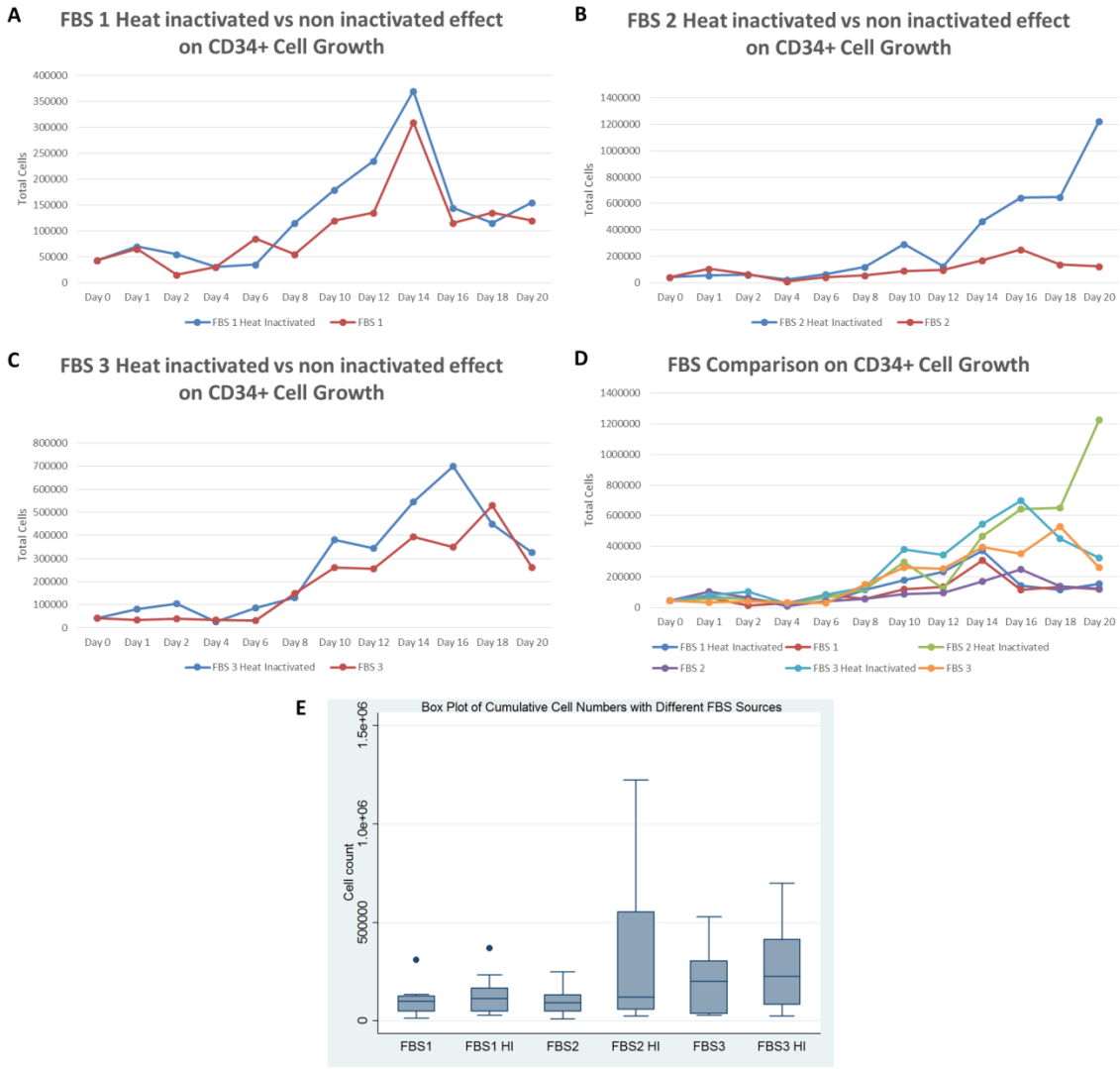


Figure 4.10. Absolute cell counts of CD34+ cells examining the effects of three FBS batches and the effects of heat inactivation (HI) on proliferation of cells grown with SCF, TPO, Flt3-L, IL-3 and IL-6 cytokines at 10ng/ml over 20 days.

A .FBS batch 1. B. FBS batch 2. C. FBS batch 3. D. Comparison of all FBS batches on cell proliferation. E. Box plot of cell counts across all FBS variants, showing distribution and mean values, FBS3 and FBS3 HI show the highest average cell counts in comparison to other FBS batches.

4.3.5 Experimental set up of Retroviral Transduction Experiments

Over the course of this study, four transduction experiments were performed. Experiments 1 and 2 were designed to investigate the effects of *CEBPD* alone and in combination with IK6 in human CD34+ cells. However, focus was shifted to *CEBPD* overexpression only for experiments 3 and 4, as IK6 showed little effect on transduced cells (Table 4.1) (Sections 4.3.5.3). All transduced cells were divided for *in vitro* and *in vivo* studies (Figure 4.11). Cells allocated to *in vitro* studies were divided between liquid and stromal cultures. All stromal cultures were performed using MS-5 cells (Figure 4.11). Liquid and stromal culture cell

numbers varied between individual experiments (Table 4.1). Xenograft injections were performed intrafemoraly on Busulfan conditioned (Section 2.3.7) female NSG mice using 75,000 transduced cells. Experimental variation between studies is displayed in Table 4.1.

Location	**Experiment No.**	**CD34+ Cell Source**	**Transduced CD34+ Cell Populations**	**Experiments *in vitro* / *in vivo***	**Mice per Population**	**Cells *in vitro* Culture**	**Tissue Culture Setting**
Cincinnati	1	Fresh	CEBPD	Both	3	40,000	6 well tissue culture plates
			IK6	*in vitro*	0		
			CEBPD + IK6	*in vitro*	0		
			MIGR1 + MiT	Both	3		
Cincinnati	2	Fresh	CEBPD + MiT	Both	5	40,000	6 well tissue culture plates
			IK6 + MIGR1		5		
			CEBPD + IK6		5		
			MIGR1 + MiT		5		
Newcastle	3	Frozen	CEBPD	Both	3	40,000	6 well tissue culture plates
			MIGR1		2		
Newcastle	4	Frozen	CEBPD	Both	3	100,000 *	10 cm tissue culture dish
			MIGR1		3		

Table 4.1. Table displaying variations in experimental set up of all transduction studies.

Experiments 1 and 2 were carried out in Cincinnati, using freshly isolated CD34+ cells for transduction experiments, Experiments 3 and 4 replicated the Cincinnati protocols in Newcastle with frozen CD34+ cell aliquots. For Experiment 1, CD34+ cells were divided into four subsets for transduction with the following retroviral particle combinations; CEBPD alone, IK6 alone, CEBPD+IK6 double transduced, MIGR1+MiT double transduced, and a fraction of non-transduced CD34+ cells. For Experiment 2 transduced populations were altered with empty vector controls added to the targeted retroviral populations, giving the following populations; CEBPD+MiT, IK6+MIGR1, CEBPD+IK6, and MIGR1+MiT. Experiments 3 and 4 only focused on CEBPD and MIGR1 retroviral vectors. . All experiments were performed in vivo and in vitro, with the exception of IK6 and CEBPD + IK6 cells from Experiment 1 due to low NSG mouse numbers available. All in vitro experiments were conducted in 6 well tissue culture plates with the exception of Experiment 4 which was performed in 10cm tissue culture treated dishes to increase the number of viable transduced cells for downstream applications. Experiment 4 stromal culture was seeded with 100,000 cells, liquid culture was seeded with 40,000 cells.*

Cell numbers were assessed twice weekly and maintained at 5×10^5 cells per ml. Cell surface marker expression was analysed once each week using flow cytometry to determine the effects of retroviral vectors on lineage differentiation.

From this point forward, the MIGR1-*CEBPD* construct will be referred to as CEBPD, the MiT-IK6 construct as IK6, the MIGR1 empty vector control as MIGR1, and the MiT empty vector control as MiT.

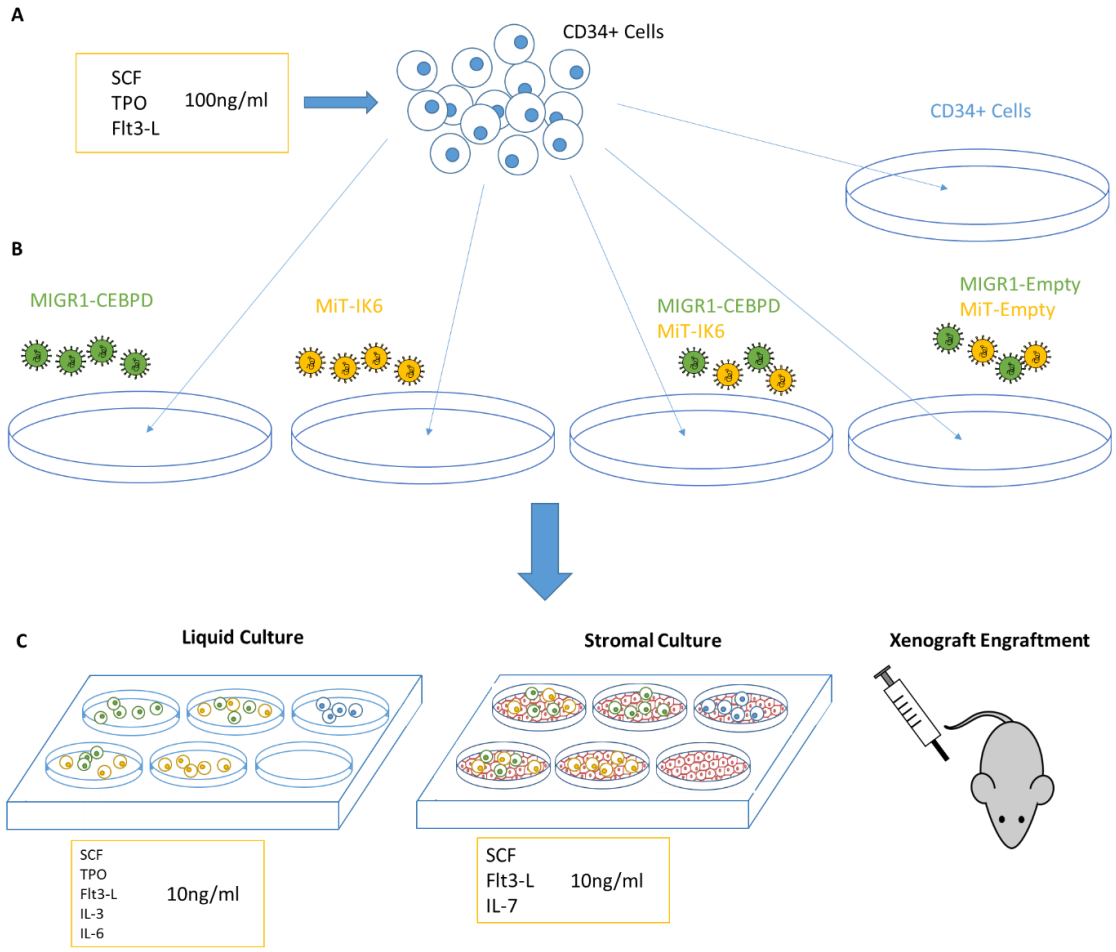


Figure 4.11. Transduction of CD34+ Cells, Experiment 1.

A. CD34+ cells are stimulated in a SCF, TPO, FLT3L 100ng/ml cocktail for 16 hours to stimulate growth. B. CD34+ cells were divided into 5 fractions and transduced with the created retroviral particles and retroviral controls over three days, some cells were left untransduced as controls. C. Cell fractions were divided for three experiments, in vitro suspension culture in standard suspension culture with SCF, TPO, FLT3L, IL-3 and IL-6 at 10ng/ml, in vitro stromal culture conditions with SCF, FLT3L and IL-7 at 10ng/ml, and for xenograft experiments.

4.3.5.1 Cellular Transduction

Transduction efficiency of retroviral vectors was analysed using flow cytometry (Section 2.3.8.3). GFP expression represented cells transduced with CEBPD and MIGR1 vectors, and Thy1 represented cells transduced with IK6 and MiT vectors. Expression of the Thy1 marker was visualised by conjugation with the PE

fluorochrome (Section 2.3.8.4). GFP expression in CEBPD transduced cells was interpreted as expression of the CEBPD protein, Thy1 expression in IK6 transduced cells was interpreted as expression of the IK6 protein isoform. Marker expression was analysed using the geometric mean, which is a more appropriate method of analysis for data points such as percentages gauged from human observation. The geometric mean is calculated by taking the same data points as the arithmetic mean, but instead of adding the points together and dividing them by the number of data points, the geometric mean multiplies the data points together and divides them by the nth root of the product, with n being the number of data points.

4.3.5.2 GFP expression rapidly decreases in CEBPD transduced cells but not in MIGR1 control cells suggesting that CEBPD expression is hindering proliferation

Transduction rates of CEBPD and MIGR1 were high for the majority of experiments, with an average of 56% and 47%, respectively. One notable exception was MIGR1 cells in Experiment 4 (Figure 4.12 G&H). GFP expression displayed a similar trend across three of the four experiments. Experiments 1, 2, and 3, all showed a more rapid loss of GFP expression in cells transduced with the CEBPD retrovirus in comparison to MIGR1 empty vector controls (Figure 4.12 A-F). Additional expression of IK6 or MiT with CEBPD did not show any additional trends, with GFP only affected by CEBPD expression (Figure 4.12 A-D). Experiment 4 displayed differing results to previous findings, with MIGR1 GFP expression being lower than that of CEBPD transduced cells. This was due to a very low starting transduction level of 1%. However GFP expression in these cells did increase over the course of the experiment, while Experiment 4 CEBPD cell behaviour was identical to the previous three experiments, starting with a high GFP level which was lost over the course of the experiment. GFP expression was generally higher in stromal cultures (geometric mean (GM) = 16%) in comparison to liquid cultures (GM = 8%). These findings suggest that CEBPD expressing cells show decreased proliferative capabilities, while the expression of IK6 exerted no additional biological effects

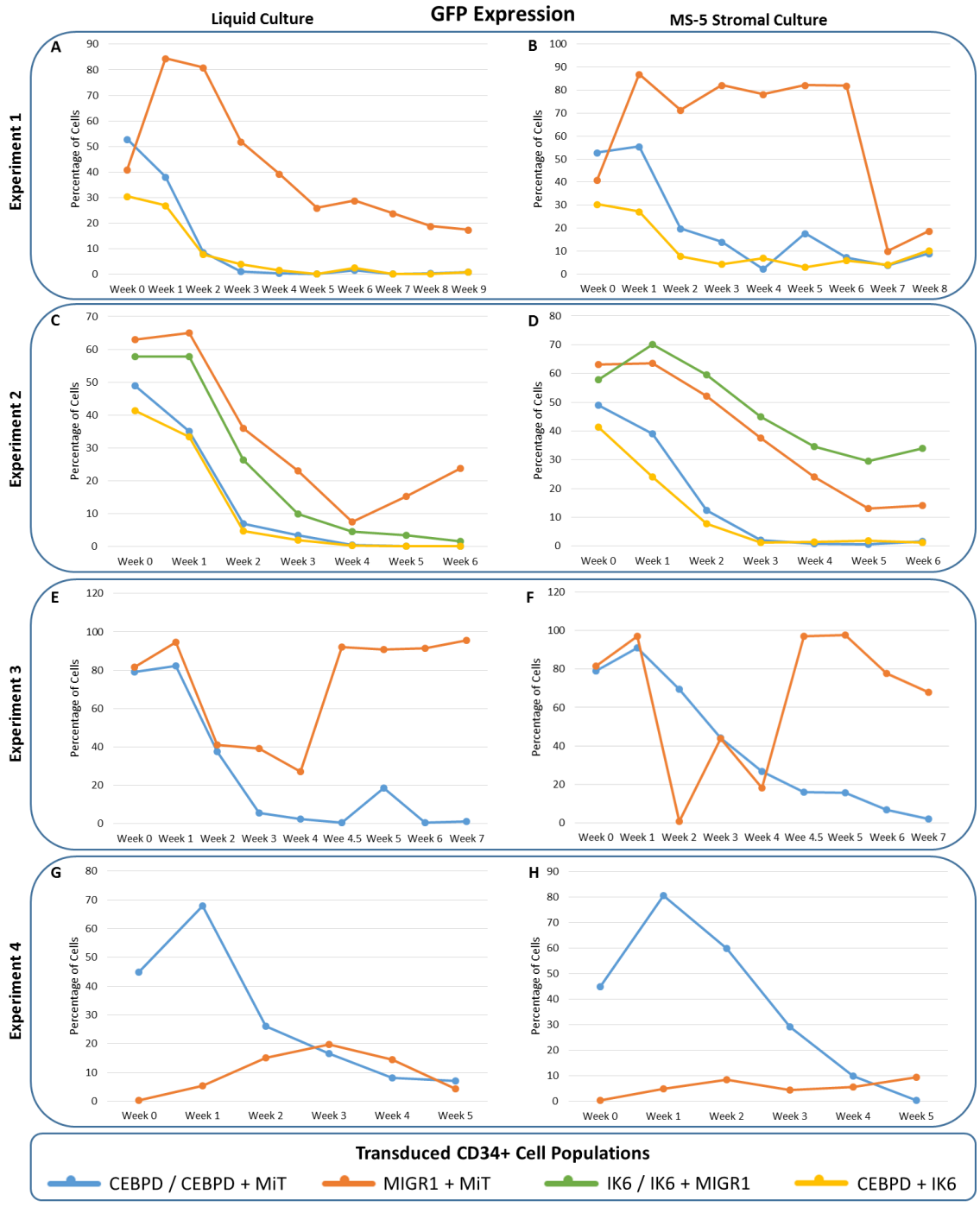


Figure 4.12. Transduction levels of GFP in CD34+ cells across four experiments under two different in vitro culture conditions.

A. Experiment 1: GFP percentages in CEBPD, CEBPD+IK6 and MIGR1+MiT CD34+ transduced cells across 9 weeks in suspension culture. B. Experiment 1: GFP percentages in CEBPD, CEBPD+IK6 and MIGR1+MiT CD34+ transduced cells across 9 weeks on MS-5 stromal co-culture. C. Experiment 2: GFP percentages in CEBPD+MiT, CEBPD+IK6, IK6+MIGR1 and MIGR1+MiT CD34+ transduced cells across 9 weeks in suspension culture. D. Experiment 2: GFP percentages in CEBPD+MiT, CEBPD+IK6, IK6+MIGR1 and MIGR1+MiT CD34+ transduced cells across 9 weeks on MS-5 stromal co-culture. E. Experiment 3: GFP percentages in CEBPD and MIGR1 CD34+ transduced cells across 7 weeks in suspension culture. F. Experiment 3: GFP percentages in CEBPD and MIGR1 CD34+ transduced cells across 7 weeks on MS-5 Stroma co-culture. G. Experiment 4: GFP percentages in CEBPD and MIGR1 CD34+ transduced cells across 5 weeks in suspension culture. H. Experiment 4: GFP percentages in CEBPD and MIGR1 CD34+ transduced cells across 5 weeks on MS-5 Stroma co-culture.

4.3.5.3 IK6 Expressing CD34+ cell populations did not show a clonal advantage alone or in combination with CEBPD

Presenting Experiment 2 as an example, expression of IK6 did not influence cell populations. Initial integration rates of cells transduced with the Thy1 vector were all above 50% (Figure 4.13). As with GFP expression, Thy1 expression was found to be higher in stromal cultures (GM = 40%) than liquid cultures (GM = 18%). Also, expression in suspension cultures showed a more severe decline (GM = 71% - 6%) in comparison to stromal cultures (GM = 71% - 21%) over 6 weeks (Supplementary Figure 7.3). As with GFP alone and Thy1 alone GFP+Thy1 double positive populations were higher in stromal cultures, no other significant trends were observed.

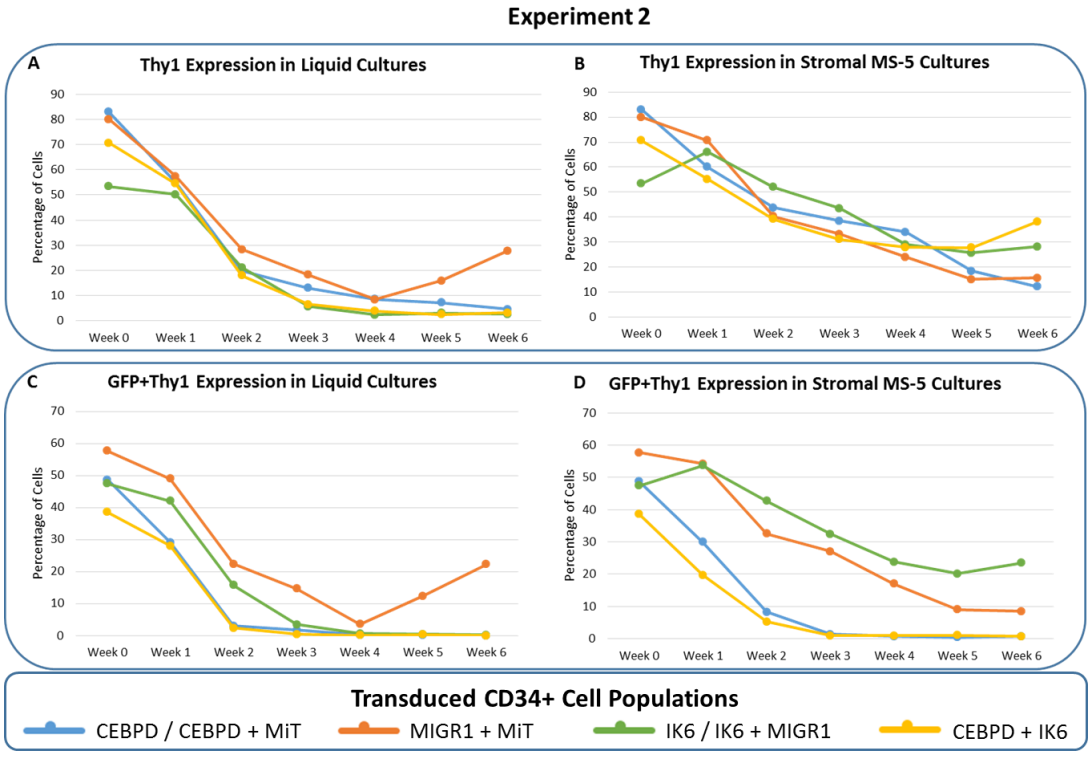


Figure 4.13 Transduction levels of Thy1 transduced and Thy1+GFP transduced in CD34+ cells in Experiment 2 in two culture conditions.

4.3.5.4 CEBPD expression shows consistent growth inhibition

Cell counts were performed weekly to track the proliferative effects of the retroviral particles. Cumulative counts are displayed for both suspension and

stromal cultures across both experiments (Figure 4.14). In both Experiment 1 and 2, little difference was observed in proliferation rates of CD34+ cells in suspension, with the exception of a decline in IK6 transduced cells in Experiment 1 (Figure 4.14 A), a trend not replicated in Experiment 2 (Figure 4.14 C). Growth on stroma showed greater variation. In Experiment 1, proliferation rates clearly divided the transduced populations with the CEBPD cells showing least growth; a trend replicated in Experiment 2 (Figure 4.14B&D). Higher proliferation rates were observed in suspension cultures than stromal cultures, which was expected due to CD34+ cell migration under the MS-5 stroma.

CEBPD proliferation also declined in Experiments 3 and 4, with MIGR1 populations showing higher growth across all experiments and culture conditions (Figure 4.14E-H), although a less pronounced difference was observed in Experiment 4 (Figure 4.14G&H). Expression of CEBPD was shown to consistently decrease cell proliferation, with the effects being more pronounced in stromal cultures.

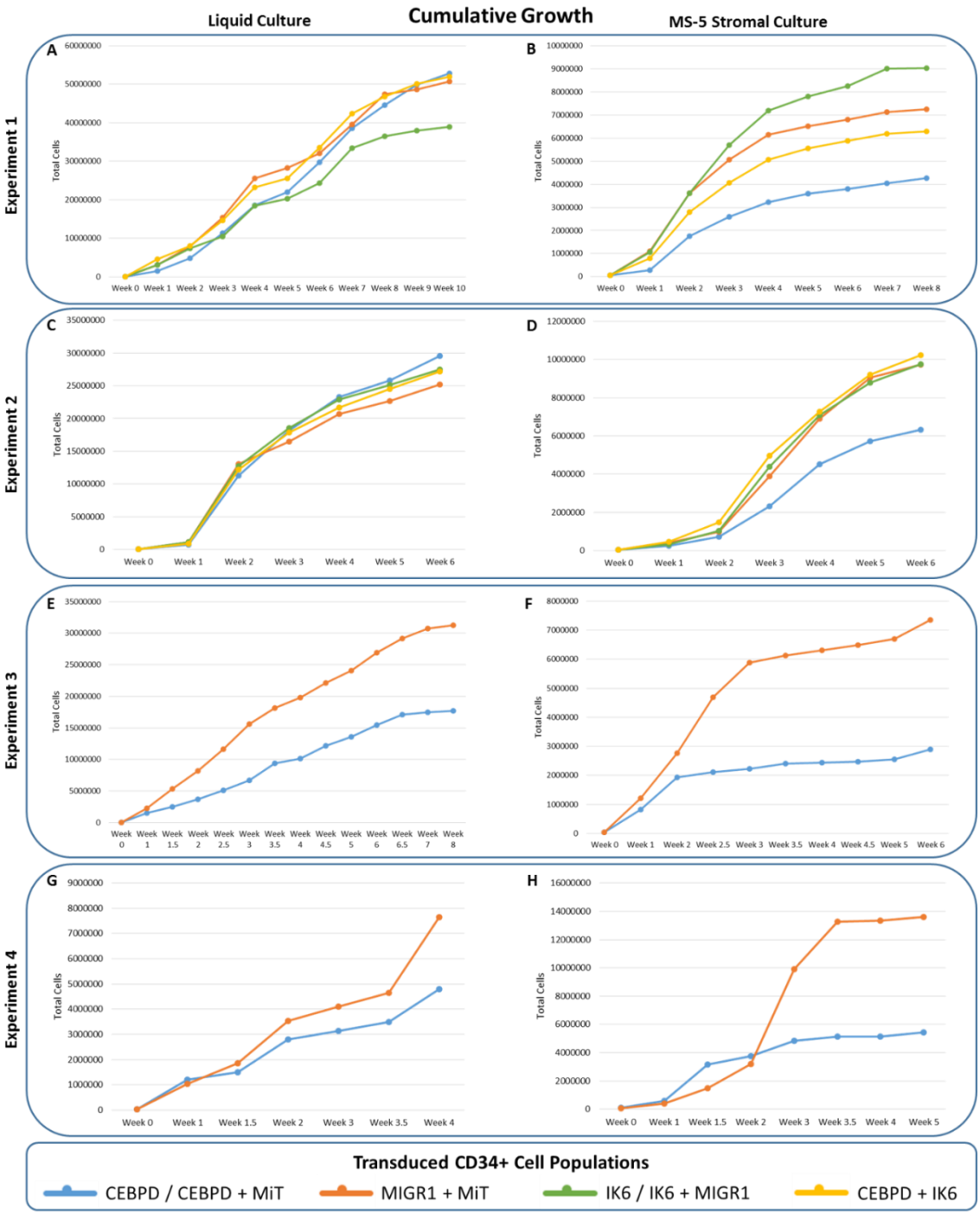


Figure 4.14. Cumulative growth of CD34+ cells across two experiments in two different in vitro culture settings.

A. Experiment 1: growth of transduced CD34+ cells in suspension culture. B. Experiment 1: growth of transduced CD34+ cells on MS-5 stromal co-culture. C. Experiment 2: growth of transduced CD34+ cells in suspension culture. D. Experiment 2: growth of transduced CD34+ cells on MS-5 stromal co-culture. E. Experiment 3: growth of transduced CD34+ cells in suspension culture. F. Experiment 3: growth of transduced CD34+ cells on MS-5 stromal co-culture. G. Experiment 4: growth of transduced CD34+ cells in suspension culture. H. Experiment 4: growth of transduced CD34+ cells on MS-5 stromal co-culture.

4.3.6 Cell Surface Analysis

Cell surface markers were chosen to investigate the consequence of retroviral expression on lineage commitment. The markers consisted of the early haematopoiesis marker CD34, myeloid markers CD11b, CD13, CD14, CD16 and CD33, and lymphoid markers CD10 and CD19.

4.3.6.1 Initial CD34 expression is high declining over time, indicating loss of haematopoietic progenitors

Two main trends were observed in CD34 marker expression, varying between experiments but not culture conditions. One trend was observed in Experiment 1 (Figure 4.15A&B), with alternating waves of high and low expression for the duration of the experiment. The alternating high / low expression was more uniform in the suspension culture (Figure 4.15 A), with all transduced cells showing similar levels of expression. Experiment 1 stromal culture displayed more variation among the cell populations with the IK6 cells showing a peak of expression at week 5, while the CEBPD population began to decline in week 4 (Figure 4.15B). A different trend was observed in Experiments 2-4, where a high starting expression dropped rapidly over the duration of all experiments in both suspension and stromal cultures (Figure 4.15 C-H), with little heterogeneity across populations. Overall CD34 expression showed more variation between Experiment 1 and Experiments 2-4 suggesting one reason for differential expression of the CD34 marker could be due to cord blood heterogeneity.

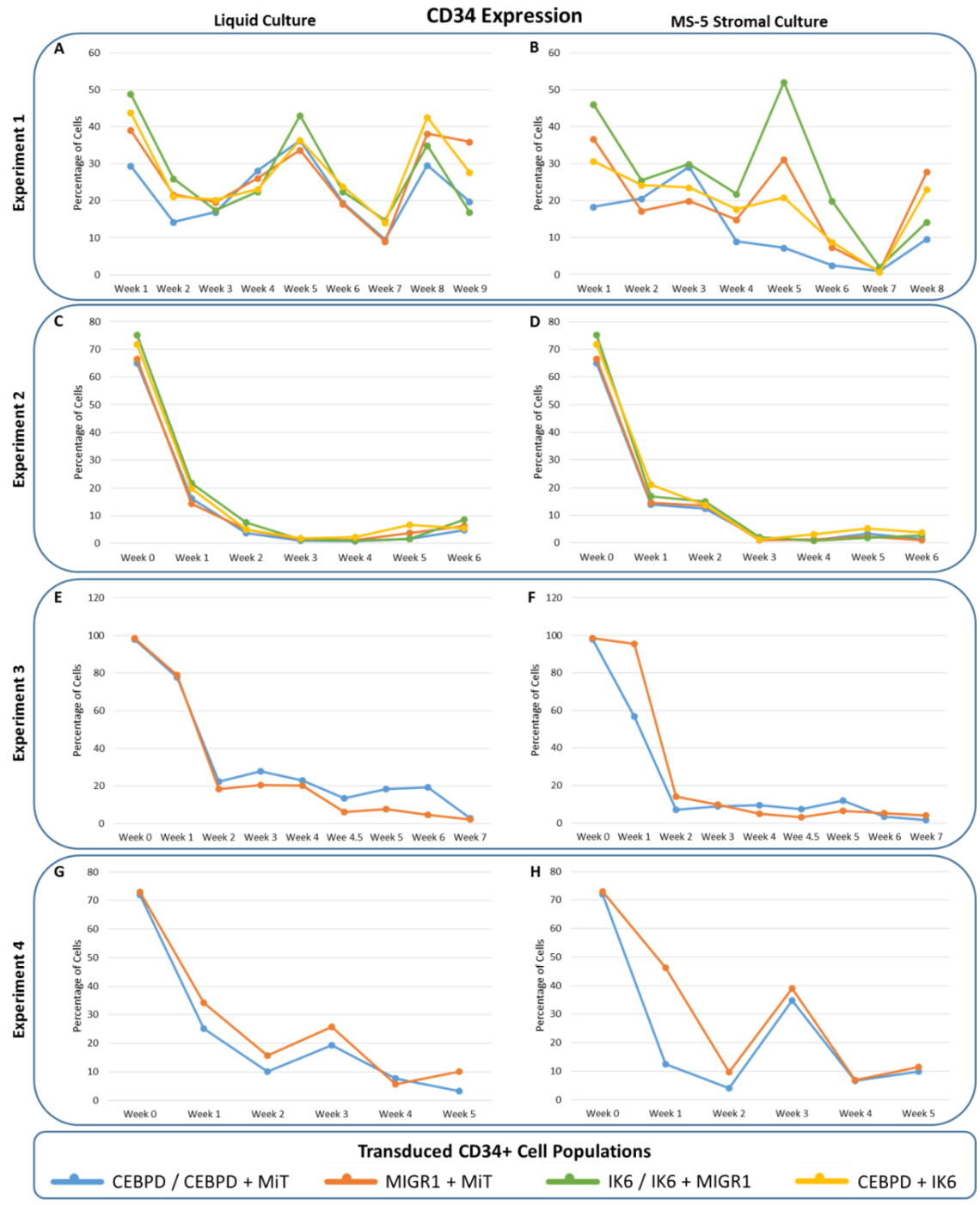


Figure 4.15. Cell surface expression of CD34 in Experiment 1 transduced populations CEBPD + IK6, CEBPD, IK6, MIGR1 + MiT and Experiment 2 transduced populations CEBPD + IK6, CEBPD + MiT, IK6 + MIGR1 and MIGR1 + MiT, and Experiment 3 and 4 transduced populations CEBPD and MIGR1.

A. Experiment 1 CD34 marker expression over 9 weeks in suspension culture. B. Experiment 1 CD34 marker expression over 8 weeks in MS-5 stromal co-culture. C. Experiment 2 CD34 marker expression over 6 weeks in suspension culture. Experiment 2 CD34 marker expression over 6 weeks in MS-5 stromal co-culture. E. Experiment 3 CD34 marker expression over 7 weeks in suspension culture. F. Experiment 3 CD34 marker expression over 7 weeks in MS-5 stromal co-culture. G. Experiment 4 CD34 marker expression over 5 weeks in suspension culture. H. Experiment 4 CD34 marker expression over 5 weeks in MS-5 stromal co-culture.

4.3.6.2 CD33 expression is consistently high throughout all experiments, indicating a strong myeloid bias in culture

Expression patterns of CD33 markers displayed variation between all experiments and cultures, particularly liquid and stromal cultures. Although overall expression was high, it was not surprising, due to the presence of myeloid biasing cytokines in the culture medium (Section 2.2.4.3, Section 2.3.6). (Figure 4.16). Experiment 1 showed low expression at the start. It should be noted that the first two weeks of CD33 analysis in Experiment 1 was analysed using the Pacific Blue fluorochrome, which had been shown to have lower expression by the Mulloy lab. As a result, the CD33 marker was changed to the APC fluorochrome at week 3, resulting in higher expression, which remained unchanged. As a result Experiment 1 weeks 1 and 2 in both liquid and stromal cultures were omitted from this work.

Expression of CD33 was consistently high across all liquid culture experiments, a decline in CD33 expression in stromal cultures was expected as the culture medium of these experiments was designed for lymphoid expansion. In addition to cytokines, contact with the stromal feeder layer and additional secretion of the IL-7 cytokine into the medium by stromal cells would have provided additional cues for lymphoid differentiation, lowering the level of the myeloid cell population.

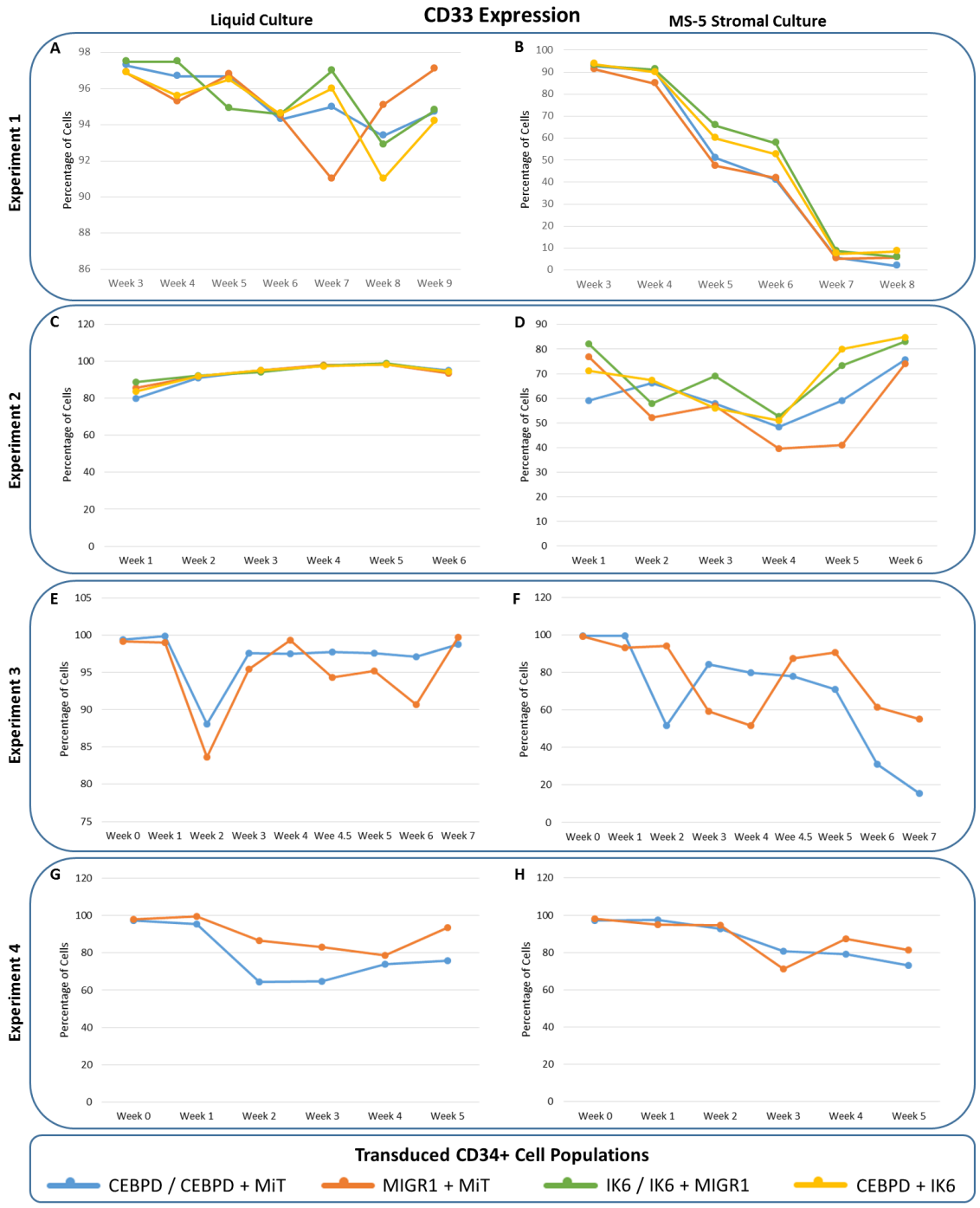


Figure 4.16. Cell surface expression of CD33 in Experiment 1 transduced populations CEBPD + IK6, CEBPD, IK6, MIGR1 + MiT and Experiment 2 transduced populations CEBPD + IK6, CEBPD + MiT, IK6 + MIGR1 and MIGR1 + MiT.

A. Experiment 1 CD33 marker expression between weeks 3- 9 in suspension culture. Weeks 1-2 were omitted due to differing FACS fluorophores giving greatly contrasting marker expression. B. Experiment 1 CD33 marker expression between weeks 3-8 in MS-5 stromal co-culture. Weeks 1-2 were omitted due to differing FACS fluorophores giving greatly contrasting marker expression. C. Experiment 2 CD33 marker expression over 6 weeks in suspension culture. Experiment 2 CD33 marker expression over 6 weeks in MS-5 stromal co-culture. E. Experiment 3 CD33 marker expression over 7 weeks in suspension culture. F. Experiment 3 CD33 marker expression over 7 weeks in MS-5 stromal co-culture. G. Experiment 4 CD33 marker expression over 5 weeks in suspension culture. H. Experiment 4 CD33 marker expression over 5 weeks in MS-5 stromal co-culture.

4.3.6.3 CD19 expression was observed in CEBPD expressing cells

Expression of CD19 was generally low across all experiments (Figure 4.17). However, suspension cultures displayed very low CD19 expression, apart from the empty vector controls, MIGR1+MiT, which showed small peaks at several time points, determined to be nonspecific binding of the CD19 antibody (Figure 4.17A&C). Stromal cultures showed higher levels of CD19 expression, with Experiments 1, 2 and 4, all showing increasing levels up to the final week of the experiment (Figure 4.17B, D&H). In Experiment 1, stromal cultures displayed small increases of CD19 from week 6 to week 8, with one large increase of the marker at weeks 4-6 observed in the CEBPD population (Figure 4.17B). This population will be discussed further below.

As expected, overall expression of CD19 was low in the myeloid biasing liquid culture conditions and slightly higher in the stromal cultures, where both IL-7 secreted by the stromal cells and supplemented IL-7 were present.

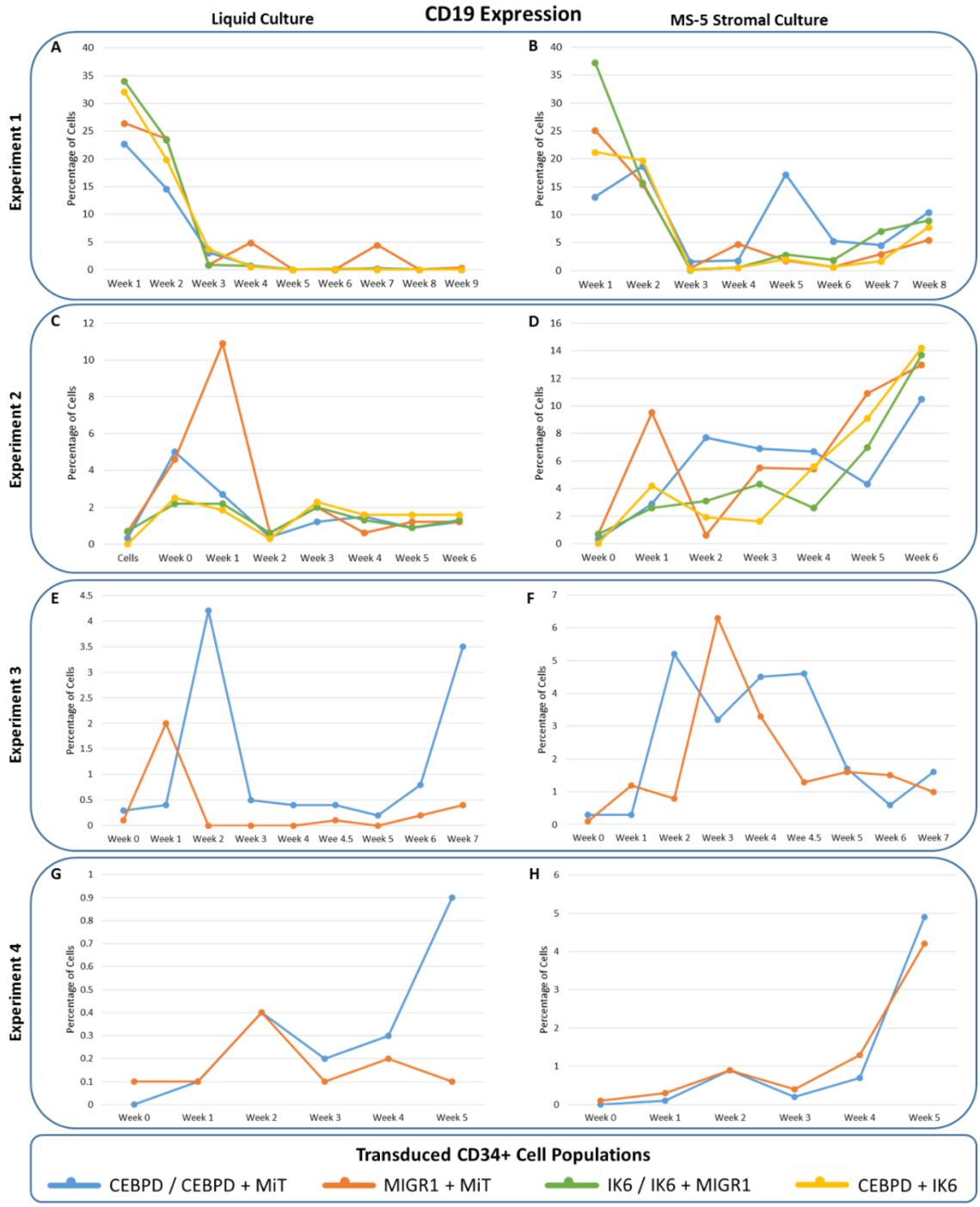


Figure 4.17. Cell surface expression of CD19 in Experiment 1 transduced populations CEBPD + IK6, CEBPD, IK6, MIGR1 + MiT and Experiment 2 transduced populations CEBPD + IK6, CEBPD + MiT, IK6 + MIGR1 and MIGR1 + MiT.

A. Experiment 1 CD19 marker expression over 9 weeks in suspension culture. B. Experiment 1 CD19 marker expression over 8 weeks in MS-5 stromal culture. C. Experiment 2 CD19 marker expression over 6 weeks in suspension culture. Experiment 2 CD19 marker expression over 6 weeks in MS-5 stromal culture. E. Experiment 3 CD19 marker expression over 7 weeks in suspension culture. F. Experiment 3 CD19 marker expression over 7 weeks in MS-5 stromal co-culture. G. Experiment 4 CD19 marker expression over 5 weeks in suspension culture. H. Experiment 4 CD19 marker expression over 5 weeks in MS-5 stromal co-culture.

In Experiment 1 stromal culture, expression of CD19 between weeks 3-6 (Figure 4.18) was further analysed, as these cells uniquely showed co-expression of GFP. A clear CD19 positive population developed at week 3 (1.2%), increasing by week 4 (2.1%) and 5 (15.4%), and decreasing at week 6 (7.3%) (Figure 4.18A). Expression of the CD19 marker against GFP displayed separate populations with expression of either GFP or CD19 at week 3. By week 4, a small but distinct population appeared which was positive for both CD19 and GFP (0.6%). This population was significantly increased by week 5 (10.7%), which began to decrease at week 6 (3.6%). By week 7, the GFP positive population had disappeared (Figure 4.18B). This population was unique as it was the only *in vitro* GFP, CD19 double positive population observed in this study. The short life span of the transduced population was in keeping with previously reported physiological effects of *CEBPD* expression due to cell cycle arrest (Gery *et al.*, 2005).

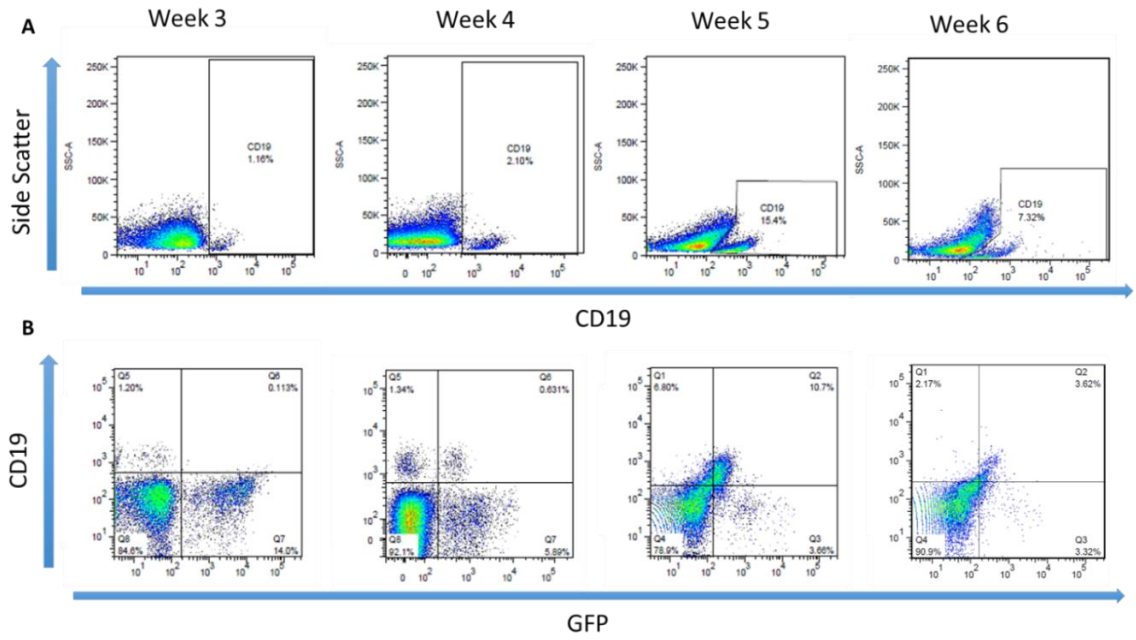


Figure 4.18. FACS plots of Experiment 1 CEBPD transduced CD34+ cells over a four week period. A. Expression of CD19 marker against side scatter in CEBPD transduced CD34+ cells. B. Expression of GFP marker against CD19 marker in CEBPD transduced cells.

4.3.6.4 Other flow cytometry markers

Myeloid markers CD11b, CD13, CD14, and CD16, and lymphoid marker CD10 were also tracked during the course of Experiments 1 and 2, however no trends were observed and these data are not reported in detail.

4.4 Retroviral Transduction of CD34+ Cells *in vivo* Assays

Xenograft experiments were set up alongside the *in vitro* studies described above (Table 4.2). After injection with transduced CD34+ cells, tail bleeds and bone marrow aspirates were performed on recipient mice every two weeks to assess engraftment status.

Experiment No.	Transduced CD34+ Cell Populations	No. of Mice per Population	Experiment Ended
1	CEBPD	3	Week 20
	IK6	0	
	CEBPD + IK6	0	
	MIGR1 + MiT	3	
2	CEBPD + MiT	5	Week 20
	IK6 + MIGR1	5	
	CEBPD + IK6	5	
	MIGR1 + MiT	5	
3	CEBPD	3	Week 22 (n=1)
	MIGR1	2	NA
4	CEBPD	3	NA
	MIGR1	3	

Table 4.2. Xenograft experiments performed for this study.

Experiment 1 was ended at week 20 after bone marrow aspirates showed complete loss of the GFP+CD19+ population, data discussed below. Experiment 2 was ended at week 20 where all mice were sacrificed to harvest GFP+CD19+ cell populations observed via flow cytometry. Experiment 3 is continuing, one MIGR1 mouse died as a result of Busulfan toxicity in the first week of the experiment. One mouse, containing CEBPD CD34+ transduced cells, was sacrificed and cells harvested, after displaying weight loss as a result of engraftment. All other animals in the experiment remain alive and have shown no engraftment via flow cytometry. Experiment 4 is continuing, no mice have shown engraftment to date.

4.4.1 Experiment 1

4.4.1.1 All CEBPD and some MIGR1 mice showed human CD34+ cell engraftment

Cells were taken at weeks 8 (PB), 15 (PB) and 18 (bone marrow, BM) and analysed using flow cytometry. Little/no engraftment was observed in week 8 (data not shown). Flow cytometry at week 15 displayed both murine CD45 (mCD45) and human CD45 (hCD45) positive populations in all CEBPD mice, and one MIGR1 mouse (Figure 4.19A). Following this observation, mice were maintained for a further three weeks prior to culling to allow these successfully engrafted populations to expand. Mice were sacrificed at week 20 of Experiment 1, spleens and BM were collected (Section 2.3.5.4) and tested for hCD45

expression. Engraftment was observed primarily in bone marrow (Figure 4.19 C), with some expression in splenic tissue (Figure 4.19B).

Human CD45+ expression decreased in two CEBPD mice from week 15 to week 18. Two mice, one CEBPD and one MIGR1, exhibited an increase of hCD45+ cells between weeks 15 and 18. The remaining MIRG1 mouse displayed a small number of hCD45+ cells in the bone marrow, suggesting that human cells were beginning to engraft (Figure 4.19). Overall, transduced CEBPD CD34+ cells appeared to engraft more efficiently than the empty vector MIGR1 controls.

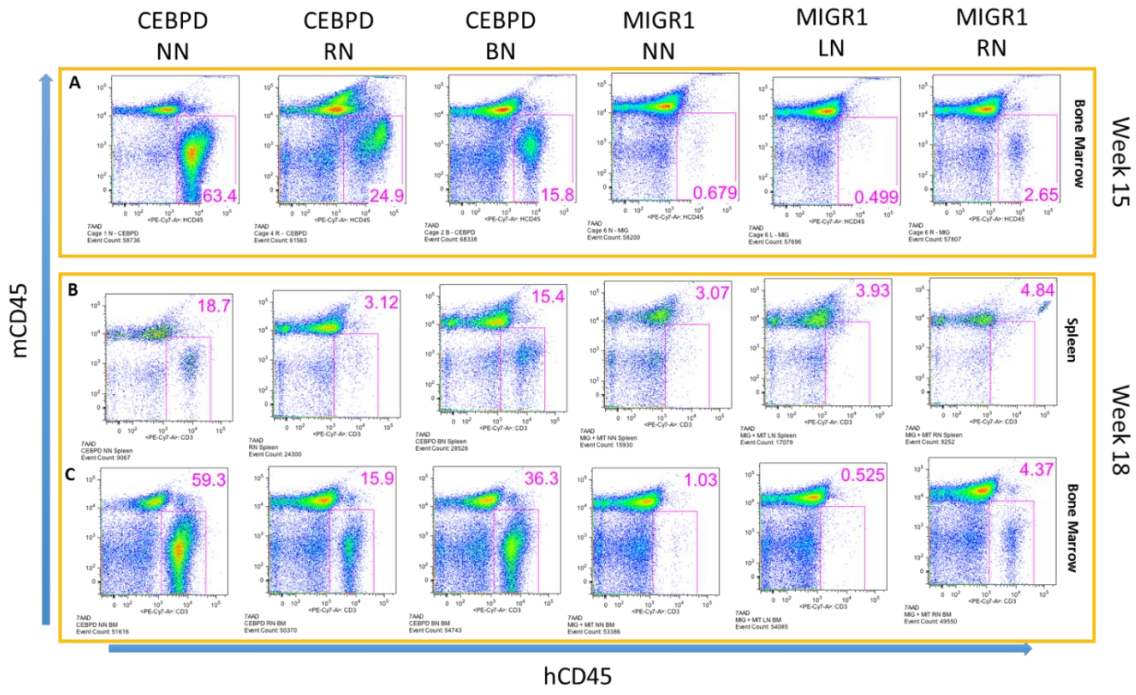


Figure 4.19. FACS plots of NSG mice intrafemorally injected with CEBPD transduced CD34+ cells (CEBPD NN, CEBPD RN and CEBPD BN) and mice intrafemorally injected with MIGR1 transduced CD34+ cells (MIGR1 NN, MIGR1 LN, and MIGR1 RN).

A. FACS plot analysing hCD45 expression against mCD45 expression of NSG mice bone marrow aspirates taken at week 15 post intrafemoral injection. FACS plot analysing hCD45 expression against mCD45 expression of NSG mice spleens and bone marrow 18 weeks post intrafemoral injection.

4.4.1.2 Transduced cells disappeared between weeks 15 and 18.

At week 15, small amounts of nonspecific GFP expression was observed in all mice, with the exception of CEBPD mouse RN, which displayed two distinct GFP+

populations with different intensities (Figure 4.20A). However, at week 18 all GFP+ cells had disappeared (Figure 4.20B).

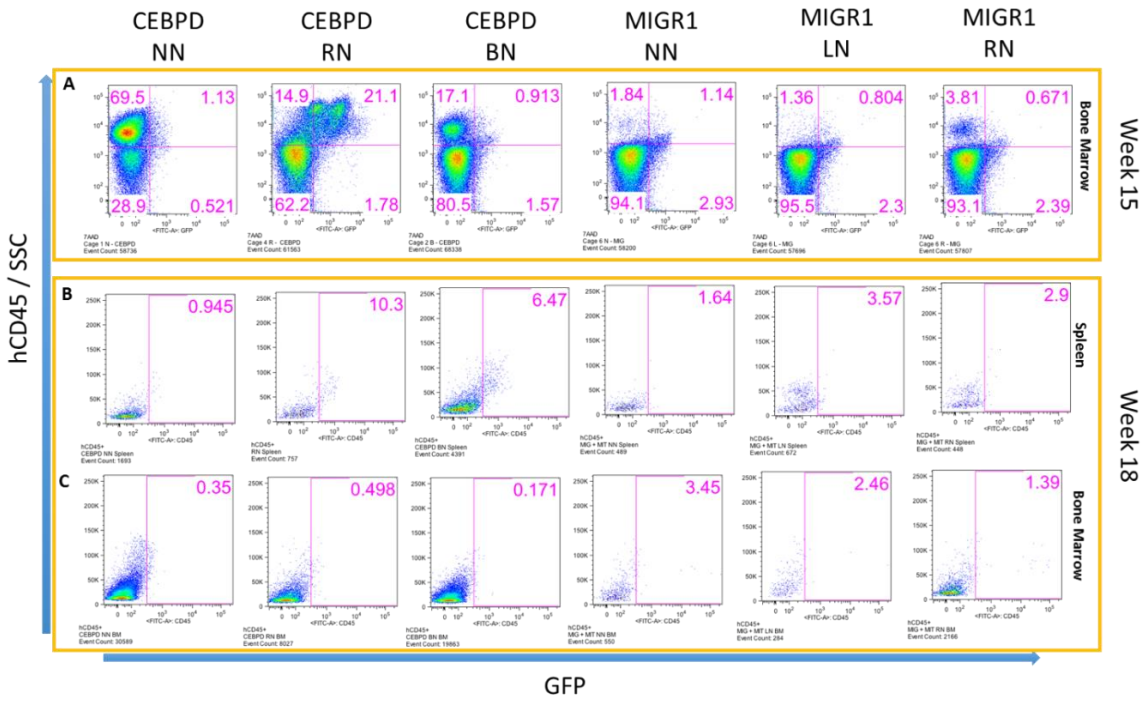


Figure 4.20. FACS plots of NSG mice intrafemorally injected with CEBPD transduced CD34+ cells (CEBPD NN, CEBPD RN and CEBPD BN) and mice intrafemorally injected with MIGR1 transduced CD34+ cells (MIGR1 NN, MIGR1 LN, and MIGR1 RN).

A. FACS plot analysing GFP expression against hCD45 expression of NSG mice bone marrow aspirates taken at week 15 post intrafemoral injection. FACS plot analysing GFP expression against hCD45 expression of NSG mice spleens and bone marrow 18 weeks post intrafemoral injection.

4.4.1.3 CD19 expression increased between weeks 15 and 18.

Mice with CEBPD transduced cells displayed varying levels of CD19 expression at week 15, mouse NN exhibited high CD19 expression (40%), while the remaining two mice both displayed low levels of CD19 positivity (5.5% and 6.8%) (Figure 4.21A). MIGR1 mice NN and LN showed some nonspecific expression of the marker, while mouse RN exhibited a clear CD19+ population (1.6%) (Figure 4.21A). At week 18, expression of this marker was more pronounced in all murine bone marrow and to a lesser extent in spleens, with CEBPD mice all displaying expression levels at ~60%, as did MIGR1 mouse RN.

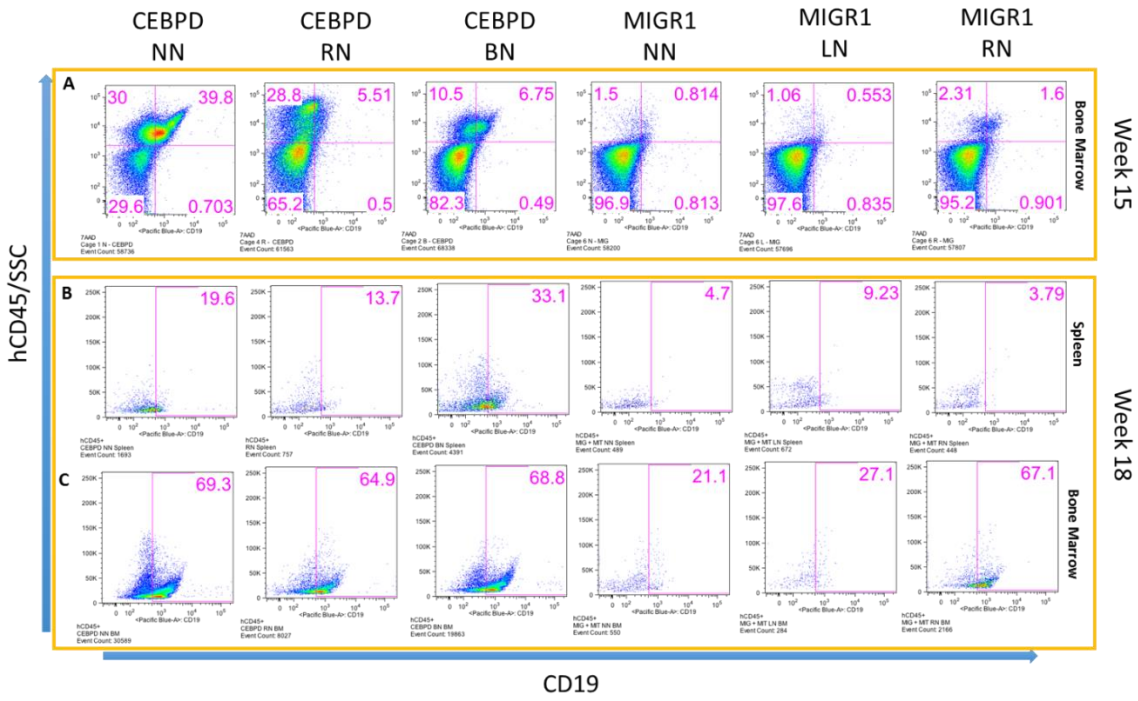


Figure 4.21. FACS plots of NSG mice intrafemorally injected with CEBPD transduced CD34+ cells (CEBPD NN, CEBPD RN and CEBPD BN) and mice intrafemorally injected with MIGR1 transduced CD34+ cells (MIGR1 NN, MIGR1 LN, and MIGR1 RN).

A. FACS plot analysing CD19 expression against hCD45 expression of NSG mice bone marrow aspirates taken at week 15 post intrafemoral injection. FACS plot analysing CD19 expression against hCD45 expression of NSG mice spleens and bone marrow 18 weeks post intrafemoral injection.

Cell selection for Experiment 1 was not performed as all CD19+ cells disappeared between weeks 15 and 18.

4.4.2 Experiment 2

4.4.2.1 Transduced cells were found primarily in MIGR1+MiT mice

Mice displayed good engraftment of human cells with expression of hCD45 observed in all mice (Figure 4.22). CEBPD+IK6 mice showed the highest average hCD45 expression (GM = 25.7%), followed by MIGR1 + IK6 (GM = 20.1%), CEBPD+MiT (GM = 19.7%) and MIGR1+MiT (GM = 15.5%).

GFP expression was low in most of the mice at week 18, despite high expression of hCD45 (Figure 4.22). Eight mice showed GFP expression, including five MIGR1+MiT empty vector control, two CEBPD+MiT and one CEBPD+IK6 mice. Cells from mice 1N, 2N, 3N, 1R and 1B were selected for cell sorting (Figure

4.22) (Section 4.5.1). The results suggested that, as observed for the *in vitro* culture and Experiment 1 murine data, GFP expression was short lived.

4.4.2.2 Cells selected for sorting showed high CD19 expression and low CD13&CD33 expression

DAPI negative, hCD45 positive and GFP positive cells from the five mice mentioned above were examined for CD19 and CD13+CD33 expression, to determine the lineage of the cells expressing the CEBPD and MIGR1 vectors. No lineage difference was observed between CEBPD and MIGR1 transduced cells. All mice expressed high levels of CD19 and low levels of CD13+CD33. CD13+CD33 expression was higher in mouse 1B (27%), however cell numbers were low for all FACS plots (Figure 4.23).

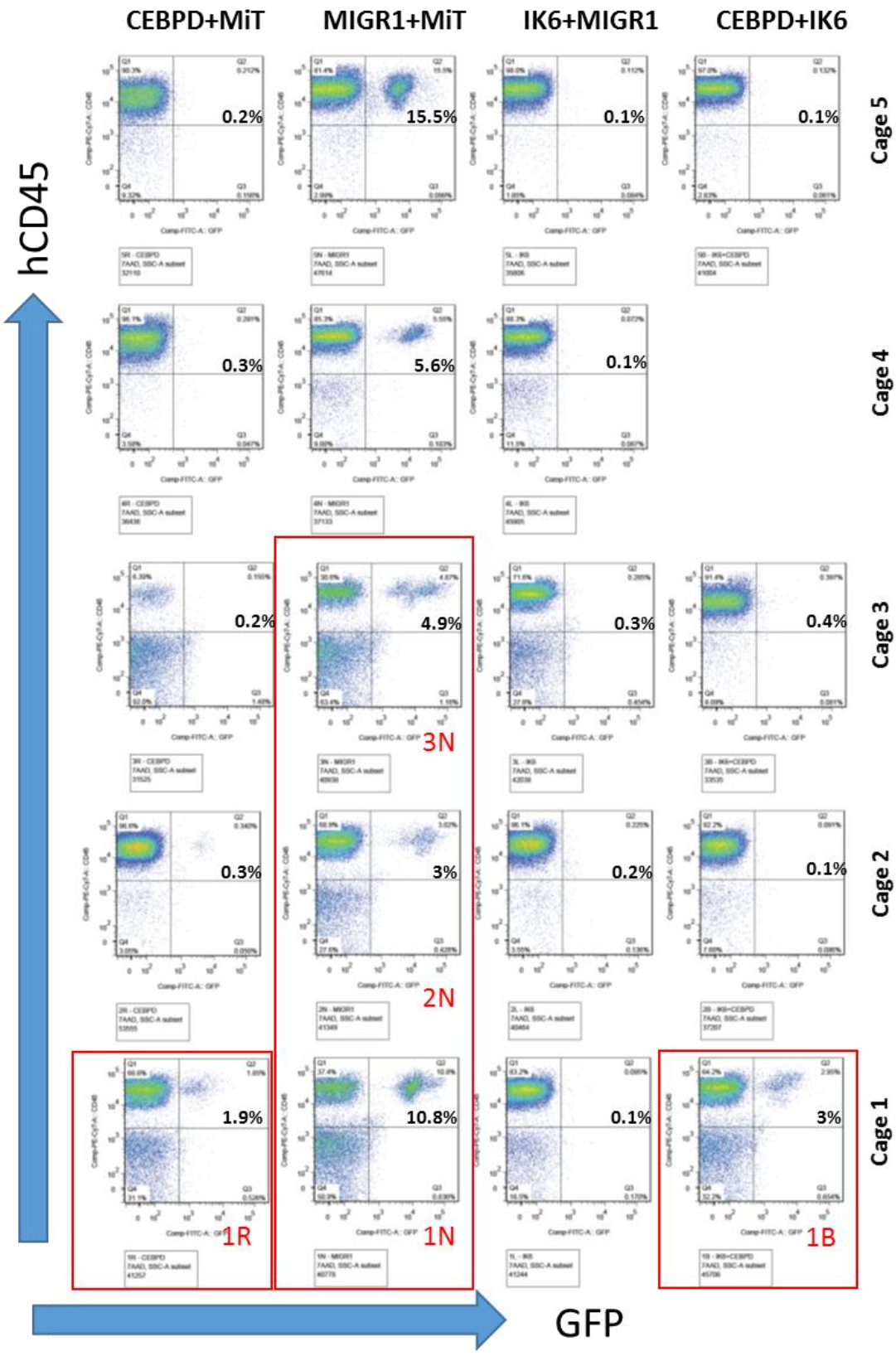


Figure 4.22. FACS plot GFP vs hCD45 from NSG mice intrafemorally injected with transduced CD34+ cell populations (CEBPD+MiT, MIGR1+MiT, IK6+MIGR1 and CEBPD+IK6), week 18. Plots highlighted by red were selected for viable cell sorting two weeks later, mouse ID in red.

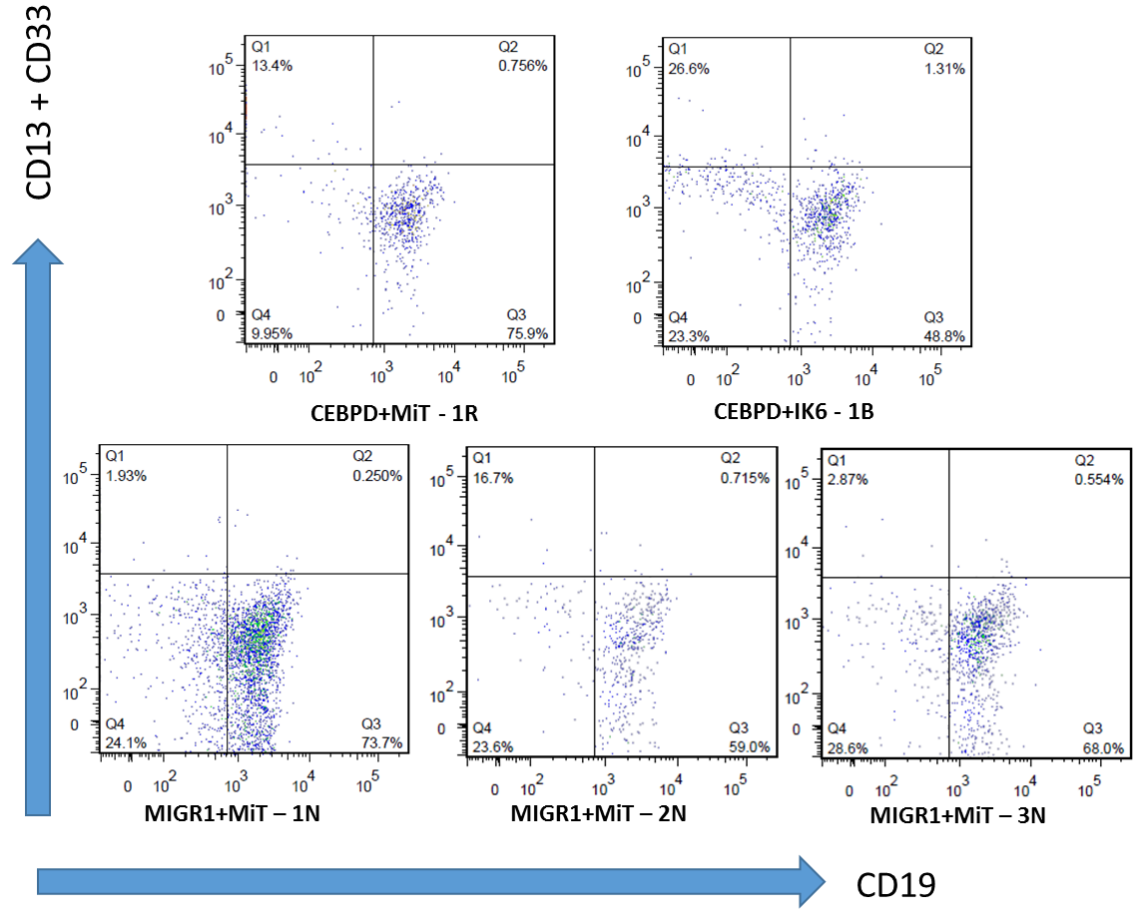


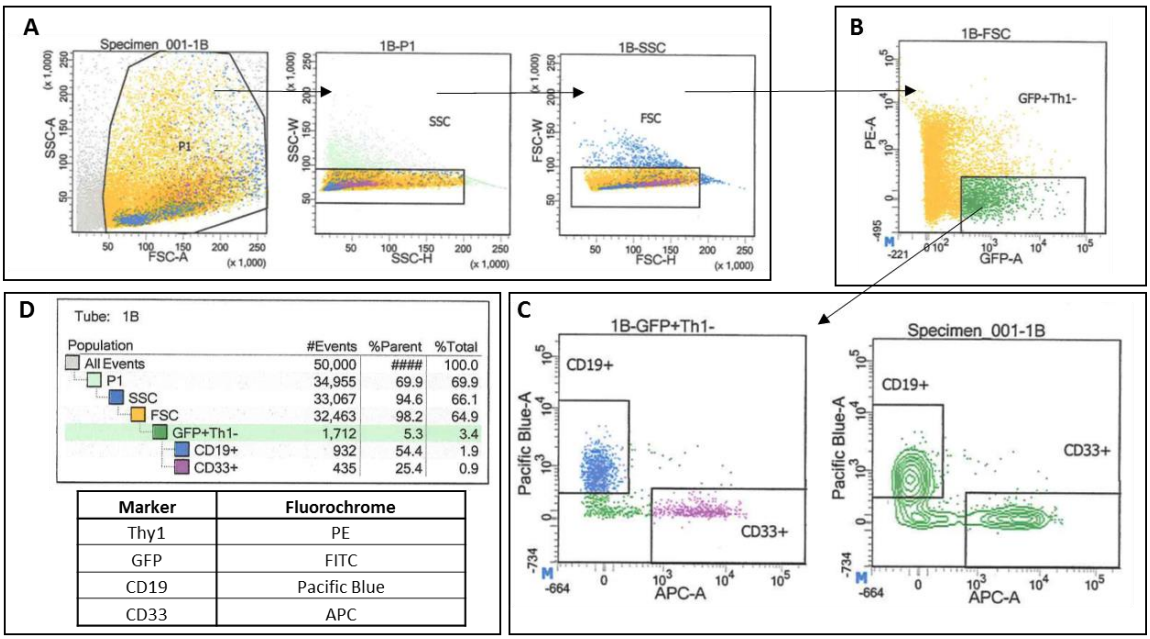
Figure 4.23. FACS plot of CD19 vs CD33 from BM cells gated for DAPI negativity, hCD45 positivity, and GFP positivity, plots from mice selected at Cincinnati for viable cell sorting, week 18. All mice display similar expression patterns of markers, with strong expression of CD19 and very little CD33 positivity observed.

4.5 *In vivo* and *in vitro* Cell Sorting

4.5.1 Experiment 2: in vivo Cell Sorting

FACS analysis and viable cell sorting was kindly performed by Mark Wunderlich in Cincinnati Children's Hospital, as I had returned to Newcastle prior to murine engraftment. NSG mice were sacrificed at week 18, bone marrow was collected (Section 4.4.2.2) and cells were sorted using the BD FACSAria II machine. The first stage involved gating of appropriately sized cells, using forward and side cell scatter to exclude cell doublets and debris (Figure 4.24 A). These selected cells were gated for GFP positivity and Thy1 negativity, isolating only human transduced cells (Figure 4.24 B). Finally, the GFP+ Thy1- cells were separated into two fractions, CD19+ cells and CD33+ cells, which were resuspended in RLT buffer and stored at -20^{o}C (Figure 4.22 C). Final CD19+ and CD33+ fractions comprised a small percentage of the total sorted population (Figure 4.24D). This

process was performed on five mice, 1 CEBPD (mouse 1R), 1 CEBPD + IK6 (mouse 1B), and 3 MIGR1 (mice 1N, 2N and 3N), which were sufficiently well engrafted for a successful cell sort. Sorted cell numbers showed variation between mice (Table 4.3). Unfortunately the RNA harvested from the sorted cells was not of a sufficient quantity for downstream applications.



Marker	Fluorochrome
Thy1	PE
GFP	FITC
CD19	Pacific Blue
CD33	APC

Figure 4.24. Cell sorting protocol performed on NSG bone marrow samples with fluorochromes used for the process, image from sample 1B. Cell sorting images provided by Mark Wunderlich.

A. First selection gate isolating cells by side scatter and forward scatter. B. Second selection gate isolating cells for GFP positivity and Thy1 negativity. Third selection gates isolating cells into two populations, CD19 positivity and CD33 positivity. D. Table displaying all populations in the FACS process. Cells were sorted using the BD FACSAriaII machine.

Mouse	Transduced Cells	Lineage	Cell Count
1R	CEBPD	CD19	25,000
1R	CEBPD	CD33	19,000
1B	CEBPD + IK6	CD19	1,100,000
1B	CEBPD + IK6	CD33	360,000
1N	MIGR1	CD19	2,100,000
1N	MIGR1	CD33	350,000
2N	MIGR1	CD19	160,000
2N	MIGR1	CD33	200,000
3N	MIGR1	CD19	2,000,000
3N	MIGR1	CD33	530,000

Table 4.3. Flow sorted cells harvested from NSG mouse bone marrow with mouse identification, transduced cells, lineage of sorted cells and final cell count after sort.

4.5.2 Experiment 3 and 4: in vitro Cell Sorting

The original aim of this study was to create and isolate CD19+ GFP+ cells. This population was observed in Experiment 1 *in vitro*, but declined before cells could be sorted (Section 4.3.6.3). The same phenomenon was also observed in Experiment 1 mice, with the CD19 population disappearing between weeks 15 and 18 (Figure 4.21). The CD19+ GFP+ population was not observed in Experiment 2 *in vitro* cultures. However these cultures showed strong engraftment and thus were FACS sorted (see below). When the work was repeated in Newcastle, the criteria for sorting of CD19+ GFP+ cells was changed as no population expressing the CD19 marker appeared. As a result, only GFP+ CD33+ cells were isolated for extraction. The first selection gate was by size, using forward scatter (FSC) vs side scatter (SSC) (Figure 4.25 A), followed by hCD45 marker expression (Figure 4.25B), with positive cells being gated by GFP expression (Figure 4.25C). GFP negative cells were collected for RNA and protein extraction, GFP positive cells were taken forward for the final selection step, which analysed CD19 and CD33 expression. As there were no CD19 + cells observed, the cells were isolated into CD33 + and CD33 - populations (Figure 4.25D).

A small side population was observed in the hCD45+ selection gate (P6) (Figure 4.25 B), these cells were selected and also sorted by GFP, CD33 and CD19 expression and combined with the existing CD33+ and CD19+ populations. Cells

isolated from Experiment 4 were sorted using the same selection criteria (Table 4.5). Unfortunately isolation of RNA from the sorted cells of both experiment was not of a high enough quality and quantity for downstream analysis.

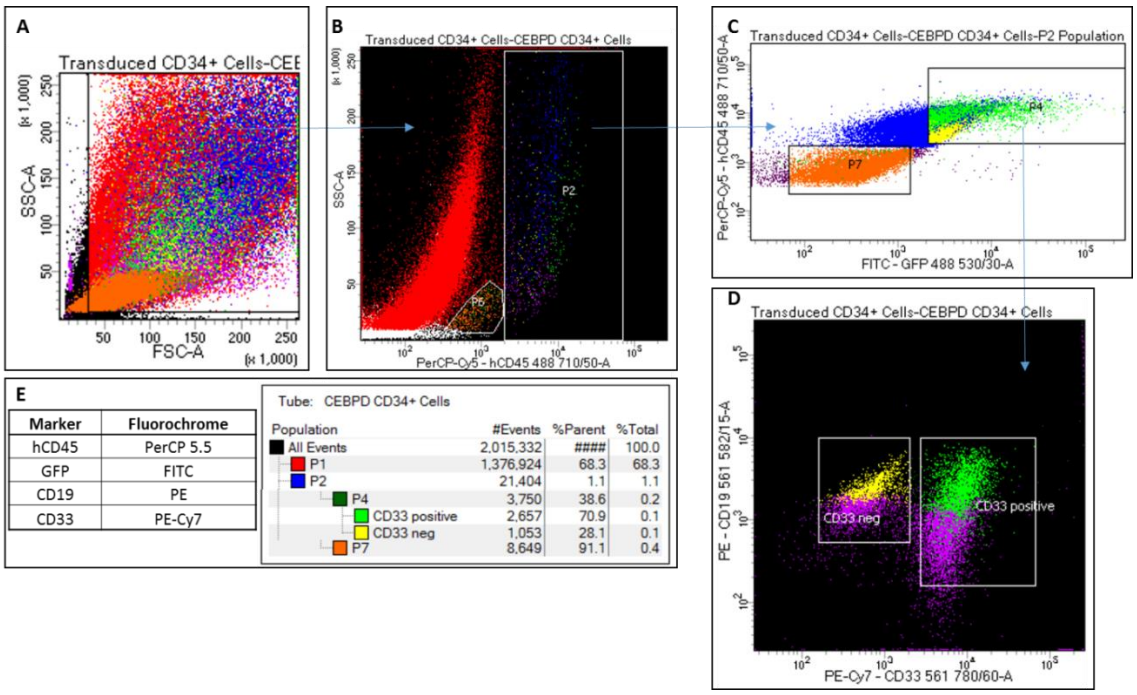


Figure 4.25. Cell sorting of CEBPD transduced CD34+ cells cultured on MS-5 stroma at week 8.

A. First selection gate by FSC and SSC (P1). B. Second selection gate collecting hCD45+ cells (P2). C. Third and fourth selection gates, GFP+ cells selected for further sorting (P4) and GFP- cells sorted and stored for RNA and protein extraction (P7). Final selection gates separating cells by CD19 and CD33 expression, as there was no CD19+ cells populations were divided into CD33+ and CD33- populations for RNA and protein extraction. E. Modified table displaying selection gates of the flow sort and FACS markers and respective fluorochromes.

Transduced Cells	Lineage	Cell Count
CEBPD	GFP-	5973
CEBPD	GFP+CD33+	8720
CEBPD	GFP+ CD33-	3058
MIGR1	GFP-	2908
MIGR1	GFP+CD33+	6292
MIGR1	GFP+ CD33-	2033

Table 4.4. Experiment 3 viable cell sort from CD34+ cells transduced with CEBPD retroviral vector.

Transduced Cells	Lineage	Cell Count
CEBPD	GFP-	11,313
CEBPD	GFP+CD33+	39,301
CEBPD	GFP+ CD33-	5615
MIGR1	GFP-	374
MIGR1	GFP+CD33+	21,756
MIGR1	GFP+ CD33-	2379

Table 4.5. Experiment 4 viable cell sort from CD34+ cells transduced with CEBPD retroviral vector.

4.6 Mouse Xenografts – Skull Sectioning and Immunohistochemistry

An aim of this study was the expansion of viable patient cells in NSG mice, to increase material for study. Initially, three viable patient samples were injected into selected NSG females (Table 4.6). Unfortunately upon sacrificing of mice, no engraftment was observed, with the exception of low levels of hCD45 in mouse 20580 (*IGH-CEBPD*) BN. Splenic cells from this mouse were serially transplanted into a further three NSG mice (Table 4.6). The skulls of the xenograft mice were sent to Dr. Halsey, a collaborator in Glasgow University, investigating the metastatic potential of ALL subgroups.

Patient ID	Source	Mouse ID	Cell Numbers Injected	Date Injected	Date Ended	Material
19734	Primograft	RN	per mouse: ~1.8 x 10^5	13/01/2012	11/10/2012	Spleen, BM
19734	Primograft	LN	per mouse: ~1.8 x 10^5	13/01/2012	18/08/2012	BM
19734	Primograft	BN	per mouse: ~1.8 x 10^5	13/01/2012	13/09/2012	Spleen, BM, R.Kidney, L.Kidney
19794	Primograft	LN	per mouse: ~8.7 x 10^5	13/01/2012	01/02/2012	Spleen, BM
19794	Primograft	RN	per mouse: ~8.7 x 10^5	13/01/2012	22/02/2012	Spleen, BM
19794	Primograft	LN	per mouse: ~8.7 x 10^5	13/01/2012	30/10/2012	Spleen, BM, R.Kidney, L.Kidney, Liver
20580	Primograft	BN*	per mouse: ~7 x 10^5	13/01/2012	15/11/2012	Spleen, BM, R.Kidney, L.Kidney, Liver
20580	Primograft	2LN	per mouse: ~7 x 10^5	13/01/2012	12/10/2012	Spleen, BM, R.Kidney, L.Kidney, Liver
20580	Primograft	NN	per mouse: ~7 x 10^5	13/01/2012	28/11/2012	Spleen, BM
20580	Xenograft	RN	per mouse: ~5x10^5	28/11/2012	08/02/2013	Spleen, BM
20580	Xenograft	LN	per mouse: ~5x10^5	28/11/2012	18/02/2013	Spleen, BM
20580	Xenograft	BN	per mouse: ~5x10^5	28/11/2012	18/02/2013	Spleen, BM

Table 4.6. Patient cells selected for engraftment into NSG mice. Samples denoted as primografts were taken from viable cells. Samples denoted as xenograft were selected from mouse BN which showed potential engraftment in FACS analysis (data not shown). All primary patient samples were from IGH-CEBPD translocated patients.*

Work in Glasgow included paraffin embedding, sectioning and immunohistochemistry staining of received skulls. Invasion of the mouse calveria was observed in 3 of the 4 skulls sent. All infiltrations were in mice engrafted with the same viable sample, originating from patient 20580. Skulls tested negative for human CD19 and human CD45. However, a clearly invasive cellular population was observed (Table 4.7) (Figure 4.26), suggesting that some of the viable *IGH-CEBP* patient cells may have crossed the blood brain barrier in three of the sectioned skulls.

Patient ID	Source	Mouse ID	Calvarial BM	Meninges/CNS
20580	Xenograft	RN	Heavy Engraftment	Rim of meningeal infiltration grade 3-4
20580	Xenograft	LN	Heavy Engraftment	Rim of meningeal infiltration grade 3-4
20580	Xenograft	BN	Heavy Engraftment	Very heavy meningeal infiltration grade 5

Table 4.7. IGH-CEBPD xenograft murine skulls sectioned and analysed in Glasgow University.

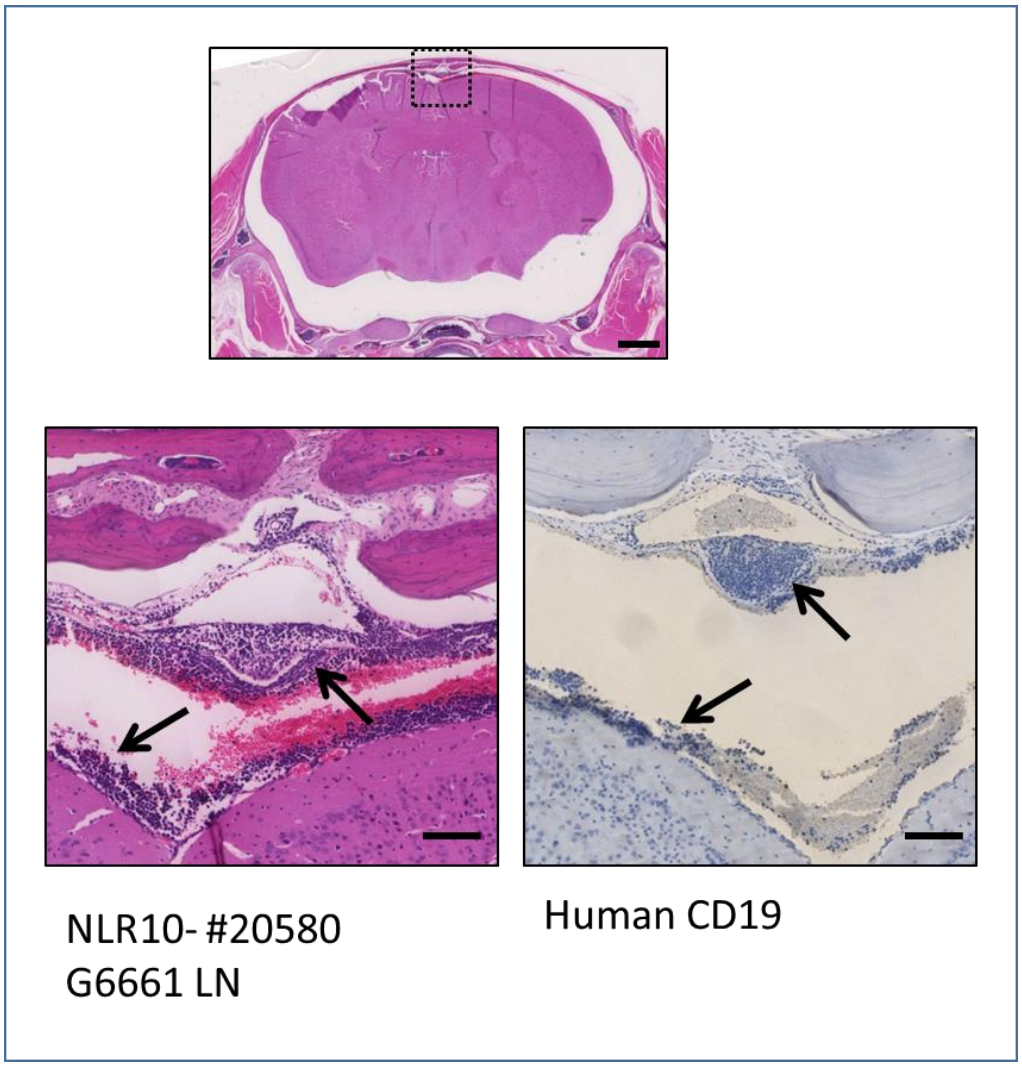


Figure 4.26. Mouse 20580 LN skull sections. Mouse was engrafted with viable cells from IGH-CEBPD patient 20580. Engraftment of cells in meningeal vein showing negativity for human CD19. Image created by Dr. Halsey.

The failure of initial primary samples to engraft was a hindrance to this project. However, during my time in Cincinnati, I was able to observe and use the chemotherapeutic agent, Busulfan, in conditioning of NSG mice (Section 2.3.7.2). After returning from Cincinnati, this technique was used in Newcastle in three NSG mice, when a new viable *IGH-CEBP* sample was received (patient 23395). The viable sample was injected after conditioning with Busulfan, currently all three mice have been sacrificed, having suffered from splenomegaly and anaemia. Spleens and BM have been collected and are being prepared for FACS analysis to confirm engraftment. Busulfan conditioning could be considered for future use with viable patient samples, to decrease the incidence of engraftment failure, which is more common among viable samples from low and intermediate risk ALL patients.

4.7 Discussion

The aim of this section of the project was to overexpress *CEBPD* within an early haematopoietic setting, both alone and in combination with IK6, to expand the transduced cells *in vivo* and *in vitro* and to observe the effects of the chosen retroviral vectors on proliferation and lineage commitment. The hope was to generate a CD19+ lineage among *CEBPD* transduced cells, to confirm that overexpression of this transcriptional regulator is an oncogenic event which can drive lymphoid differentiation.

The process of transducing CD34+ cells was successfully achieved in Cincinnati and, after several optimisation steps, was successfully recapitulated in Newcastle. Some of the issues encountered in Newcastle warrant further discussion.

The FBS batch used for the culture of CD34+ cells proved to be crucial for cell expansion. Initial attempts to expand CD34+ cells in Newcastle failed. One cause of failure may have been the presence of high levels of endotoxin in a particular FBS batch. Endotoxin is a complex lipopolysaccharide, which is a main component of the outer membrane of most gram-negative bacteria. Levels of

endotoxin had been found to vary greatly between FBS batches in the 1980s, going from 0.006 ng/ml to 800ng/ml (Case-Gould, 1984). Although improved handling protocols and filtration have since decreased overall endotoxin levels, it was discovered that concentrations as low as 1ng/ml could inhibit colony formation of murine erythroid cells (Case-Gould, 1984).

Failure of the initial CD34+ cells to expand may have been due to naturally varying endotoxin levels between FBS batches, as ultimately it was discovered that both high performance and standard FBS could be used to culture CD34+ cells (Figure 4.10). The origin of FBS was shown not to be important as both non U.S. (FBS 2) and U.S. (FBS 3) FBS batches resulted in cell growth. Future culturing of CD34+ cells should only be performed with tested FBS batches. Analysis also identified heat inactivation of FBS as a positive factor in CD34+ cell expansion. Originally FBS heat inactivation was performed to inactivate complement and remove mycoplasma. The use of the technique is debated as improved filtration of FBS has led to the suggestion that heat inactivation is not only unnecessary, but deleterious to cell expansion. This hypothesis was tested, with heat inactivation at 56°C proving to negatively impact on most cell lines used (Scientific, 1996). It was hypothesised that the FBS batches used, was negatively influencing CD34+ cell proliferation, as heat inactivation at 56°C for 30 mins was shown to improve CD34+ cell proliferation with all FBS batches (Figure 4.10). The handling of CD34+ cells prior to transduction was a potential factor. In Cincinnati, cells were isolated from fresh cord blood, often less than 24 hours old. In Newcastle frozen cells were used (Allcells, England), which may have impacted on cell proliferation and decreased viability of CD34+ transduced cells at later weeks of the *in vitro* experiments.

Ultimately production of retroviral vectors for the transduction of CD34+ cells was successfully performed after optimisation of the protocol from the Mulloy group.

CEBPD was selected for transduction as it was the most commonly occurring *CEBP* gene within the *IGH-CEBP* cohort and it had been possible to inject two viable *IGH-CEBPD* positive patient samples into NSG mice for *in vivo* expansion. The original plan had been to compare the gene expression profiles of CEBPD transduced CD34+ cells with expanded viable patient samples. However, both viable samples failed to engraft.

CEBPD expression is documented to occur in late stage myeloid differentiation, functioning in granulocyte (Wang and Friedman, 2002) and macrophage maturation (Gery *et al.*, 2005). Expression of *CEBPD* in early haematopoiesis has not been reported, with *CEBPA* being the only *CEBP* gene expressed at the HSC stage (Zhang *et al.*, 2004) (Figure 1.10). *CEBPA* positively regulates and is positively regulated by *PU.1*, a master regulator responsible for myeloid and lymphoid differentiation over that of *GATA-1* driven erythroid commitment (Arinobu *et al.*, 2007).

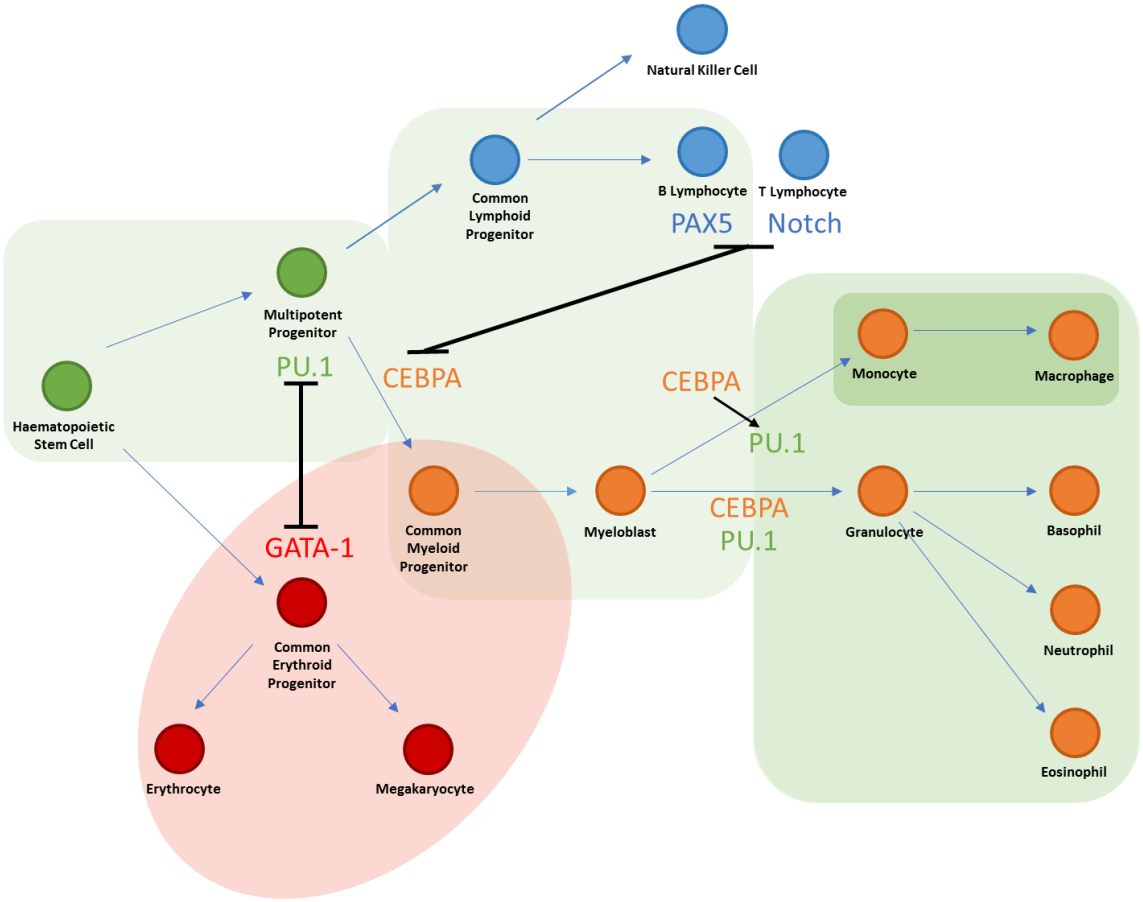


Figure 4.27. Interplay of CEBPA with PU.1 in haematopoiesis.

GATA-1 and PU.1 negatively regulate each other, controlling lineage commitment between GATA-1 positive erythroid cells or PU.1 positive myeloid and lymphoid cells. PU.1 expression increases with terminal differentiation of myeloid cells. Monocytes show the highest expression of PU.1. CEBPA interacts with PU.1 at multiple stages, including committing multipotent progenitors to the myeloid pathway, upregulating PU.1 as part of monocyte commitment, and also acting independently of PU.1 for granulocyte commitment. In lymphoid commitment the gene down regulates and is down regulated by PAX5 in B lymphoid cells, and Notch in T lymphoid cells. Green indicates expression of PU.1 in cells, with the darker green representing higher expression of PU.1. Red indicates expression of GATA1. Adapted from (Arinobu et al., 2007) and (Friedman, 2007).

The act of overexpressing CEBPD during early haematopoietic differentiation poses interesting questions due to the co-dependent nature of the CEBP proteins. CEBP ratios would be altered, and as a consequence so would *CEBP* function. Overexpressed CEBPD would likely form heterodimers with other available CEBP proteins, altering function and either knocking out the wildtype action of CEBPA homodimers, or functioning interchangeably, as has been shown with CEBPA and CEBPB (Jones *et al.*, 2002; Chiu *et al.*, 2004). The result of such heterodimerisation would be; either myeloid commitment, suggesting that CEBPD would function in the same role as CEBPA, supporting previously documented examples of *CEBP* gene functions as paralogues (Jones *et al.*, 2002): or lymphoid commitment, as in *IGH-CEBPD* BCP-ALL patients, suggesting that CEBPD is either inactivating myeloid commitment function of CEBPA by inhibiting its interaction with *PU.1* (Figure 4.27), or by encouraging lymphoid commitment through a different mechanism. The *PU.1* inactivation theory is supported by previous reports of *CEBPA* interaction with the gene. *CEBPA* and *PU.1* upregulation is reciprocal, without upregulation by *CEBPA*, *PU.1* would be expressed but not up-regulated (Figure 4.27) (Reddy *et al.*, 2002). This is important as high levels of exogenous *PU.1* in *PU.1* negative haematopoietic progenitors resulted in myeloid differentiation, while low levels of *PU.1* resulted in B-cell differentiation, (DeKoter and Singh, 2000). Thus, low levels of *PU.1* would theoretically be sufficient to commit haematopoietic progenitors away from the erythroid lineage, while not providing sufficient levels of *PU.1* to commit to myeloid differentiation, leading to lymphoid commitment, suggesting a lineage biasing mechanism in this leukaemia type. Inactivation of CEBPA by CEBPD may also be supported by CEBPA down regulation of the B-lymphoid differentiation gene, *PAX5*. Should this block be removed, lymphoid differentiation would occur more readily. There is a fallacy in this argument because if this were so then *IGH-CEBPA* patients would present with a myeloid leukaemia. As such it is more likely that *IGH-CEBP* patient lineage commitment is not directly influenced by the *CEBP* genes, potentially due to prior lineage commitment caused by an associating genetic lesion, such as a *CDKN2A/B*, *PAX5, IKZF1*, and *ETV6* deletions. In the clonal evolution FISH studies that were performed, *IGH-CEBP* translocations were observed to arise both prior to, and after associating CNAs.

For this study it was decided that screening by flow cytometry would be the most efficient method for analysis of retroviral expression in CD34+ cells. Marker, GFP and Thy1 expression showed variation between individual experiments. However some interesting trends were identified;

Over expression of CEBPD in CD34+ cells hinders cell proliferation, as observed in stromal cultures in Cincinnati, and stromal and suspension cultures in Newcastle (Figure 4.12). Rapid loss of GFP expression was exhibited in *CEBPD* transduced cells, both alone and in combination with IK6 (Section 4.3.5.3). The same observation was made in xenograft Experiment 1, where GFP positive cells were lost between weeks 15 and 18 (Figure 4.17). The failure of these cells to expand supports the suggestions that *CEBPD* negatively influences cellular proliferation, and while the transduced population in these cultures was short lived, they had a detrimental effect on the population overall. As *CEBPD* has not been previously overexpressed at the HSC stage, discussion as to how this effect is exerted must include observations made for other *CEBP* partners. If it can be assumed that *CEBPD* controls cell cycle in a haematopoietic setting in a similar or identical manner to the other *CEBP* genes (Johansen *et al.*, 2001; Gutsch *et al.*, 2011), or in the same manner as it has been shown in other cellular environments (O'Rourke *et al.*, 1999; Pawar *et al.*, 2010), then mechanistically this outcome can be explained by *CEBPD* mediated down regulation of cyclin complexes, *MYC*, *E2F1* and upregulation of *CDKN1B* (Figure 1.11) (Heavey *et al.*, 2003), which resulted in cell cycle arrest at the G0/G1 phase. Within an early haematopoietic setting, the findings relating to the effects of *CEBPA* on HSCs have been mixed. In a murine haematopoietic setting, *CEBPA* deletion resulted in a pronounced expansion of adult HSCs, as *CEBPA* was shown to function as a proliferative regulator of foetal to adult HSC differentiation, through down regulation of *MYCN* (Ye *et al.*, 2013). These findings have been disputed in another study, showing loss of *CEBPA* leading to loss of self-renewal potential and exhaustion of HSCs (Hasemann *et al.*, 2014).

Ultimately these findings may support either theory behind *CEBPA* function in HSCs, as *CEBPD* may be exerting an influence by inactivating *CEBPA* through protein heterodimerisation, therefore removing its function in HSCs, supporting the theory that *CEBPA* is important in HSC renewal. Alternatively, *CEBPD* may be functioning in tandem with *CEBPA* and exerting cell cycle control through

MYCN downregulation. To answer this question, the function of the CEBPA-CEBPD heterodimer would need to be ascertained: by tandem overexpression of both genes in target host cells, the presence of the heterodimer could be determined through Western Immunoblotting; cells could be sorted using FACS and gene expression or RNA expression studies performed.

Over expression of *CEBPD* alone and in tandem with IK6 was not sufficiently powerful to create an oncogenic phenotype in CD34+ cells. These double positive CD34+ cells displayed no improvement in proliferative or clonal potential. Expression of IK6 declined at a slower rate than the empty vector controls in experiments, suggesting that although IK6 was not boosting proliferation, it was not hindering cell division, especially when compared to CEBPD expression. However this trend was only observed once, while IK6 transduced cells did not expire as quickly as their CEBPD counterparts, no proliferative advantage was observed *in vitro* or *in vivo*. This negative observation was disappointing as *IKZF1* has been shown to inhibit cell cycle progression by: up regulation of *CDKN1B* kinase inhibitor, arresting cell line growth at the G0/G1 stage (Kathrein *et al.*, 2005), down regulation of *MYC* and *cyclin D3* expression, and up-regulation of *CDKN1B* in pre-B cells (Ma *et al.*, 2010). Similar work published during the course of this study investigated the interplay of the BCR-ABL1 fusion protein with IK6 in a CD34+ cell setting, with IK6 only cells found to have no proliferative advantage over controls (Theocharides *et al.*, 2015), supporting the findings in this study. More recent work on disease modelling in BCP-ALL has shown that more than two hits are required for oncogenesis, with pre-B cells expressing BCR-ABL1, *CDKN2A* (ARF) and IK6 being used to model ARF loss (Churchman *et al.*, 2015).

Transduced cells also failed to achieve consistent lymphoid transformation, as indicated by CD19 expression and GFP positivity. Although this effect was observed in CEBPD overexpressed CD34+ cells in one of four experiments (Figure 4.17B), the results were not successfully recapitulated and the observed population was minor and temporary (Figure 4.17). However, a similar observation was made in xenograft samples, where loss of CD19 expressing cells was observed between weeks 15 and 18. There are several potential reasons why an isolated experiment exhibited CD19+GFP+ cells. The first is retroviral integration. The MSCV retroviral vector was developed from the MLV virus. Studies into the integration sites of this virus and MLV-derived vectors found the

preferential integration sites were; 20% in the 5' end of transcription units, 17% in the vicinity of CpG islands (Felice *et al.*, 2009), 11% in the vicinity of DNase I-hypersensitive sites (Lewinski *et al.*, 2006) with the remaining integration sites being random (Mitchell *et al.*, 2004). Integration into the 5' region of a gene, such as *RUNX1*, *IKZF1*, *EBF1*, *E2A* or *PAX5,* could push commitment into the B-lymphoid lineage

Natural genetic variation between haematopoietic cell donors may be another cause. CD34 marker expression from Experiment 1 showed a large degree of deviation in both suspension and stroma cultures, with fluctuating patterns of high and low expression suggesting activation of CD34+ cells from a quiescent to a proliferative state several times over. All other experiments exhibited a pattern that was initially high CD34 expression followed by a steady decline over the course of the experiment, as the haematopoietic cells differentiated.

Additionally the integration of the viral vector in cells at different differentiation stages may play a large part in development of downstream clones. Integration of the CEBPD retrovirus into a pro-B cell of origin could be a strong push for myeloid differentiation, while integration into more mature pre-B cell or later, could fail to achieve the same differentiation due to V(D)J rearrangement resulting in commitment of the cell to the lymphoid lineage, resulting in a different oncogenic effect. Thus, transducing more mature B-cells may have resulted in more CD19+ CEBPD expressing cells.

CD19 expression was virtually non-existent in suspension cultures. This result was expected due to a lack of lymphoid lineage supporting cytokines. Low CD19 expression was observed in all transduced populations cultured on stroma, with a slight increase in marker expression observed in three of the experiments, over the course of each experiment. However only the Experiment 1 *CEBPD* population displayed a CD19+ GFP+ population, while CD33+ GFP+ populations were present in all experiments.

Currently *in vivo* experiments in Newcastle have yielded one mouse with potential engraftment of *CEBPD* transduced CD34+ cells, displaying a large spleen and anaemia. However this mouse was sacrificed only recently and no experimental data are yet available. Thus the Cincinnati *in vivo* experiments only can be discussed. Xenografts set up in Experiment 1 included three *CEBPD* mice and

three MIGR1 mice. All three *CEBPD* mice engrafted in comparison to one MIGR1 mouse.

Cell sorting of xenografts from Experiment 2 was successful, with both CD19 and CD33 positive cells isolated from five mice (Table 4.3). Overall numbers were low however, resulting in low levels of RNA and no available protein for downstream applications. Cells were also isolated from *in vitro* cultures in Experiments 3 and 4, for gene array analysis, but the resulting RNA was not of sufficient quality and/or quantity.

Several challenges were highlighted during this study, the greatest being low output of GFP+ cells at the end of the modelling process, with *CEBPD* transduced cells having a brief and limited proliferative life span. There was insufficient RNA and protein due to the low number of cells, which resulted in downstream applications, such as qPCR expression analysis, immunoblotting for protein expression, and cell cycle analysis, yielding poor quality data. This aspect was particularly challenging in stromal cultures, in which cells migrated under the MS-5 stromal layer, making it difficult to track cell numbers accurately and to harvest cells.

Other issues emerged from attempts to reproduce the same experiments in Newcastle. Stromal cultures did not show the same viability as those in Cincinnati, for which there were several potential reasons: in Cincinnati, high levels of CD34+ cells were observed to migrate and detach the stromal feeder layer. This problem was avoided in Newcastle by maintaining cell density below 5 $x10^5$/L. However the stromal feeder layer did not appear to support expansion of transduced cells as efficiently in Newcastle as Cincinnati, with slower growth of transduced cells observed. MS-5 cells cultured in Cincinnati and Newcastle were from different stocks, although both were less than 10 passages. Potentially identical sources of MS-5 cells should have been used, with an aliquot of cells sent from Cincinnati to Newcastle, if only to standardise experimental settings.

CD34+ cell source was also different between Cincinnati and Newcastle, with CD34+ cells in Cincinnati being isolated from fresh cord blood, and undergoing transduction within several days of extraction. In Newcastle, frozen cells were purchased and thawed for use. This may have impacted on later cell proliferation

and the decreased viability of CD34+ transduced cells in the later weeks of *in vitro* experiments.

Analysis of transduced cell populations was performed by FACS, however the methods varied between sites, with different fluorochromes, machines, lasers, and optimisation used. In Cincinnati three fluorochrome tubes were set up (Supplementary Table 7.2) with differing fluorochromes to test all lineage markers. Prior to every FACS run, machine voltage and compensation was set in real time. This adjustment was necessary, as the voltage of the BDS Canto II machine varied day to day, due to heavy use, making it impossible to use one standardised set of compensation and voltage parameters. It manifested as variation in marker expression from week to week, as shown in CD19 expression in Experiments 1 and 2 (Figure 4.17A-D), although this variation was minimised by using IgG and unstained controls (Section 2.3.8.1). In Newcastle, initially the BDS Canto II machine was used to investigate five fluorochrome markers (Section 2.2.10.1) (Supplementary Table 7.3), one compensation programme was set up for use across the entire experiment, with the aim of maintaining stable voltages to give clearer, less erratic marker expression. However week to week variation in GFP expression between CEBPD and MIGR1 transduced cells in Experiment 3, with CEBPD cells declining in GFP expression and MIGR1 cells increasing in expression (Figure 4.12 E&F), resulted in emission spectra of the FITC fluorochrome (490-525nm) overlapping with the dead gate, Zombie Aqua (425-516nm) marker, and CD11b Pacific Blue (410-455nm) marker. This problem was solved by changing FACS machines: the three laser BD Canto II machine was replaced by the five laser BD FortessaX20 machine, allowing fluorochromes to be separated onto individual lasers, removing the possibility of spectral spill over. The Zombie Aqua dead gate marker was also replaced by DAPI, due to the availability of a UV laser on the new machine. The switch to the FortessaX20 improved the quality of the overall data and the necessity for compensation adjustments.

As the aim was to recapitulate CEBPD overexpression in a haematopoietic environment similar to that of patients, use of more mature haematopoietic cells may have been appropriate, mimicking the natural development of ALL more closely. This could have been achieved through several methods: Extending the culture time of CD34+ cells to allow for differentiation into pre/pro-B cells before

retroviral transduction may have been more representative of true leukaemic development, where the activating genetic lesion likely arose in committed lymphoid cell. These cells could have been initially tracked by flow cytometry and transduced upon expression of early B-cell markers, such as CD10, CD24, and CD5. The same protocol could have been performed with inducible retroviral vectors, activated upon expression of the same markers.

Inducible vectors could have been used with multiple genes of interest, allowing for expression of selected genes in a controlled step wise manner. Such an experiment would allow us to implement the knowledge gained in Chapter 3, during analysis of the clonal evolution of *IGH-CEBP* BCP-ALL, investigating whether the order of acquired genetic lesions influenced cell lineage and proliferation. Another option would be the use of a lineage specific vector, such as a modified self-inactivating woodchuck hepatitis virus with a CD19 promoter, designed to specifically express only in B-lymphoid cells (Werner *et al.*, 2004).

A different option may have been expression of CEBPD in pre-B cell lines, such as REH, NALM16 or MHH-CALL-2. However each of these cell lines had an oncogenic phenotype from other genetic abnormalities such as *ETV6-RUNX1* (REH) or high hyperdiploidy (NALM16 and MHH-CALL-2). This option was not considered as Experiment 1 exhibited the increase in GFP+ CD19 + population, giving hope that such a result could be repeated.

In retrospect, the use of the IK6 dominant negative isoform for transduction at the HSC stage was an error, as knocking out of the *IKZF1* gene prior to lineage commitment of CD34+ cells has now been shown not to be conducive to lymphoid differentiation. Recent work performed in CD34+ cells, investigating co-expression of BCR-ABL1 and IK6, resulted in cells differentiating into myeloid leukaemia, with up-regulation in granulocyte-monocytic progenitor and stemness transcriptional programmes (Theocharides *et al.*, 2015). These results were unexpected, as the authors were originally attempting to create an ALL disease model. However it was discovered that IK6 expression strongly supressed B-cell progenitor programmes.

This problem may have been avoided by transducing pre/pro-B lineage cells, which were already committed to lymphoid differentiation, or avoiding IK6 expression and working with a different associating abnormality such as

CDKN2A/B knock down. However, knock down of genes is a less elegant method of genetic alteration, with difficulties in long term assays. Knock down of CDKN2A/B or *p16* has not been achieved over long periods of time. siRNA technology has achieved knock down of p16 for ~96 hours (Bond *et al.*, 2004). Improved RNAi technology has been developed, which persists for several days (Pieraets *et al.*, 2012). A potential method would be CRISPR-Cas technology, which has yielded guide RNAs for *CDKN2A* knock out, and transcription activator–like effector (TALE) technology, imitating epigenetic suppression, has been successfully carried out in fibroblast cells targeting *CDKN2A* (Bernstein *et al.*, 2015). However both of these methods are more recent developments, which were not available at the time of planning this study. Another option would have been the use of transgenic mice knockout models for CDKN2A/B, which have been developed for two decades (Serrano *et al.*, 1996). Such mice have been used in retroviral models to investigate the consequences of additional genetic lesions with *CDKN2A/B* loss (Williams *et al.*, 2006), resulting in transplantable leukaemia. A different choice would have been the knockdown / knockout of PAX5. As an important B lineage regulator, *PAX5* has been shown to function as a tumour suppressor in BCP-ALL (Dang *et al.*, 2015). A potential downside for use of PAX5 in this context may have been the loss of an important B-cell commitment gene in tandem with a myeloid gene resulting in strong myeloid commitment, although this hindrance could have been mitigated by introducing *PAX5* loss later in cell development and after initial lymphoid commitment.

Further work is also needed to determine the source of the cells detected in NSG mice calvaria sent to Glasgow. These cells were tested for hCD45 and hCD19 and tested negative for both, suggesting either a lineage switch of *IGH-CEBP* cells (hCD45- hCD19-), or infiltration of foreign mouse cells. To determine the source of the cells, staining with the murine splenic marker, CD45R, should first be performed as these cells tested negative for hCD45, hCD33 and hCD19 with flow cytometry (data not shown). In the event of a negative stain for the murine marker, a more widely expressed human marker must be considered for staining in order to determine the origins of the cells. An observed lineage switch would be particularly interesting, as there are no reports of *IGH-CEBP* translocations outside of BCP-ALL. Additional work could also be performed as spleen and BM viable cells were frozen for all three mice.

Ultimately although several challenges were presented in the expansion of transduced cells, some were successfully sorted for further analysis into the effects of *CEBPD*. Additionally, the potential engraftment of transduced CD34+ cells in the xenograft model in Newcastle provides further potential for additional work.

Chapter 5 Functional Characterisation of *IGH-CEBP* translocated B-cell precursor acute lymphoblastic leukaemia

5.1 Introduction

The use of gene expression arrays and RNA-seq technology has advanced scientific understanding of leukaemia at the transcriptional level. This technology has uncovered important pathway signatures through identification of vital transcriptional regulators. Such identification has improved stratification of patients who do not fit into the well-defined cytogenetic subgroups. These studies have led to the identification of more appropriate therapeutic agents, resulting in improved survival and reduced treatment toxicity. One example is the identification of the *BCR-ABL1*-like, or Ph-like ALL subgroup and treatment with TKI.

Identification of deregulated pathways in cancer is becoming increasingly important, as more targeted therapeutic agents are identified. Agents which are effective in specific cancers have also been shown to function in non-related malignancies, such as the successful use of the TKI, imatinib, in the treatment of gastrointestinal stomach tumours (Din and Woll, 2008), due to identical signalling pathways being altered.

Identification of specific deregulated pathways in *IGH-CEBP* BCP-ALL would elucidate both the development of the disease and provide potential targets for treatment. One obvious link with *CEBP* overexpressing leukaemias is the *RB*/*E2F* pathway, and inflammatory and immune system signalling, which have all been established as important CEBP functional domains (Section 1.5), providing potential pathways of oncogenic deregulation.

The mechanism behind *RB/E2F* pathway oncogenesis has been discussed in previous chapters, the function of inflammation in oncogenesis has not been discussed in great detail. Chronic inflammation is considered to be the hallmark of malignancy (Colotta *et al.*, 2009). The link between inflammation and cancer can be described in two main mechanisms.

1. Extrinsic mechanisms involving immune and microenvironment factors, where a constant inflammatory state contributes to the initiation and progression of the cancer.
2. Intrinsic mechanisms including acquired genetic lesions affecting oncogenes, tumour suppressors and genes involved in genome stability that contribute to the activation of the inflammatory pathways.

Multiple links between inflammation and leukaemia have been reported.

Extrinsic factors include regulation of tumour activated macrophages (TAMs), which have been shown to destroy AML cells in mouse models. Upregulation of the marker, CD47, which confers protection against TAMs, was shown to inhibit TAM function. Conversely AML engrafted mice, treated with an anti CD47 antibody, showed improved clearance of AML and improved survival (Jaiswal *et al.*, 2009; Majeti *et al.*, 2009; Chao *et al.*, 2011).

However, the majority of links between inflammation and leukaemia are intrinsic and linked to cytokine and chemokine signalling. Activation of cell surface receptors by cytokines, chemokines and growth factors regulate signal transduction activity and the interaction between cells and the bone marrow microenvironment (Ferretti *et al.*, 2012). The frequency of activation of signal transduction in AML exceeds the levels of observed mutations or genetic lesions found in pathway receptors, suggesting an alternative mechanism of stimulation in pathway signalling. This signalling provides leukaemic cells with selective advantages through inhibition of apoptosis, stimulation of proliferation and blocking differentiation (Van Etten, 2007).

There is evidence of abnormal cytokine signalling in AML, which has been tested via single cytokine serum level analysis (Schwieger *et al.*, 2009). In depth studies have been performed on 27 patients with AML and myelodysplastic syndromes. Expression of specific cytokines and chemokines were found to be predictive of outcome including *CCL5*, *IL-8*, *IL-2*, *CCL4* and *IL-5* were predictive of outcome, and a panel of 11 cytokines and chemokines (including *CCL3* and *CCL5*) that stratified patients into favourable, intermediate and adverse groups with statistically significant median survival rates ($p = 0.003$) (Kornblau *et al.*,

2010). Cytokine expression levels decreased in leukaemia patients upon remission (Van Etten, 2007).

The main regulators of inflammation are the *NF-κB* transcription factor family. They have been shown to activate more than 200 genes, including inflammatory cytokines, adhesion molecules and *COX2* genes (Sun 2011). Many promote survival of malignant cells through induction of anti-apoptotic genes, such as *BCL2*. *NF-κB* is sequestered in the cytoplasm and held in an inactive state by Inhibitor-κB (*IκB*). Signalling from inflammatory cytokines or chemokines, activation of the immune system, or DNA damage, results in phosphorylation of IκB and release of *NF-κB* for entry into the nucleus and initiation of signalling. Downstream targets include activation of the bZIP genes, *JUN* and *FOS.* They combine post transcriptionally to create the AP1 transcription factor protein, which exerts a direct effect on cell proliferation, differentiation, and apoptosis (Hess *et al.*, 2004). Constitutive activation of *NF-κB* is frequent in malignant cells (Giles *et al.*, 2014). In leukaemia, upregulation of *NF-κB* has been observed in multiple studies, but no direct observation on survival has been shown (Estrov *et al.*, 1999; Guzman *et al.*, 2002; Frelin *et al.*, 2005). In *in vitro* studies deregulation of *NF-κB* signalling in CD34+ cells alone was insufficient to induce oncogenesis (Romano *et al.*, 2003; Schepers *et al.*, 2006). It has been hypothesised that *NF-κB* activation is the result of autocrine production of cytokines (Dokter *et al.*, 1995), mainly expressed in LSCs. This activation confers chemotherapeutic resistance to these cells leading to relapse of the disease (Guzman *et al.*, 2001). Specifically in ALL, constitutive activation of *NF-κB* occurred in Ph+ samples, while Ph- samples and B-precursor cell lines had normal expression (Munzert *et al.*, 2004). Multiple other genes and gene families have been implicated in regulation of the *NF-κB* signalling pathway, including vascular endothelial growth factors, tumour necrosis factors, toll like receptors, and matrix metalloproteases (Giles *et al.*, 2014).

STAT3 is a member of the *STAT* family, which is affected by multiple inflammatory signals and are involved in hematopoietic cytokine receptor signalling pathways, which control cell proliferation, differentiation and survival. This gene family is deregulated in AML, and constitutive activation has been associated with poor outcome (Benekli *et al.*, 2002). These genes are activated by multiple cytokines (IL-5 and IL-6) which interact to control expression of a

number of genes, such as *PAI-1*, which is involved in cancer cell metastasis and angiogenesis (Placencio and DeClerck, 2015), and anti-apoptotic genes *BCL2* and *BCL3* (Grivennikov and Karin, 2010).

The *COX2* signalling gene, functions by initiating the formation of prostanoids, and is another major mediator of inflammation observed in multiple malignancies. Increased expression of *COX2* has been identified in CML and CLL, with correlation to prognosis (Bao *et al.*, 2007). *AICDA* (*AID*) provides another potential link between inflammation and leukaemia. This gene is involved in somatic hyper mutation and class switch recombination (Wang *et al.*, 2014) , chronic inflammation has been shown to trigger aberrant *AID* expression in B-cells. *NF-κB* has been shown to mediate expression of *AID (Mechtcheriakova et al., 2012)*, leading to increased oncogenic aberrations *in vivo* (Shimizu *et al.*, 2012).

Multiple other genes and gene families have been implicated in regulation of the *NF-κB* signalling pathway, including vascular endothelial growth factors, tumour necrosis factors, toll like receptors, and matrix metalloproteases (Giles *et al.*, 2014).

The ability to identify the aberrant pathway signalling in leukaemia is the cornerstone of modern cancer treatment, both through stratification and targeted therapy.

Due to the complex and highly interactive nature of molecular pathways, multiple downstream genes can be targeted for therapy. One example is *KMT2A-AF4* rearranged ALL, which expresses high levels of *BCL2*. *KMT2A-AF4* positive cells have been successfully eliminated by the *BCL2* antagonist, ABT-199, in cell lines and xenograft models (Benito *et al.*, 2015). Such approaches directed to downstream targets are highly specific to leukaemic cells, thus more protective of normal cells.

The ever increasing numbers of targeted chemotherapeutics has made identification of affected molecular pathways highly valid option for the improvement of patient treatment and in turn survival.

5.2 Aims

Having performed genetic screening of *IGH-CEBP* patient samples, this section of the project aimed to investigate the consequences of *CEBP* overexpression at the molecular level. RNA-seq technology was used to identify potential genes and pathways affected in available samples. Analysis of *CEBP* expression was also performed by qPCR to investigate *CEBP* expression fold changes in comparison to both patient and normal controls. Protein analysis was performed using SDS PAGE to investigate *CEBP* protein isoforms in available samples, to elucidate post transcriptional *CEBP* expression.

Aims

- To investigate *CEBP* expression levels in patients using qPCR.
- To investigate CEBP protein isoform expression in relation to mRNA expression in available patient samples.
- To analyse RNA sequencing data of two *IGH-CEBP* patients to identify differentially expressed genes and pathways.

5.3 Results

5.3.1 Patient Analysis - qPCR shows CEBP mRNA expression is upregulated in IGH-CEBP BCP-ALL patients

The expression of *CEBPs* was investigated in two patients using the qPCR TaqMan platform to assess the extent of *IGH* mediated overexpression (Figure 5.1). Patient 11739 had the *IGH-CEBPB* translocation in 45% of blasts, and patient 23395 the *IGH-CEBPD* translocation in 61% of blasts. The samples were compared against other BCP-ALL patients (Supplementary Table 7.26). The following controls were chosen: normal liver was a positive control for *CEBPB* and *CEBPG* and a negative control for *CEBPD*; normal kidney as a positive control for *CEBPD* and negative for *CEBPB* and *CEBPG*; BM was selected as a reference sample. Genes analysed were the known *IGH* partners of the two selected patients and the *CEBPG* gene, which functions as a dominant negative isoform.

Expression of *CEBP* genes among control BCP-ALL patients was varied (Figure 5.1). *CEBPB* and *CEBPD* are typically expressed in a tissue and context specific manner for short periods of time, while *CEBPG* is generally expressed across all tissue types. *CEBPG* expression was observed across all patients and controls. Expression was particularly high in patients 2058, 4679, 23395 and 11739 (increase of 14, 15, 5 and 3 fold, respectively) when compared to BM cDNA.

IGH-CEBPB patient 11739, showed a high level of expression of *CEBPB* (increased 2.8 fold) in comparison to other patients. High expression was also observed in the liver positive control. *CEBPB* expression was low among all other patient and control cDNAs. *CEBPD* expression was low and restricted to the *IGH-CEBPD* patient 23395 (increased 1.5 fold) and positive control kidney tissue. These data indicate that respective *CEBP* genes were overexpressed in the patients analysed.

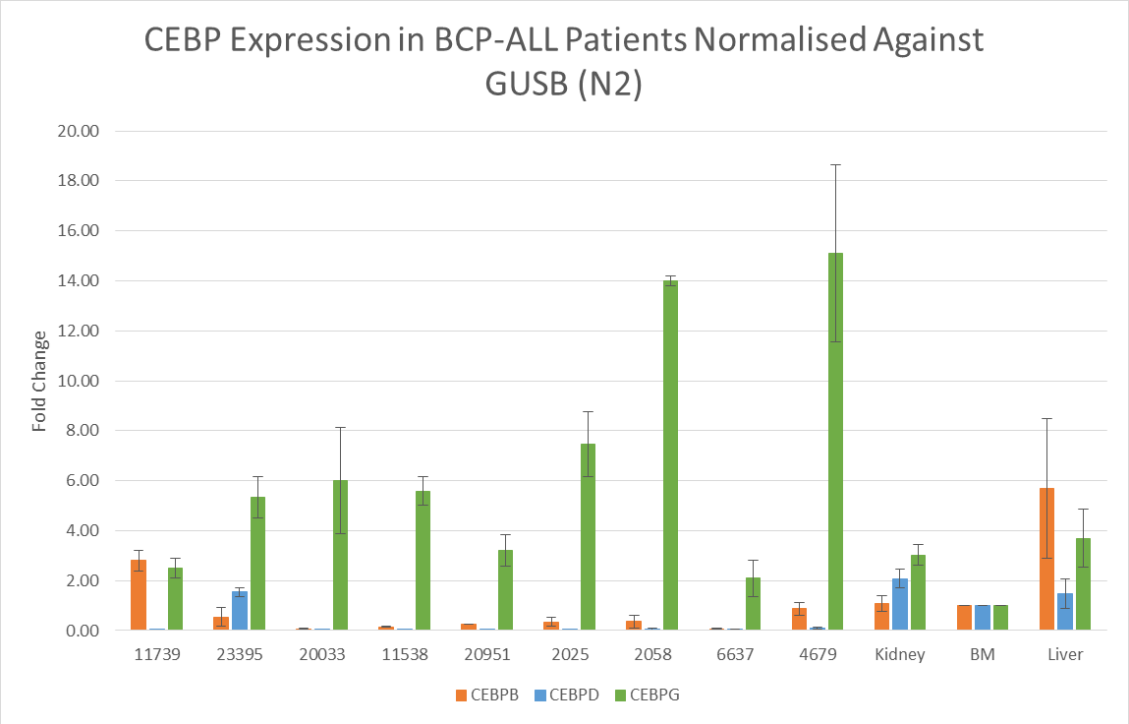


Figure 5.1. Expression of CEBP genes in IGH-CEBP patients, with BCP-ALL and tissue controls, normalised against the GUSB house keeping gene with BM used as the reference sample. Experiment performed in duplicate.

5.3.2 Patient Analysis – CEBPB Protein Expression using SDS PAGE Shows dominant expression of the LAP and LAP isoforms*

Identifying the CEBP protein isoforms expressed in patient samples may elucidate the mechanism of action behind *IGH-CEBP* translocated BCP-ALL. Protein samples were only available for patient 11739 for investigation (Figure 5.2). Other BCP-ALL samples from patients without an *IGH-CEBP* translocation were used as controls (Supplementary Table 7.27). Protein concentration was low, between 30-40ng/µl for control samples and 46ng/µl for sample 11739 (Supplementary Table 7.27); α-tubulin expression low in control 1 and 2, with no protein in control 3. Although α-tubulin expression was low in Patient 11739, CEBPB expression was evident in comparison to patient controls. All three isoforms were present, with the 44kDa LAP* and 42kDa LAP proteins being the most highly expressed. The dominant negative isoform LIP was visible at 20kDa at a lower concentration. Other bands were also visible, suggesting post transcriptional modification of the CEBPB protein. These data suggested that CEBPB was not being inactivated through LIP inactivation in patient 11739.

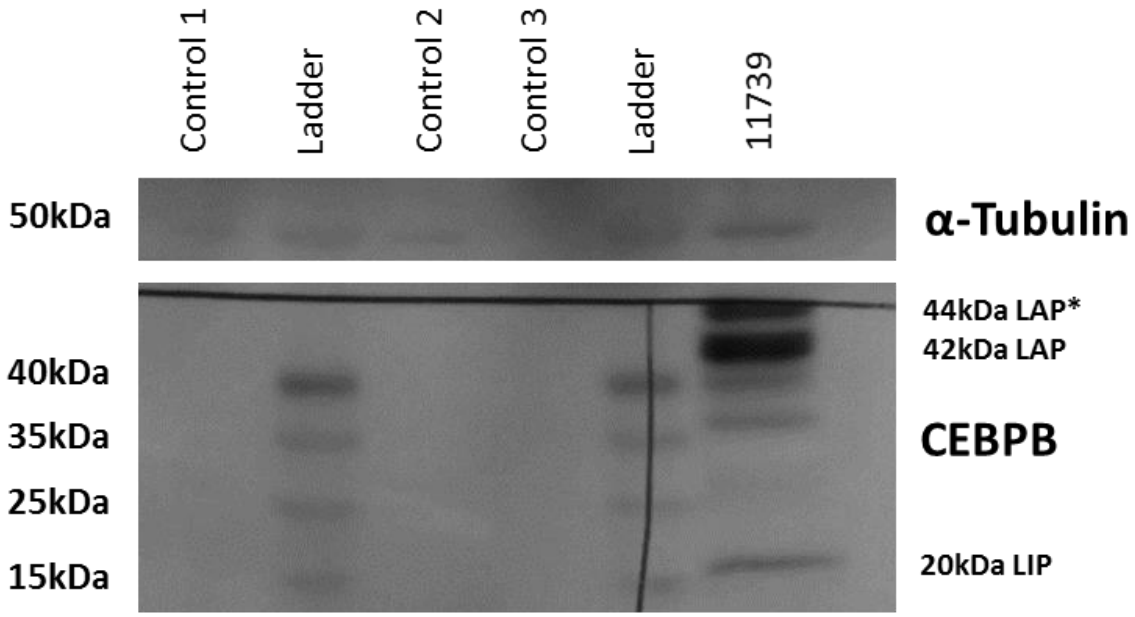


Figure 5.2. Immunoblot for CEBPB protein of IGH-CEBPB 11739, and BCP-ALL control sample lysate, with α-Tubulin loading control.

Although all samples show low levels of α-tubulin loading control CEBPB protein is clearly expressed in patient sample 11739. All three CEBPB protein isoforms are visible, LAP at 44kDa, LAP at 42kDa and LIP at 20kDa. The sample shows that the IGH mediated overexpression of the CEBPB gene is mainly driving expression of the functional LAP* and LAP proteins.*

5.3.3 Patient Analysis - RNA-seq

Two patients, 11739 and 23395, had sufficient RNA for RNA-seq. These samples were normalised against 14 BCP-ALL control samples, comprising *IGH*-other, iAMP21, and B-Other patients (Supplementary Table 7.28). RNA reads were converted to $\log_2$ fold change values, using the DESeq2 method, which minimises individual reads exponentially, in order to tackle the problems inherent with small replicate numbers, large dynamic range of reads and outliers (Love *et al.*, 2014). These values were then subtracted from controls to give the final $\log_2$ fold change values, discussed below in relation to transcript level. While data obtained from this analysis cannot be considered significant, due to a low sample size, some interesting trends were observed.

5.3.3.1 *IGH-CEBPB Patient 11739 shows variation in downstream effects on CEBPB targets*

The first focus of the RNA-seq data analysis; the fold change of *CEBP* family members was investigated. *IGH-CEBP* Patient 11739 showed a 2.8 fold increase of *CEBPB*, while the other *CEBP* genes were downregulated, with the exception of *CEBPZ* (Figure 5.3 A). This result conflicted with expression of *CEBPG* on the qPCR platform, where expression was higher (Figure 5.1). RNA-seq data however must be considered against the controls it is normalised to, as other BCP-ALL patients may have high expression of *CEBPG.* Indeed, such a trend was observed in the qPCR analysis of BCP-ALL controls, in which *CEBPG* was highly expressed across all patients. Additionally, normal BM comprises a mixture of haematopoietic cell types, with varying expression levels of *CEBPG*, while patient controls would be predominantly immature B-cells.

Genes known to be involved in haematopoietic differentiation were investigated in an attempt to elucidate the mechanism of linage commitment. Patient 11739 showed decreased expression of both erythroid *GATA* factors and myeloid *ID* family genes, while lymphoid commitment genes, *IKZF1*, *PAX5*, and *EBF1,* showed increased expression (Figure 5.3 B). A 2.1 fold increase in *IKZF1* expression was notable as this patient had the 4-7 *IKZF1* deletion, which leads to expression of the IK6 isoform. Overexpression of this dominant negative isoform may be considered as an oncogenic hit, contributing to the development of the leukaemia. One copy number gain of *EBF1* correlated with a 0.8 fold increase in RNA transcripts.

Early cell cycle phase genes, comprising predominantly of the *RB*/*E2F* pathway, were investigated in more detail due to the documented interactions with the cyclin kinase inhibitors, cyclins, *RB* family genes, and *E2F* genes (Tsukada *et al.*, 2011). Patient 11739 showed upregulation of *RB1*, *MYC* and cyclin / cyclin kinase complexes (Figure 5.4 A). The deleted cyclin inhibitors, *CDKN2A* and *CDKN2B,* showed reduced expression of both transcripts at -1.1 and -1.7 fold, respectively. Due to the importance of the *CDKN2A*/*B* pathway there are many regulatory mechanisms influencing the expression of these genes. *CDKN2A*/*B* genes are directly expressed erythroblast transformation- specific (ETS) transcription factors *ETS1*, *ETS2*, and *E47*, in response to oncogenic and senescent signalling (Ohtani *et al.*, 2001), while *ID1* functions to repress *CDKN2A*/*B* expression.

Lowered expression of the *ID* genes in patient 11739 may be factor influencing expression of *CDKN2A*/*B*. *JUN* is another repressor of *CDKN2A*/*B* expression which shows lowered expression in patient 11739.

Downstream, the *E2F* family is divided into *E2F* activator genes, *E2F1-E2F3,* which function in cell cycle progression, and *E2F* repressor genes, *E2F4-E2F8*, which function to downregulate *E2F* activator genes (Polager and Ginsberg, 2008). These genes showed mixed levels of expression. *E2F* activator genes showed low expression, while *E2F* repressor genes were predominantly upregulated, suggesting proliferative inhibition (Figure 5.4 C). Direct downstream targets of the *E2F* family, linked with cell cycle progression, showed decreased expression, while *E2F* targets linked with cell synthesis were expressed normally, supporting observation that cell cycle may be inhibited in this patient (Figure 5.4 D). Several cell cycle genes showed variant expression, with genes directly promoted by *CEBPB*, *MYC* and *MYBL2* showing downregulation.

As *CEBP* genes are heavily regulated post transcriptionally, other *CEBP* targets were assessed to determine whether *CEBP* overexpression was translating into visible downstream effects. Several direct targets of *CEBPB*, broadly divided between functioning in the immune system and in haematopoiesis, were analysed. An interesting trend was observed, in which a number of genes reported to be upregulated by *CEBPB* were found to have low expression in patient 11739. This was found in 3/4 immune system genes and 2/6 haematopoiesis genes comparing against previously published data (Figure 5.5 C&D) (Supplementary Table 7.29). The two genes with lowered expression in haematopoiesis, *ID1* and *ID2*, were of particular interest as they have been shown to function as part of the B-cell differentiation pathway, facilitating differentiation beyond the pro-B stage. Lowered expression of these genes could be suggestive of a B-cell differentiation block.

Genes with the greatest differential expression showed few obvious trends. In patient 11739, both oncogenes and tumour suppressors were upregulated, all of which were shown to be associated with prognostic outcome in solid tumours such as head and neck cancer, due to strong links in metastasis (Supplementary Table 7.30). Three genes showing significant downregulation were involved in the inflammatory pathway, two of them, *JUN* and *FOS,* heterodimerise to create the AP1 transcription factor, which is involved in the inflammatory cascade,

proliferation and cell differentiation (Angel and Karin, 1991; Zenz *et al.*, 2008). The third one, *CXCL8,* is a potent pro-inflammatory signalling chemokine (Gales *et al.*, 2013).

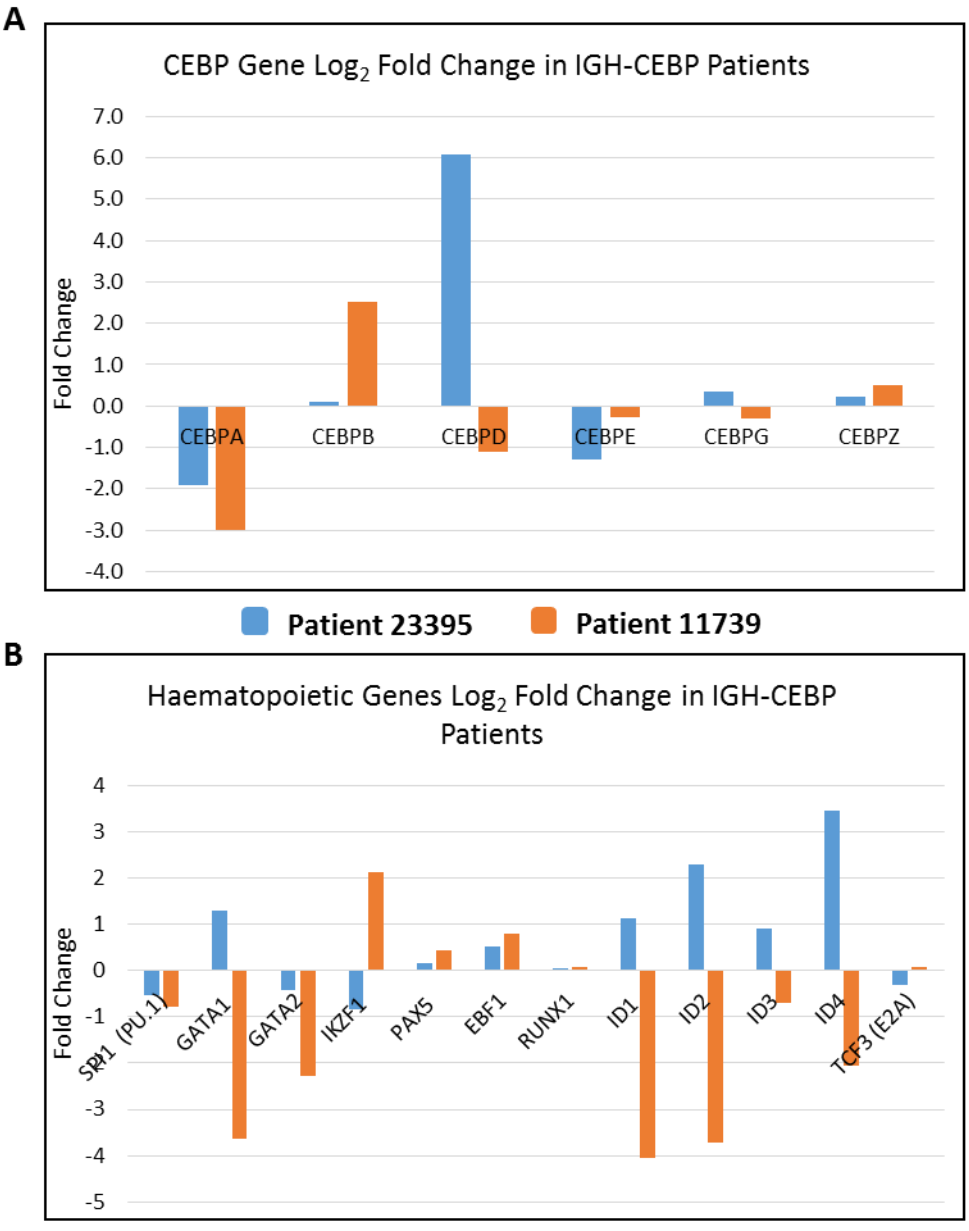


Figure 5.3. RNA-seq expression of CEBP family genes in IGH-CEBPD patient 23395 and IGH-CEBPB patient 11739. Values were calculated from RNA-seq transcript reads and converted into fold change values. Data was normalised against control BCP-ALL patient samples.

A. The expression of the CEBP genes shows a variation between patients, however there were some trends observed. Over expression of the IGH partner is visible in both patients. Myeloid biasing CEBPs; CEBPA, CEBPB, CEBPD, CEBPE are either downregulated or show very low expression, with the exception of the IGH partner CEBPB and CEBPD in the respective patients. B. Expression of haematopoiesis differentiation genes. Patient 11739 shows down regulation of

multiple myeloid differentiation genes with all GATA and ID family genes showing down regulation, and IKZF1 upregulated. Patient 23395 shows the opposite regulation of myeloid genes and down regulation of IKZF1.

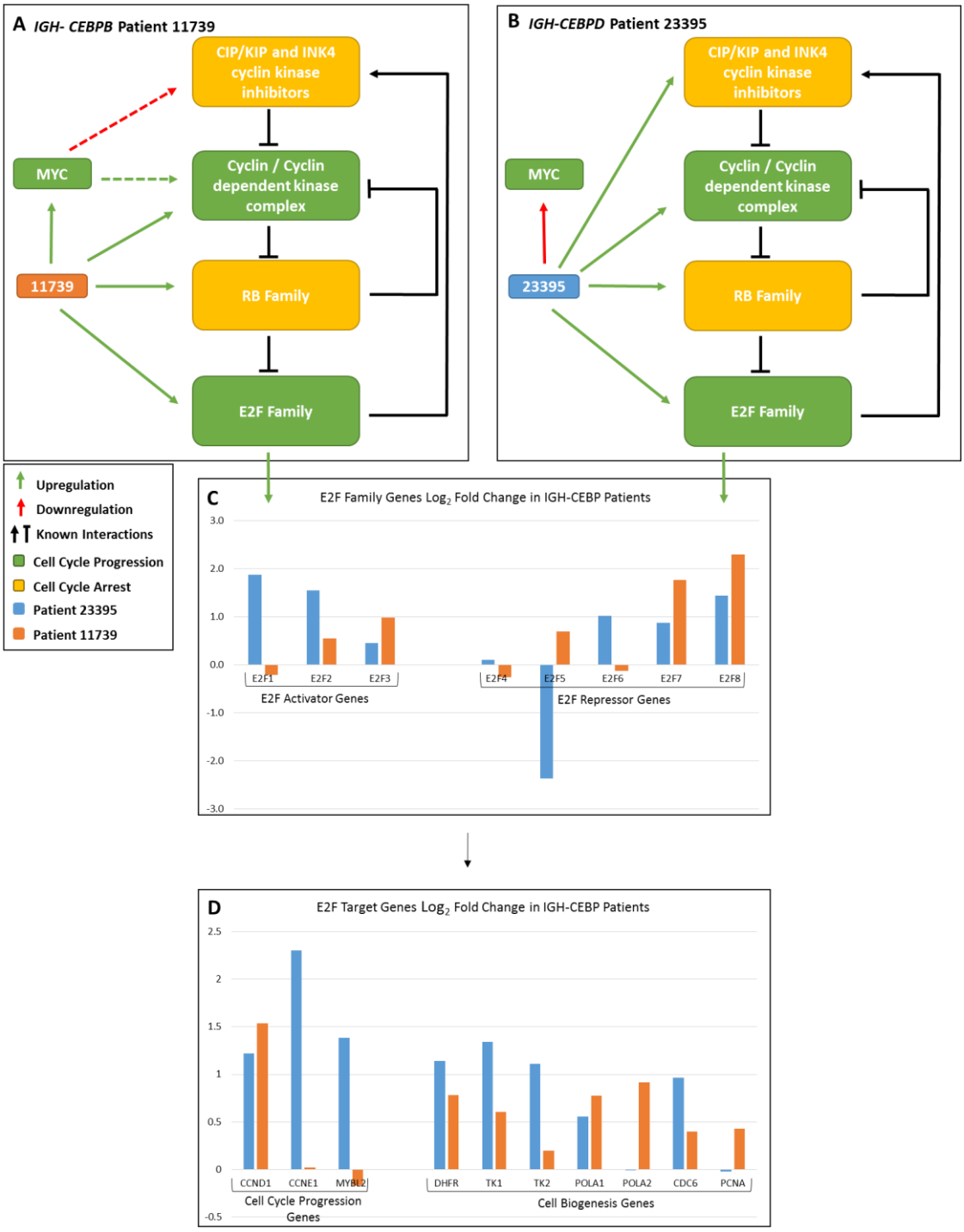


Figure 5.4. RNA sequencing displaying log_2 fold change in the RB/E2F pathway and downstream targets dictating cell cycle progression from G1 to S phase.

A. Patient 11739 shows marginal upregulation of MYC (increase 1.8 fold), Cyclin and Cyclin dependent kinases, RB family and E2F family genes, while cyclin inhibitors were downregulated. Dashed arrows show potential mechanisms of upregulation or downregulation as a result of MYC upregulation. B. Patient 23395 shows upregulation of all gene groups with the exception of MYC which was downregulated (decrease of 0.2 fold). C. A detailed view of the E2F family and individual gene log_2 fold changes in patients 23395 and 11739. E2F genes are divided between E2F activator genes E2F1, E2F2, and E2F3, and E2F repressor genes E2F4, E2F5, E2F6, E2F7, and E2F8. Patient 23395 shows higher expression of E2F activators while patient 11739 shows downregulation of the primary activator E2F1 and upregulation of multiple E2F repressor genes. D. Downstream pre-transcriptional targets of the E2F family divided by function into cell cycle progression genes and cell biogenesis genes. Direct E2F targets show upregulation in patient 23395, particularly cell cycle progression genes, while patient 11739 shows low expression of cell cycle genes and high expression of cell biogenesis genes.

5.3.3.2 *IGH-CEBPD Patient 23395 shows expected downstream effects on CEBPD targets*

IGH-CEBPD Patient 23395 displayed no CNAs detectable by MLPA and had insufficient DNA for SNP6.0 analysis. RNA-seq showed the expected upregulation of the *IGH* partner *CEBPD*, while the *CEBP* with myeloid bias; *CEBPA*, *CEBPB* and *CEBPE*, were either down regulated or showed only minor transcriptional alterations (Figure 5.3 A).

Expression of haematopoiesis differentiation gene *GATA.1* and the myeloid committing *ID* family were high, while *IKZF1* and *TCF3* were low (Figure 5.3B). The patient predominantly showed high expression of both cell cycle progression and cell cycle arrest genes, with *MYC* being the most notably downregulated gene (Figure 5.4B). Despite *MYC* downregulation patient 23395 showed high expression of *E2F* activator genes and low expression of *E2F* repressors, in particular *E2F5* (Figure 5.4C). High expression was observed in downstream *E2F* activator target genes, in particular cell cycle progression genes; *CCNE1*, and *MYBL2* (Figure 5.4D), suggesting that despite increased expression, the *RB* family were not functioning to inactivate *E2F* activators. Direct *CEBPD* targets (including *CREB1*, *CREBBP*, *PPARG*, *IL-6* and *PER2*), with a number of different functions, predominantly showed transcript levels to be at expected from previous publications (Supplementary Table 7.31), with only 2/11 genes showing variation from expected transcript levels (Figure 5.5A&B) (Supplementary Table 7.31).

Of note, genes involved in *PI3K*/*Akt* signalling pathway (*FN1*, *TNC*, *COL1A*, *COL3A1*, *COL11A1* and *IBSP*) were amongst the most highly expressed genes in this patient. The majority were collagens and fibronectins, likely functioning through binding of the *PTK2* (*FAK*) receptor, leading to activation of *PI3K* (Xia *et*

al., 2004; You *et al.*, 2015). Multiple oncogenes associated with solid tumours, where they enhance metastasis and angiogenesis, were also observed (Supplementary Table 7.32). Genes involved in inflammatory pathways were also upregulated, including *MMP13* a matrix metalloprotease, *GJA1* involved in intracellular communication, and the *CXCL12* chemokine.

5.3.3.3 Comparisons between patient 11739 and patient 23395 shows altered CEBP function.

There was a high level of variability between the two patient samples analysed on the RNA-seq platform. Initial comparison of these data was focused on the *CEBP* genes and their direct targets.

In both patients the respective *CEBP* gene partnered with *IGH* (*CEBPB* for patient 11739, and *CEBPD* for patient 23395) was overexpressed, while the early myeloid biasing gene *CEBPA* was downregulated in both patients (Figure 5.3 A).

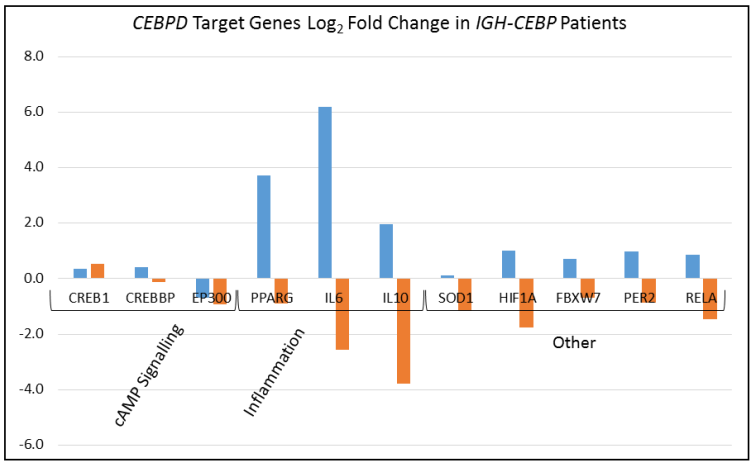
Several *CEBP* target genes showed differential expression between the two patients. Reduced expression levels of *MYC* were observed in patient 23395, supporting suggestions of *CEBPD* mediated *MYC* downregulation, observed in myeloid cell lines (Gery *et al.*, 2005). However, *MYC* showed high expression in patient 11739 (Figure 5.4 A). Other genes that showed differential expression were *ID1* and *ID2*, downregulated and upregulated in patient 11739 and 23395, respectively (Figure 5.5 C&D).

Genes involved in the inflammatory response that are activated by *CEBPD* and *CEBPB* (Yan *et al.*, 2012) were downregulated in patient 11739, suggesting that *CEBPB* function was impaired (Figure 5.5 A). A similar trend was observed for another *CEBPB* target, *PPARG*. This gene is found to be upregulated by both *CEBPB* and *CEBPD* in the literature, and was upregulated in patient 23395, but not in patient 11739, again suggesting that *CEBPB* function was impaired (Figure 5.5 C).

A

Pathway / Function	Gene	CEBPD Role	23395 Log_2 Fold Change	CEBPB Role	11739 Log_2 Fold Change
cAMP signalling	CREB1	Promoter	0.3	NA	0.5
	CREBBP	Promoter	0.4	NA	-0.1
	EP300	Promoter	-0.7	NA	-0.9
Inflammation	PPARG	Promoter	3.7	Promoter	-0.9
	IL6	Promoter	6.2	Promoter	-2.6
	IL10	Promoter	2.0	Promoter	-3.8
Apoptosis	SOD1	Promoter	0.1	NA	-1.1
Hypoxia Signalling	HIF1A	Promoter	1.0	NA	-1.7
Tumour Supressor	FBXW7	Repressor	0.7	NA	-0.7
Ubiquitination	PER2	Promoter	1.0	Promoter	-0.9
Immune System	RELA	Promoter	0.8	NA	-1.5

B



C

Pathway / Function	Gene	CEBPB Role	11739 Log_2 Fold Change	CEBPD Role	23395 Log_2 Fold Change
Immune System	CD14	Promoter	0	NA	0.8
	CXCL8	Promoter	-7.5	NA	3.1
	PPARG	Promoter	-0.9	Promoter	3.7
	CHIT1	Promoter	0.8	NA	0.7
Haematopoiesis	ID1	Promoter	-4	NA	1.1
	ID2	Promoter	-3.7	NA	2.3
	GATA2	Repressor	-2.3	NA	-0.4
	ZFPM1	Repressor	-1.6	NA	-1
	MYB	Promoter	0.4	NA	1
	SMARCA2	Promoter	0.9	NA	0.5

D

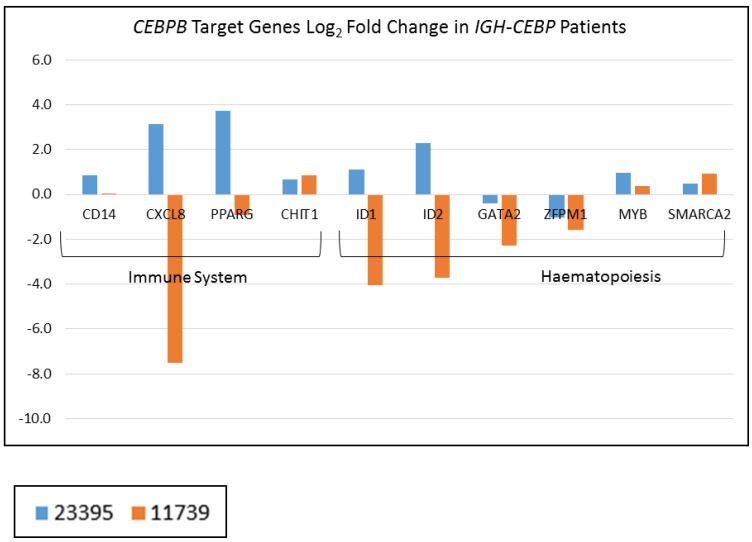


Figure 5.5. RNA sequencing fold variations of CEBP target genes in the two IGH-CEBP patients.

A. Table showing CEBPD pre-transcriptional target genes divided by function, with the role of the CEBPD gene next to each target. Genes found to be expressed differentially to previously published findings (Supplementary Table 7.32) are highlighted in yellow. This table shows that CEBPD broadly functions as expected in patient 23395, while IGH-CEBPB

patient 11739 shows altered function of genes from expected CEBP effects, such as downregulation of inflammation genes and PER2 despite previous reports as a promoter. B. Graphical representation of table A. C. Table showing CEBPB pre-transcriptional target genes divided by function, with the CEBPB gene role next to each target. Genes found to be expressed differentially to previous findings (Supplementary Table 7.29) are highlighted in yellow. This table shows that CEBPB is showing mixed expression of target genes with 3/4 and 2/6 genes differentially expressed in immune system and haematopoiesis gene subsets respectively. D. Graphical representation of table C.

In patient 23395, *CEBPD* downstream targets involved in inflammation showed strong upregulation. This upregulation would confer advantages to the recipient blasts in the form of increased mutagenesis through DNA damage (*CXCL8*, *IL-6*), increased chemotaxis, which has been shown to increase resistance to chemotherapy (*CXCL12*, *MMP13*, *IL-6*), and increased proliferation (*FOS*, *PPARG*, *IL-6*) (Giles *et al.*, 2014). Patient 11739 exhibited the opposite expression profile, with multiple inflammatory genes downregulated (Figure 5.6 B&C). It is interesting to note that *NF-κB* factors showed no upregulation in these patients, suggesting that the inflammatory upregulation observed in patient 23395 is mediated through another pathway.

A **Patient 23395**

Top 15 Upregulated Genes	Log_2 Fold Change	Top 15 Downregulated Genes	Log_2 Fold Change
MMP13	9.4	ZNF793	-6.1
POSTN	8.0	CALN1	-5.6
TNC	8.0	ZNF135	-5.5
MYF6	7.7	ZNF582-AS1	-5.2
COL3A1	7.7	CLECL1	-5.2
IBSP	7.6	LOC100128398	-5.1
LUM	7.3	PNMA3	-5.0
COL1A2	7.2	CSMD1	-5.0
FN1	7.1	EVC2	-5.0
GJA1	7.1	MYO7B	-4.9
IGFBP5	7.0	GPA33	-4.8
STC1	7.0	BMPR1B	-4.8
ZNF804A	7.0	RGMA	-4.7
COL11A1	6.9	SEMA6A	-4.7
CXCL12	6.8	PLEKHG4B	-4.7

B **Patient 11739**

Top 15 Upregulated Genes	Log_2 Fold Change	Top 15 Downregulated Genes	Log_2 Fold Change
SCN5A	9.2	MCAM	-8.4
MYF6	8.0	RGS1	-7.8
BMP8B	7.8	BTG3	-7.6
HS3ST4	7.6	CXCL8	-7.5
XIST	7.3	JUN	-7.5
LAMP5	6.9	SCN3A	-6.8
DFNA5	6.9	ATF3	-6.6
DCHS2	6.4	NEU4	-6.6
EFEMP1	6.4	MYO18B	-6.5
NCKAP5	6.1	FFAR1	-6.4
HCG22	5.8	MSR1	-6.3
SPRY4	5.7	RPS4Y1	-6.1
SYNPO	5.6	FOS	-6.1
CSPG4	5.6	ARPP21	-6.1
CHST3	5.5	ZNF492	-6.1

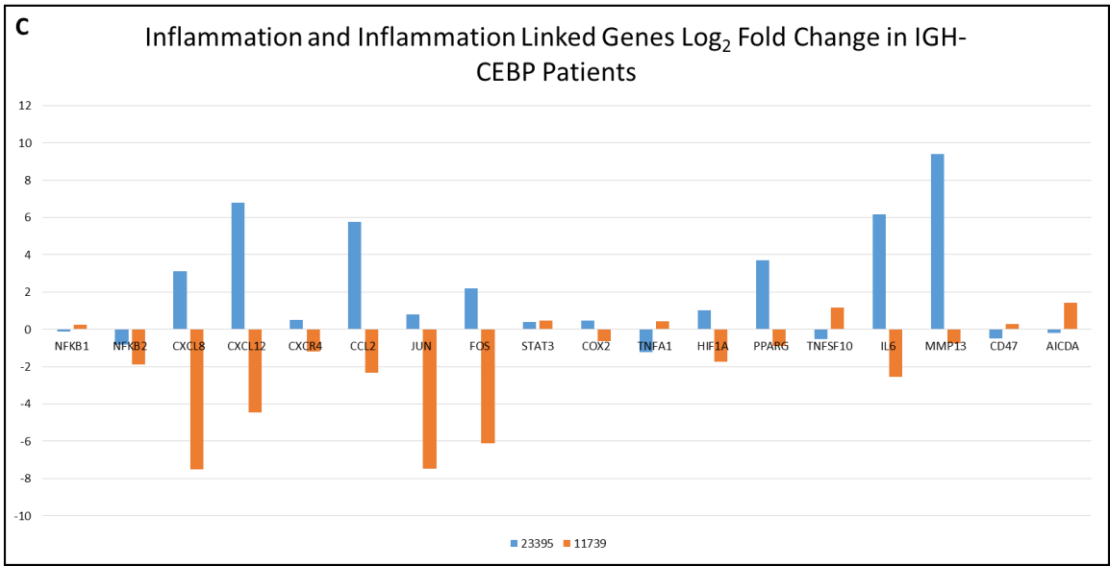


Figure 5.6. Top dysregulated genes in IGH-CEBP patients identified by RNA-seq.

A. Top and bottom 15 dysregulated genes in patient 23395, with highlighted genes involved in inflammation. B. Top and bottom 15 dysregulated genes in patient 11739, with highlighted genes involved in inflammation. Patient 23395 shows upregulation of multiple inflammatory genes, while patient 11739 shows the opposite with downregulation of inflammatory genes.

5.4 Discussion

The molecular analysis of *CEBP* deregulation was limited due to the small number of patient samples, hampered further by the necessity to analyse patient samples using different platforms and different controls. The low numbers of viable CD34+ transduced cells was also a limiting factor. Thus this chapter should be viewed as gathering pilot data for future work. Data from RNA-seq was analysed in the greatest depth, as the controls for these samples comprised a range of BCP-ALL subgroups.

IGH-CEBPB Patient 11739

Expression of CEBP family genes in patient 11739 showed upregulation of *CEBPB* across all experimental platforms. All other *CEBP* genes showed downregulation or only marginal expression. This patient also had available protein lysate to analyse expression of CEBPB protein isoforms, whose interplay determines the function of the gene. As *CEBPB* is coded from a single exon, gene transcript analysis was unable to determine which protein isoform was transcribed. Western immunoblotting showed that the major protein isoforms in this patient were LAP* and LAP activating proteins, while the dominant negative LIP protein was expressed at a lower level. This expression pattern indicated a high LAP/LIP ratio, which would result in *CEBPB* mediated transcriptional regulation. This regulation has been shown to halt cellular proliferation, as well as increase cellular differentiation, inflammation and phagocytosis. However, the opposite was observed in patient 11739, in which direct *CEBPB* target genes were inactivated, including genes involved in both haematopoiesis and inflammation.

Multiple genes involved in the immune system and inflammatory process showed down regulation despite CEBPB LAP* and LAP upregulation. The *CXCL8* gene, a potent pro-inflammatory signalling chemokine, whose expression has been shown to promote angiogenesis and metastasis in solid tumours, was significantly downregulated (Gales *et al.*, 2013). Downregulation of *ID1* and *ID2* was also observed, suggesting that a post transcriptional regulatory mechanism may inactivate *CEBPB*, as *CEBPB* has been shown to directly upregulate both of these genes (Sun, 1994; Saisanit and Sun, 1997). It is unlikely that the interplay between *CEBP* genes is playing a role in patient 11739, as all other family members were downregulated. Inconsistencies between *CEBPG* qPCR and RNA-seq data were observed. However different controls were used between experiments, which may explain the different expression levels. If qPCR *CEBPG* expression data is to be considered in place of RNA-seq data, the fold of *CEBPG* is at a similar level to *CEBPB*, this would be sufficient to block functionality of the activating CEBPB proteins. This is unlikely however as multiple *CEBPB* downstream targets are observed to be upregulated in the RNAseq data.

Genes involved in haematopoietic differentiation were analysed in order to investigate the consequences of *CEBP* overexpression on important transcription

factors. Patient 11739 showed a lymphoid primed expression profile, with erythroid driving *GATA* genes being downregulated, as well as the myeloid genes; *PU.1* and *ID* family members. The *ID* family has been hypothesised to facilitate the development of *IGH-CEBP* BCP-ALL (Nerlov, 2007), as these genes function to inactivate *TCF3* and are directly upregulated by *CEBPA* and *CEBPB* (Sun, 1994; Saisanit and Sun, 1997). It has been suggested that initial *ID1* expression is required for pro-B lymphoid differentiation, while *ID1* must then be downregulated by *TCF3* for continued B-cell development. Should an *IGH-CEBPB* translocation take place in a pro-B cell, the *CEBP* gene would promote expression of *ID1* and *ID2*, inhibiting further progression down the lymphoid pathway. This inhibition would occur due to failure to induce *TCF3,* which in turn would fail to function along with *EBF1* to induce V(D)J recombination in affected cells, as shown experimentally in murine haematopoietic cells (Sun, 1994). However, the decreased expression of *ID1* and *ID2* in patient 11739 suggests that this is not the mechanism behind leukaemic development in this patient. Lymphoid commitment genes, *IKZF1*, *PAX5,* and *EBF1,* all showed varying levels of upregulation. Previously published work showed that forced expression of *CEBPB* in precursor B cells leads to differentiation into myeloid precursors, through *CEBP* mediated downregulation of *PAX5* and synergy with endogenous *PU.1* (Xie *et al.*, 2004). Further work in pro-B *PAX5* null cells showed that forced expression of *CEBPA*, *GATA1*, *GATA2* and *GATA3* resulted in a myeloid switch (Heavey *et al.*, 2003), also seen in T-cells (Laiosa *et al.*, 2006). This phenomenon was not observed in this study, where continued expression of *PAX5* and low levels of *PU.1* and *GATA* factors likely were responsible for B lineage commitment.

The effects of *CEBP* transcription factors on the *RB*/*E2F* pathway was investigated, because it is the main mechanism of *CEBP* mediated cell cycle control. *E2F* expression in patient 11739 showed low levels of activator *E2Fs*, but increased expression of regulator *E2Fs*. *E2F7* and *E2F8* control proliferation either by repressing *E2F1* to induce apoptosis, or downregulating *MCM* and *CDC6* cell cycle control genes to promote proliferation (Lammens *et al.*, 2009). Neither *MCM* nor *CDC6* were down regulated in patient 11739 (data not shown), suggesting that the regulatory E2Fs were downregulating *E2F1*, potentially explaining the low expression level of the gene transcript. Direct *E2F* targets

included upregulation of biogenesis genes and the early G1 phase cell cycle progression gene, *CCND1,* while late cell cycle progression genes, *CCNE1* and *MYBL2,* showed low expression. The cyclin genes are upregulated at G1 and S phases of the cell cycle, providing a feedback loop to drive proliferation forward. Increased expression of *CCND1* is interesting as this gene is not normally expressed in a haematopoietic setting. Cylin genes *CCND2* and *CCND3* are typically found in lymphoid tissue (Lam *et al.*, 2000; Sicinska *et al.*, 2003), while expression of *CCND1* in the same setting is only observed in cancerous cells with mantle cell lymphoma (Rosenwald *et al.*, 2003), multiple myeloma (Bergsagel *et al.*, 2005) and hairy cell leukaemia (de Boer *et al.*, 1996) all showing de-regulation of this gene. This can be brought about by multiple methods including translocation with the *IGH* locus. Expression of *CCND1* in this novel setting is likely to disturb the balance between *CCND2* and *CCND3* leading to the de-regulation of cell cycle control. While*MYBL2* drives late cell cycle progression interacting with the downstream target, *FOXM1,* to improve binding to late cell cycle promoters (Sadasivam and DeCaprio, 2013). Other late cell cycle progression genes, such as *CCNA1*, *CCNA2* and *CCNAB1,* all showed upregulation (data not shown), suggesting that direct *E2F* cell cycle progression targets are being negatively affected, but not sufficiently to halt cell cycle progression.

Among the most differentially expressed genes in patient 11739 were the bZIP family genes, *JUN* and *FOS,* which were downregulated. As bZIP family members, both have been found to heterodimerise with members of the *CEBP* family, although no direct regulatory role has been reported between the genes. Heterodimers of these genes form the AP1 activator protein, which is involved in inducing the inflammatory cascade, proliferation and differentiation, downstream of the *NF-κB* protein complex. The significant downregulation of these genes may explain the low expression of downstream inflammation genes in this patient.

In patient 11739, the *CEBPB* gene was not exerting its expected effect across multiple genes, due to a number of potential reasons: the complex genome of the patient, comprising unknown mutations, translocations, or epigenetic modifications; post transcriptional modification of CEBPB, through ubiquitination, SUMOylation; inactivation through heterodimerisation with other bZIP family proteins. *FLT3* has been shown to inactivate *CEBPA* protein through

posttranscriptional phosphorylation (Radomska *et al.*, 2006). In previously published work, we (Akasaka *et al.*, 2007) speculated that this was one of the possible methods of *CEBPA* mediated *IGH-CEBP* leukaemogenesis, as the myeloid biasing gene was being inactivated to allow for lymphoid differentiation. Recently activating mutations of *FLT3* have been found to indirectly inactivate *CEBPB* through interaction with the microRNA, miR-155 (Salemi *et al.*, 2015). *FLT3* was upregulated in patient 11739, making this theory a possibility. What must be considered when discussing these data is that only certain *CEBPB* target genes seem to be affected including those involved in: cell cycle progression, *CCNE1*, *MYBL2* and *MYC*; cell differentiation, *ID* family; inflammation, *IL-6*, *IL-10*, *PPARG*, *CD14*, and *CXCL8*. Whether these genes were randomly or specifically downregulated remains unclear. The *AICDA* gene was upregulated in patient 11739. This observation was of interest as this gene is known to be involved in induction of mutations within the somatic hypermutation mechanism. Screening of additional patients for expression of this gene would indicate whether heightened expression is a consistent or unique feature of among *CEBP* subgroups.

IGH-CEBPD Patient 23395

Patient 23395 showed upregulation of its *IGH* partner across all platforms, although the range of expression varied between techniques. qPCR analysis also showed high expression of *CEBPG* in this patient, while RNA-seq did not. These observations suggest that RNA-seq data are likely to be the more trustworthy, as downstream targets directly affected by *CEBPD* were upregulated, discounting the possibility that *CEBPD* is being inactivated by *CEBPG* expression.

Low expression of *PU*.1*, TCF3*, *GATA2*, and *RUNX1*; low expression of the lymphoid genes, *IKZF1* and *PAX5,* and high expression of *GATA1* and the *ID* family genes were observed in this patient. These findings suggest a more myeloid or, due to expression of *GATA1*, erythroid biased profile. High transcript expression of the *ID* family was of interest, suggesting that *CEBPD* functions in upregulation of this family in a haematopoietic setting, as shown for both *CEBPB* and *CEBPA* (Sun, 1994). The expression of these genes in a lymphoid committed leukaemia could give credence to the theory that *ID1* and *ID2* are able to mediate

a differentiation block of the lymphoid lineage in the pro-B stage, with *TCF3* inhibited by the *ID* genes being unable to induce expression of *EBF1*, *PAX5,* and *IKZF1,* to facilitate expansion of these cells.

Expression of downstream *CEBPD* target genes showed the expected expression, with little variation observed. Patient 23395 broadly mirrored previous findings in expression of *RB/E2F,* with *MYC* downregulated, *RB1* and several kinase inhibitors upregulated. Despite these results, *E2F* expression was high. Interestingly the regulator gene, *E2F5,* was heavily downregulated in this patient. *E2F5* functions by binding to *RBL1* and *RBL2,* and mediating cell cycle arrest at the G1 phase (Gaubatz *et al.*, 2000). The loss of *E2F5* may contribute to loss of cell cycle control, resulting in increased proliferation. This *E2F* expression profile was likely influenced by the upregulation of *E2F* targets involved in cell cycle progression, biogenesis and apoptosis. Downstream *E2F* targets also showed higher overall expression. Cell cycle genes, *CCND1*, *CCNE1* and *MYBL2,* were all upregulated, suggesting continued proliferation. The overall profile was indicative of oncogenic deregulation with cyclin kinase inhibitors broadly downregulated and cyclins and dependent kinases upregulated.

Patient 23395 showed upregulation of multiple inflammatory and environmental cytokines and chemokines linked with collagen, fibronectin and cell extracellular matrix interaction (Supplementary Table 7.32). The majority of such genes have been implicated in metastasis and angiogenesis in a variety of solid tumours. In leukaemia such genes have been increasingly identified as being important in BM niche microenvironment, and proliferation of HSCs and potential leukaemic cells which use this niche to escape chemotherapy induced apoptosis (Chiarini *et al.*, 2016). Additionally, several of these genes are involved in the *PI3K/AKT* pathway, this is likely through fibronectin and collagen mediated activation of the *PTK2* (*FAK*) tyrosine kinase which in turn activates downstream *PI3K/AKT* genes. Upregulation of this pathway would potentially lead to increased proliferation and survival of blasts (You *et al.*, 2015). Several of the down regulated genes observed in this patient are known tumour suppressors in solid tumours, linked to prevention of metastasis and angiogenesis, while three zinc finger proteins are potentially involved in gene expression control. Such a profile viewed in solid tumours would be indicative of a highly metastatic cancer. However, in the

context of this leukaemia, it is more likely the consequence of *CEBPD* mediated inflammatory pathway upregulation.

Inflammation and leukaemia

In depth studies have been performed investigating the link between expression of inflammatory chemokines and cytokines in leukaemia. In AML, specific expression patterns of CCL chemokines and interleukins were resulted in higher remission and survival rates (Kornblau *et al.*, 2010). However, this interplay was highly complex with multiple factors influencing expression levels.

In this project, key inflammatory regulators showed limited upregulation in both *IGH-CEBP* patients. *NFκB* factors were almost universally downregulated in both RNA-seq patients, a surprise as *NFκB* is a downstream target of multiple genes found to be upregulated in the *IGH-CEBP* patients. Other inflammatory regulators included the *STAT3* transcripts which were slightly upregulated in both patients, while *HIF-1α* showed upregulation in *CEBPD* patient 23395 and down regulation in patient 11739, showing no single pathway was involved in inflammatory regulation in these patients. This supports the theory that *CEBPB* is not upregulating inflammatory signalling in patient 11739, and signalling that is taking place in patient 23395 appears to be independent of *NFκB*.

The genes expressed however were functionally relevant. The chemokine, *CXCL12,* is well characterised in cancer and leukaemia. Expression of this gene in AML cells, along with *CXCR4,* was shown to facilitate binding to the bone marrow stroma, with associated improvements in survival, proliferation and chemotherapeutic resistance of AML cells (Koblas *et al.*, 2007). Interestingly, a *CXCL12* analogue, Plerixafor, has been used to mobilise AML blasts into peripheral circulation, improving their removal when used in combination with chemotherapy in mice. This analogue was also used in phase I/II clinical trials and was found to improve remission rates (Uy *et al.*, 2012). Interestingly patient 23395 showed a high levels of *CXCL12* expression.

Activation of multiple chemo- and cytokines has been shown to result in increased cell survival. However, *CEBPB* patient 11739 showed no upregulation of

inflammatory factors, despite its well documented action as a mediator of inflammation.

Investigating the role of inflammation in genomic instability is of interest. Overexpression of the *AICDA* gene which is upregulated by multiple pro-inflammatory factors (Mechtcheriakova *et al.*, 2012) in CD34+ cells would be of interest, as this gene induces multiple double stranded DNA breaks. The function of AICDA in CD34+ cells could be tested through long term cell culture and subsequent analysis by karyotype, FISH with whole chromosomal paints to identify translocation events.

Variation Between *IGH-CEBP* Patient RNA-seq Data

From the RNA-seq data, both *IGH-CEBPB* patient 11739 and *IGH-CEBPD* patient 23395 showed decreased *CEBPA* expression, which is interesting in the context of the disease. Along with *CEBPB*, *CEBPA* is the most potent early myeloid commitment gene in the *CEBP* family. Both genes have been shown to function in place of each other, providing a broad range of redundancy. The down regulation of *CEBPA* in both patients, after normalising against other BCP-ALL controls, may indicate that the overexpressed *CEBP* genes may be responsible for down regulation of the *CEBPA* gene indirectly. Although *CEBPB* has been observed to directly upregulate *CEBPA* expression, it has only been reported in adipocyte differentiation (Tang *et al.*, 2003). While *CEBPA* mutations are common in AML, none have been observed in BCP-ALL, and no CNA involving this gene emerged from SNP arrays, discounting CNA driven change of expression of this gene. All other myeloid biasing *CEBP* genes were either downregulated or showed low expression, as expected in the context of this disease.

It seems that by the time these leukaemic precursors enter the pre- or pro-B stage they become locked into the lymphoid lineage. The effect of *CEBPs* as myeloid differentiators is not sufficient to transform the cells. Interestingly, in a RNA-seq study of eight BCP-ALL cases, which switched to monocytic lineage, *CEBPD* was among the most differentially expressed genes, whose expression was initially high and continued to increase as the switch occurred. *CEBPA* was also overexpressed (Fier *et al.*, 2014). These patients did not express known genomic

aberrations by MLPA or FISH. It would be of interest to know whether these patients had been investigated for the presence of *IGH-CEBP* translocations. It would also be of interest to compare my cohort with this lineage switching cohort, in an attempt to determine what is inducing one group of patients to switch lineage and the other to maintain lineage, despite seemingly strong pressure to differentiate.

The main difference between patients 11739 and 23395 from RNA-seq data was the variation in expression of direct *CEBP* targets. Overall *CEBPD* appeared to function as expected. However, *CEBPB* was different. Cell cycle progression genes showed low expression, while the expression profiles of other direct targets, such as the *ID* genes, varied markedly between the two patients. Firstly, these observations suggested that *CEBPD* functions in the same manner as *CEBPB* in working to upregulate expression of *ID1*. If *ID1* deregulation was an influencing factor in patient 23395, it would suggest that different *IGH-CEBP* leukaemias developed through different *CEBP* mediated pathways. The consistent difference in *ID* expression in both patients and committed lineage cells is worthy of further study It may be that *FLT3* is exerting a posttranscriptional effect, but *FLT3* expression was upregulated in both patients to the same degree, seemingly with no negative effects in patient 23395. This may be due to *FLT3* only exerting post transcriptional regulation on *CEBPA* and *CEBPB.* However, there are no reports to date of *CEBPD – FLT3* interactions.

High expression of the *MYF6* gene was also observed in both samples. This gene functions in myocyte differentiation with no direct link to the *CEBP* genes, although interaction with *EP300*, a *CEBPD* target gene, and *TCF3* have been reported. Further work to determine the role of this gene in BCP-ALL may be worthwhile.

It will be necessary to unravel the potential mechanisms behind *CEBP* function in these patients. Application of various protein analysis techniques may shed light on post transcriptional variation of *CEBPB*. Interactions with FLT3 and CEBPD could be analysed, to determine whether the two proteins interact. The phosphorylation status of *CEBPB* should be investigated. Chromatin immunoprecipitation assays could be used to investigate *CEBPB* interaction with target genes, such as *ID1*, *ID2*, *CXCL8,* and other differentially expressed targets.

These approaches would provide invaluable data, the limiting factor being the amount of protein lysate available from patient samples.

The impact of the *ID* family in *IGH-CEBP* requires further study. All four *ID* genes have been implicated in multiple forms of cancer including ALL (Lasorella *et al.*, 2014). If more patients are identified with forced expression of the *ID* genes, this would identify them as suitable targets to facilitate differentiation of leukaemic blasts. The *ID1* gene has been targeted in metastatic breast cancer cells after transduction with *ID1* antisense knock out (Fong *et al.*, 2003).

The analysis of these patients has highlighted the heterogeneity of transcript expression between different *IGH-CEBP* patients. Due to the function of the *CEBP* genes and their interactive nature with members of the bZIP family, further work, particularly protein analysis, needs to be performed to determine the function of these genes in BCP-ALL.

Chapter 6 Discussion

With continually improving genetic technologies the volume of genomic data is increasing. The challenge in BCP-ALL research is translation of these data into improved risk stratification and increased patient survival through identification of novel treatment options. The classification of BCP-ALL into subgroups provides the backbone for treatment selection. This system has traditionally used established cytogenetic alterations in combination with demographic and clinical data, such as age, sex and WBC, to classify patients, which has translated into improved survival rates (Harrison, 2011).

Despite this success much remains to be improved; older patients have a far lower five year event free survival rate, while relapse occurs across all age groups with high mortality (Moorman *et al.*, 2014). Another challenge is to improve survival of patients with no clear recurring genetic alterations, who may be difficult to treat (Roberts *et al.*, 2012). The use of WGS and WES has not only increased the detection rate of novel genetic lesions but, along with RNA sequencing, has helped to identify the molecular mechanisms deregulated in some of these patients, identifying new subgroups. These discoveries are providing options for more targeted treatments, contributing to decreased chemotherapy related toxicity and improved survival.

Originally included in the unclassified B-other subgroup, *IGH-CEBP* patients are predominantly teenagers and young adults, who are already associated with inferior outcome. Since the identification of this subtype of BCP-ALL (Chapiro *et al.*, 2006; Akasaka *et al.*, 2007), few studies have characterised the group further. Many have simply collected additional patients, comprising small numbers for analysis of clinical and cytogenetic data (Chapiro *et al.*, 2006; Akasaka *et al.*, 2007; Lundin *et al.*, 2009; Messinger *et al.*, 2012; Chapiro *et al.*, 2013; Russell *et al.*, 2014).

The aim of this project was to characterise *IGH-CEBP* patients using detailed genomic and functional techniques, and to make full use of the UK clinical trials database resource to create a cohort large enough to identify patterns within

clinical and cytogenetic data, which was not previously possible due to small patient numbers.

Genetic Characterisation

In this project genetic characterisation of the subgroup has revealed a number of interesting new findings. A total of 33 *IGH-CEBP* patients were identified, successfully generating the largest known cohort of this subgroup, which was found to comprise 19% of the *IGH* cohort, higher than the 11% previously reported (Russell *et al.*, 2014), and 1% of ALL as a whole. Clinical comparisons of this group with the *IGH* subgroup as a whole showed multiple similarities, with the exception of relapse rate, which was found to be twice as high in the *IGH-CEBP* patients. Despite these findings there were not sufficient patients to analyse the *CEBPE* and *CEBPG* subgroups, highlighting the need for an even larger cohort. There is an ongoing recruitment drive in the UK which will ensure the cohort will grow with time, however other options are also available. Forming collaborative links with other groups who have access to *IGH-CEBP* patients would be extremely useful. The Children's Oncology Group likely have multiple *IGH-CEBP* patients to add to this cohort, for example the Messinger cohort mentioned previously (Messinger *et al.*, 2012). Additional viable samples gained can now be expanded more reliably *in vivo*, through use of Busulfan conditioning in NSG mice, giving more material for downstream applications.

Despite an ongoing need for more patients, this *IGH-CEBP* cohort identified for the first time, novel differences between individual *IGH-CEBP* partners. *IGH-CEBPB* patients were significantly older, had a higher mortality rate and higher number of CNAs. Conversely, *CEBPD* patients had low WBC, the lowest number of CNAs, and included the youngest patients, which remained significant whether or not DS patients were included. The incidence of CNAs between different *CEBP* subgroups was similar to those observed between individual BCP-ALL subtypes: *CEBPD* patients showed low levels of CNAs similar to the HeH subgroup, while *CEBPB* patients showed a profile more similar to iAMP21 or *BCR-ABL1* patients (Schwab *et al.*, 2013). The most common alterations varied from the incidence typical of paediatric BCP-ALL. For example, the *IGH-CEBP* subgroup showed deletions of *CDKN2A*/*B* (25%),

IKZF1 (21%), *PAX5* (14%) and *ETV6* (11%) at different levels from other paediatric B-ALL subgroups at 27%, 13%, 19% and 22%, respectively (Harrison, 2013). The higher incidence of *IKZF1* deletions in *IGH-CEBP* patients was interesting due to its association with the *BCR-ABL1* and *BCR-ABL1*-like subgroups and link to a poor outcome (Den Boer *et al.*, 2009; Mullighan *et al.*, 2009b; Harvey *et al.*, 2010b). Another interesting trend relating to *IKZF1* in this cohort was its restricted occurrence to the *CEBPB* and *CEBPD* subgroups, which was a novel and statistically significant finding. Additionally, SNP6.0 arrays identified one novel aberration not previously observed in BCP-ALL: a focal deletion involving the *ABL2* gene, potentially leading to constitutive activation of this tyrosine kinase.

ABL2 and Classification of the *IGH-CEBP* Subgroup

A novel finding in this project was the discovery of four *ABL2* deleted patients within three *CEBP* subgroups (*CEBPA*, *CEBPB* and *CEBPD*). The outcome for these patients was poor, with two patients relapsing and one dying. The deceased patient however also harboured a *BCR-ABL1* translocation, which likely contributed to the poor outcome. When compared to the *IGH* group as a whole, *ABL2* deleted patients were slightly older with a similar WBC and mortality rate, but an increased incidence of relapse, and *IKZF1* deletions (Table 6.1).

***IGH-CEBP* Patients**	**No. of Patients**	**Mean Age**	**WBC x 10^9L**	**Deaths (%)**	**Relapse (%)**	**IKZF1 del**
ABL2 Deleted	4	29	36	1 (25%)	2 (50%)	2 (50%)
ABL2 Wildtype	29	22	33	8 (28%)	3 (10%)	4 (14%)

Table 6.1. Comparison of ABL2 deleted and wildtype patients in the IGH-CEBP cohort

TKIs have been used successfully in treatment of leukaemia and cancer (Greuber *et al.*, 2013), thus the presence of *ABL2* deletions highlights a potential novel treatment option for the *IGH-CEBP* subgroup. Cells from *BCR-ABL1*-like patients with *ABL2* translocations have been treated *in vitro* with the *ABL1* TKI, dasatinib, with successful response (Roberts *et al.*, 2014a). Currently

there are no *ABL2* specific TKIs, likely due to the high efficiency of those TKIs initially developed for *ABL1* inhibition (Greuber *et al.*, 2013). However, before these findings can be used clinically, additional work needs to be performed, firstly to identify the incidence of the deletion in the subgroup as a whole, and secondly to confirm the hypothesis that focal deletions are resulting in *ABL2* deregulation.

In order to identify new patients with *ABL2* deletions, a FISH screening approach could be confidently applied for larger deletions. An alternative option for those patients with smaller exon 2 deletions and/or without fixed cells is the continued development of copy number qPCR. Should additional *ABL2* deletions be found in the cohort, it would be of interest to expand screening to other BCP-ALL subgroups. An efficient high throughput method to perform such a screen would be the development of a specific MLPA kit, covering exons of interest, although this would method would not be sufficient for identification of clones below 20% incidence (Schwab *et al.*, 2010a). Targeted NGS provides another screening option, which has been shown to detect deletions at levels as low as 1% in targeted regions (Grotta *et al.*, 2015). Alternatively high density SNP arrays have been shown to efficiently identify CNA; decreasing costs and lower DNA requirements make this a valid option. Such screens will be most useful if the *ABL2* deletion is found to be functionally relevant, which could be tested by a range of methods. For example, the CRISPR-Cas9 system could be used. Guide RNAs could be selected to delete selected exons from the *ABL2* gene, recreating both the exon 2 and exons 2-7 deletions. This approach could be used on CD34+ cells, in order to investigate the consequences on a blank haematopoietic background, in a BCP-ALL cell line such as REH, or *IGH-CEBP* patient material without the *ABL2* deletion. A disadvantage of using leukaemia cell lines would be the confounding effect of additional genetic lesions. This disadvantage could be mitigated through the use of the same cancer cell lines as control samples when performing RNA-seq or gene expression array analysis, serving to remove the effects of aberrant expression resulting from other genetic lesions. The advantages of using cancer cell lines are simple cell culture protocols, more rapid expansion of edited cells, and potentially unlimited material. Clonogenic assays could also be performed to monitor effects on proliferation and survival. The initial aim of such an experiment would be to

confirm constitutive activation of the gene, and later, the consequences of oncogenic deregulation. RNA-seq / gene expression analysis in combination with western blotting would prove interaction of the modified ABL2 protein with direct downstream targets, such as *CRK*, *RIN1* or *RAC1* (Cao *et al.*, 2008; Li and Pendergast, 2011). Deregulating *ABL2* translocations in the *BCR-ABL*-like subgroup have been observed to lead to phosphorylation of STAT5 (Roberts *et al.*, 2014a), something that could be tested with Western Blotting and a phosphor-specific antibody. Such studies would help to confirm the functional consequences of the *ABL2* deletion.

If *ABL2* deletions lead to deregulation of the tyrosine kinase, then these patients may belong to the *BCR-ABL1*-like subgroup. Clinically there are not many similarities between *IGH-CEBP* and *BCR-ABL1*-like patients to determine this relationship. Examination of data published by others (Den Boer *et al.*, 2009; Roberts *et al.*, 2014b), showed that the median age of *BCR-ABL1*-like patients was younger than the *IGH-CEBP* subgroup. WBC was similar to the Roberts study, but not the Den Boer study, which was higher, while *IKZF1* deletions occurred at a lower incidence in the *IGH-CEBP* group (21%) compared to the *BCR-ABL1*-like (40%). However recurrent *IKZF1* deletions and potential tyrosine kinase deregulation suggests that *IGH-CEBP* patients may benefit from treatment protocols similar to those of *BCR-ABL1*-like patients with *ABL* class fusions.

An alternative method of classifying the *IGH-CEBP* subgroup could be through a more recent classification system, which uses bioinformatics modelling to combine cytogenetic and genomic data to create a new risk stratification model for treatment of BCP-ALL patients (Moorman *et al.*, 2014). This work has taken MLPA data on eight genes/regions commonly altered in BCP-ALL and integrated these data on a total of 1551 patients within known cytogenetic risk groups to create new genetic good and a poor risk groups, which are independent of other risk factors. Prevalence of CNAs was important in determining outcome, as incidence of three or more CNAs was classified as poor risk. Those genes which were affected were also important, as deletions of *IKZF1*, *PAR1*, *EBF1* or *RB1* were immediately classified as poor risk. Any cytogenetic profile which had not previously been classified as good risk was considered to be poor risk. Using this classification system for the *IGH-CEBP*

subgroup would result in the *CEBPB* and *CEBPD* patients classified as poor risk, due to the high incidence of *IKZF1* deletions, as well as several other *CEBPs* due to the high incidence of CNAs, while other patients would be classified as good risk as no CNAs were observed. Application of this classification system would divide the *CEBP* subgroup, resulting in higher intensity treatment for the majority of the *CEBPB* subgroup, where the majority of deaths occurred.

Clonal Evolution in the *IGH-CEBP* Subgroup

Another issue with the *IGH-CEBP* subgroup was the high rate of relapse in comparison to the *IGH* group as a whole. Relapse in ALL has several origins, with smaller clones present at diagnosis defined as one of the sources (Mullighan *et al.*, 2008b). These clones evade initial treatment and expand upon elimination of a dominant clone. One of the methods considered to decrease relapse rates is through targeting of primary and secondary abnormalities at the start of treatment, destroying smaller clones before they can expand to cause a relapse.

In this study, we discovered that *IGH-CEBP* translocations could occur as either primary or secondary events. This knowledge could be an important factor when deciding upon treatment strategies. Previously published work on clonal evolution in BCP-ALL also showed that *IGH* translocations can occur as secondary events in patients with another known established genetic abnormality, such as, *ETV6-RUNX1*, *BCR-ABL1*, iAMP21 and *KMT2A* translocations. This work included two patients from this cohort, who displayed *IGH* translocations occurring both as primary and secondary events. HeH patient 7143 (*IGH-CEBPA*) had an *IGH* translocation in the earliest known clone, in combination with additional chromosomes 4, and two copies of chromosome 21 and a *BCR-ABL1* patient 10859 (*IGH-CEBPB*), where the *IGH* translocation was secondary to *BCR-ABL1* (Jeffries *et al.*, 2014). Work performed in this project also shown variation in the occurrence of *IGH* translocations in relation to CNAs. *IGH-CEBP* translocations occurred as both primary and secondary events, with the latter slightly more common. These findings indicate that *IGH-CEBP* translocations can occur as both primary and

secondary alterations, depending upon the patient and the associating abnormalities. As the study was restricted to abnormalities which could be detected by FISH, the presence of other genetic lesions occurring at different time points cannot be ruled out. The involvement of the *ABL2* deletion was not investigated as the CNA was identified after the clonal evolution study had been performed. However it would provide interesting future work. Additionally, hypodiploid patient, 23168 (*IGH-CEBPD*), could also be analysed to determine order of occurrence.

Functional Characterisation

While functional characterisation of the *IGH-CEBP* disease proved difficult, in retrospect the results found in this project support other recently published modelling studies. Modelling *CEBPD* deregulation in CD34+ cells showed that *CEBPD* and *IK6* overexpression in tandem were insufficient for proliferative deregulation and consistent lymphoid differentiation. *CEBPD* exerted cell cycle arrest, as shown in other cell types, with reduced proliferation of transduced cells.

A study (Theocharides *et al.*, 2015) modelled BCP-ALL by expressing *BCR-ABL1* and *IK6* in CD34+ cells. Myeloid expansion was observed in double transduced cells, with cells expressing IK6 only exhibiting no oncogenic traits when compared to non-transduced CD34+ cells. Myeloid differentiation was an unexpected and novel finding in this study, particularly in cells expressing a combination of genetic alterations commonly found in BCP-ALL. It appears that IK6 expression is insufficient to push lymphoid differentiation, which may be a reason for why no visible lymphoid commitment was seen in transduced cells in my study. A more recent study modelled a range of *IKZF1* deregulation (including IK6 expression) with *CDKN2A* loss in *BCR-ABL1* cells, which resulted in development of BCP-ALL (Churchman *et al.*, 2015). Both studies showed that the oncogenic hits selected for modelling were crucially important for the development of BCP-ALL. This project indicated that *CEBPD* upregulation cannot be considered as an oncogenic hit, rather as a tumour suppressor, considering the strong anti-proliferative action observed here. As a result, *CEBPD* expressing populations were short lived, both *in vivo* and *in vitro*, and

the cells sorted by FACS were not of sufficient quantity or quality to extract the amount of RNA needed for gene expression analysis.

Future efforts in modelling of this disease would benefit from making use of WGS and WES data to select the most appropriate genetic alterations to express dependent upon those molecular pathways affected in patients. *CDKN2A* would have been a good choice as an important cell cycle control gene, the removal of this gene may have offset the cell cycle control properties of *CEBPD* (O'Rourke *et al.*, 1999).

Potential Oncogenic Mechanisms of *IGH-CEBP* in BCP-ALL

RNA-seq data analysis showed large differences in transcript expression between two *IGH-CEBP* patients. Doubtlessly the characteristic natural variation of any leukaemia will be responsible in part for differences in transcription profiles. However, despite different *CEBP* genes being overexpressed, it was surprising to see the extent of the differences in downstream *CEBP* targets between the two patients. *CEBPB* did not appear to function as expected: direct *CEBPB* targets were not affected by high *CEBPB* upregulation, with low expression of inflammation factors and down regulation of the *ID* family being particularly surprising. Conversely high expression of the *ID* family in the *IGH-CEBPD* patient suggested a direct interaction of *CEBPD* with the *ID* family, as has been reported previously for *CEBPB* (Saisanit and Sun, 1997) and *CEBPA* (Wagner *et al.*, 2006).

The downregulation of *CEBPB* targets is not the result of a dominant negative LIP protein. The western blot performed in this project showed strong expression of the LAP* and LAP isoforms, indicating an active role for *CEBPB* in this patient. This observation suggested that the gene expression patterns observed using RNA-seq are the result of posttranscriptional controls in patient 11739. Only one other study has published data on CEBP protein expression, a *IGH-CEBPA* patient who expressed the active p42 isoform of this gene (Chapiro *et al.*, 2006). These findings, now in two *IGH-CEBP* patients, indicate that the active protein isoforms of *CEBP* are overexpressed and are contributing to leukaemogenesis by activation, rather than inactivation of *CEBP* function.

It is unclear how increased expression of the *CEBP* genes may influence leukaemogenesis. However, the fact that five members of the same gene family are *IGH* partners suggests a common deregulatory mechanism between *CEBP* genes. Yet what is being observed from the RNA-seq data in this project suggests that there may be two distinct mechanisms behind *CEBP* mediated oncogenesis. Logically, inactivation is the most likely mechanism behind *CEBP* oncogenesis, as this would result in the removal of *CEBP* mediated cell cycle control and differentiation. This mechanism may be supported by the existence of the single *IGH-CEBPG* patient, as high expression of the dominant negative CEBPG protein could only function by inactivating other CEBP proteins. Inactivation of other *CEBPs* by *CEBPG* is something observed in a range of cancer types (Section 1.5.7.1).

The second potential mechanism of leukaemogenesis is through *CEBP* upregulation. One common function of multiple CEBP genes is regulation of inflammation and the immune system, typically triggered by infection (Tsukada *et al.*, 2011). Upregulation of multiple inflammatory genes has been found to contribute to multiple cancers, including leukaemia, and is a hypothesis that is explored further below.

Inflammation as an Oncogenic Hit in the *IGH-CEBP* Subgroup

High variation in expression of multiple inflammatory factors was found in the RNA-seq data analysed in this project. Inflammation genes were among the top 15 most highly expressed genes in patient 23395, among them were inflammatory factors, such as *CXCL12*, *IL-6*, *CCL2*, and *MMP13,* which are indicative of leukaemia with a strongly upregulated inflammatory phenotype (Giles *et al.*, 2014). The action of the majority of these genes provide an oncogenic effect through increased migration of leukaemic cells to the bone marrow microenvironment and adherence to this niche, providing shelter from chemotherapeutic agents. *CXCL12* and its receptor, *CXCR4*, are important in the maintenance and adherence of the haematopoietic stem cell pool (Tzeng *et al.*, 2011), and the migration of leukaemic cells to the BM through chemotaxis via *p38MAPK* and *PI3K* activation (Wang *et al.*, 2000; Juarez *et al.*, 2009; Sahin and Buitenhuis, 2012). Specifically, high expression of *CXCL12* has resulted in

increased levels of malignancy in both solid tumours (de Oliveira Cavassin *et al.*, 2004) and leukaemia (de Oliveira *et al.*, 2007). *CCL2* has also been shown to increase adhesion of ALL cells to BM stroma resulting in increased survival and proliferation of ALL cells (de Vasconcellos *et al.*, 2011).

Such a phenotype would result in improved cell survival, through resistance to apoptotic signals and chemotherapy, and increased metastasis, due to matrix metalloprotease expression and *IL-6* expression. This upregulation of inflammatory factors may be the oncogenic push conferred by the *CEBP* family in *IGH-CEBP* BCP-ALL, although it appears to be independent of *NF-κB*. The high expression of such factors may play a role in the increased relapse rate of the *IGH-CEBP* subgroup in comparison to other *IGH* patients (15% vs 7%) (Figure 6.1), but they may also provide interesting targets for therapy with the aim of decreasing relapse rates, which has been observed in AML with the use of the *CXCL12* analogue Plerixafor, which binds to the *CXCR4* receptor and prevents downstream signalling (Uy *et al.*, 2012).

Ultimately, upregulation of *CEBP* genes appears to lead to oncogenesis, either through upregulation of inflammatory factors or through deregulation of cell cycle and differentiation control.

What are the Posttranscriptional Pressures and Protein Interactions of the *CEBP* Genes?

These RNA-seq finding raise an important question, what are the post transcriptional pressures on the *CEBP* genes? The seemingly erratic function of *CEBPB* is likely due to post transcriptional regulation and not expression of the dominant negative LIP isoform. This observation suggests inactivation through other means, such as *FLT3* mediated phosphorylation, or ubiquitination / SUMOylation. SUMOylation can likely be discounted as the source of regulation because all SUMO gene transcripts are downregulated in patient 11739 (data not shown). *FLT3* has been shown to force expression of the *CEBPB* dominant negative LIP protein isoform, which is achieved by dephosphorylation of *PKR*, a phosphorylating agent which inactivates elf-2α and removes its inhibition of *elf-4E*, the translation factor responsible for expression of fully formed CEBPB and CEBPA (Calkhoven *et al.*, 2000; Haas *et al.*, 2010). Alternatively, it would be

interesting to investigate the binding partners for the CEBP genes, which could be achieved using chromatin immunoprecipitation assays to determine protein interaction with DNA. Targets such as the *ID* family and the multiple inflammatory factors could be analysed in patients 11739 and 23395, to identify whether *CEBP* binding was taking place at the desired target gene. Standard immunoblotting could be used to determine CEBP dimers and to investigate their interplay with other bZIP genes. Genes such as *FOS*, *JUN*, and *CREB*/*ATF* are *CEBP* partners, with potential to be functionally inactivated due to high expression of *CEBP* genes leading to sequestering of other functional bZIP partners in *IGH-CEBP* BCP-ALL cells.

Lineage commitment in the *IGH-CEBP* Subgroup

A question which was initially interesting in this study was attempting to identify why myeloid differentiation was blocked in *IGH-CEBP* patients, particularly with the myeloid committing role of the *CEBP* family. It is most likely that *IGH-CEBP* translocations occur later in leukaemia development, when the affected blast is already committed to the lymphoid lineage. In the two patients, *PU.1* expression was low while *PAX5* expression was normal. *CEBPs* initiate myeloid differentiation in lymphoid committed cells by working synergistically with *PU*.*1* and down regulating *PAX5*. It is surprising that with this powerful transformative ability, there is no recorded incidence of myeloid leukaemia with an *IGH-CEBP* translocation. This may be due to the natural function of the *CEBP* family; translocations occurring prior to B-lymphoid commitment undergo *RB*/*E2F* mediated cell cycle arrest and differentiate into myeloid cells, followed by cell death, preventing the expansion of a potentially oncogenic clone. The importance of *PAX5* in maintenance of B-cell identity in *IGH-CEBP* BCP-ALL could be investigated by culturing primary patient material and knocking out *PAX5* using CRISPR-Cas9 technology. The use of this technology with primary cells has been successfully tested and performed on T-cells and CD34+ cells by using modified guide RNAs (Hendel *et al.*, 2015). Modified cell lineage could be tracked using cell surface marker expression with flow cytometry.

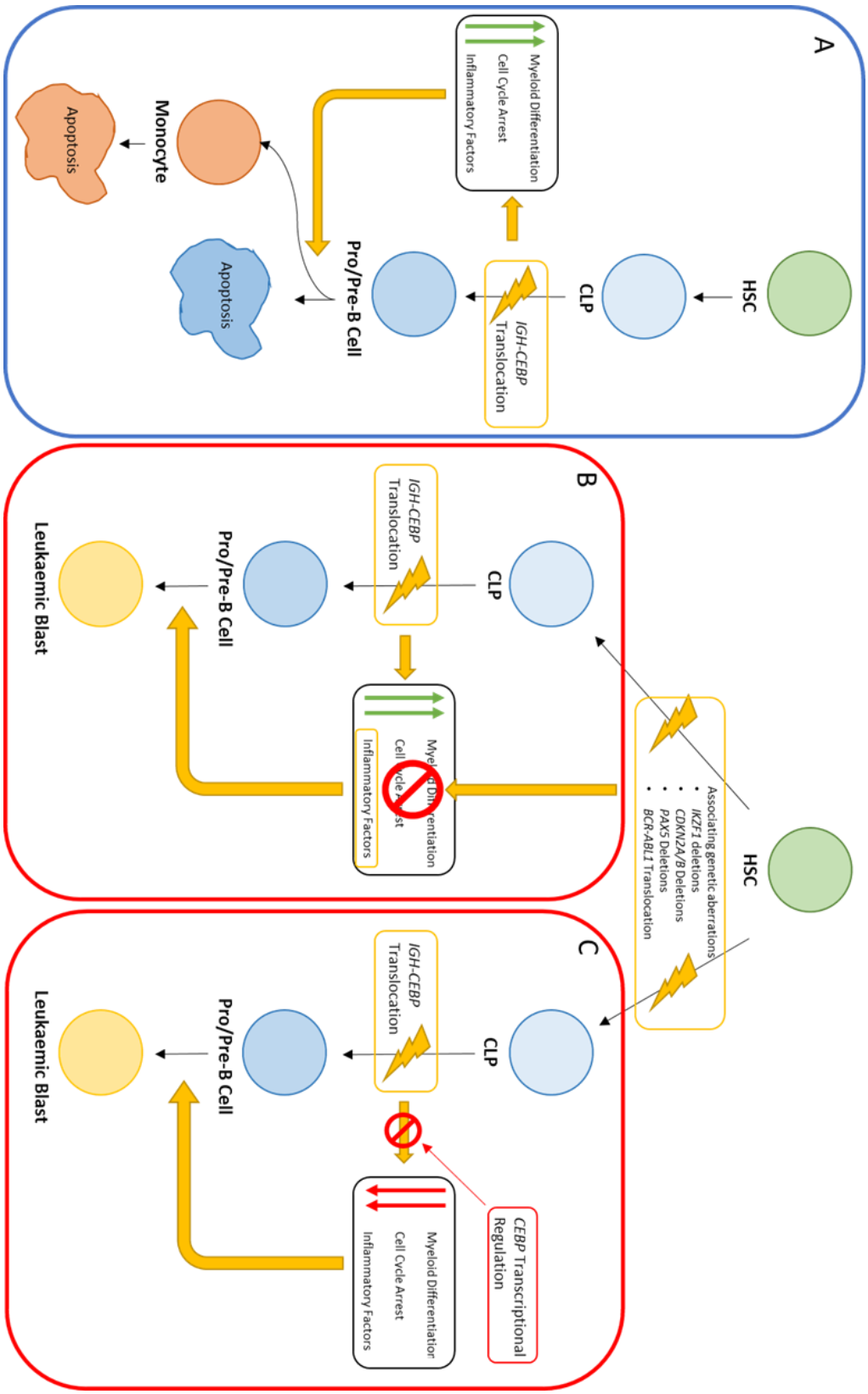


Figure 6.1. The potential mechanisms of IGH-CEBP mediated oncogenesis.

A. An IGH-CEBP translocation occurring in a CLP cell without prior genetic aberrations, leading to myeloid differentiation and apoptosis as a result of the upregulation of myeloid committing and cell cycle arrest genes. B. IGH-CEBP leukaemogenesis may develop as follows; several priming genetic alterations commit the recipient cell to the lymphoid

lineage and induce proliferative deregulation, following these events the IGH-CEBP translocation occurs. CEBP functions controlling myeloid differentiation and cell cycle arrest is blunted due to prior genetic alterations, however CEBP mediated upregulation of inflammatory factors is not affected, and an oncogenic push is exerted through upregulation of inflammatory genes leading to increased DNA damage and drug resistance. This theory fits the RNA-seq data observed in patient 23395. C. A second potential mechanism of IGH-CEBP leukaemogenesis may occur through the target cell receiving priming genetic insults, again committing the cell to the lymphoid lineage, on this occasion however CEBP function is hindered post IGH-CEBP translocation through an unknown mechanism, leading to a loss CEBP function, including control of cell cycle control which results in oncogenesis. This theory is reflected in the RNA-seq data observed for patient 11739, and the mechanism of oncogenesis observed in CEBPG deregulated cancers.

Therapeutic Targets for *IGH-CEBP* Patients

Other than the potential application of TKIs for *IGH-CEBP* patients with *ABL2* deletions, there is potential for therapeutic targets focused on inhibition of the *CEBP* genes themselves. Berberine, a well-studied plant alkaloid, indirectly downregulates the CEBP genes through upregulation of *CEBP* family repressors; *CEBPZ* and *DEC1* / *DEC2*, which in turn inhibit the function of *CEBPA* through interruption of DNA binding. This experiment was performed on 3T3-L1 preadipocytes and found to inhibit adipogenesis (Pham *et al.*, 2011). Berberine controls cell cycle progression at multiple stages, one by inhibiting expression of *CCND1, CCND2, and CCNE1* and halting cell cycle progression at the G1 phase (Mantena *et al.*, 2006), and downregulating *CDK1* and *CCNB1* expression, resulting in cell cycle arrest at the G2/M phase (Lin *et al.*, 2006). This therapeutic agent may be particularly attractive for use with certain *IGH-CEBP* BCP-ALLs, as it hits multiple inflammatory targets such as were observed to be upregulated in patient 23395; *IL-6* (Chen *et al.*, 2008a), *COX2* (Kuo *et al.*, 2005) and *HIF-1α* (Lin *et al.*, 2004). It would be of interest to investigate whether this therapeutic compound could function in inhibition of other *CEBPs,* which would be feasible as *DEC1* and *DEC2* have been shown to lead to *CEBPB* inhibition in adipogenesis (Gulbagci *et al.*, 2009).

Betulinic acid is a well-established therapeutic agent with antimalarial, anti-inflammatory and anticancer properties, which was found to inhibit *CEBPD*, *CEBPE* and *CEBPG* to a similar degree, while showing little effect on *CEBPB* (Hollis *et al.*, 2012). The drug operates by inhibiting binding of CEBP dimers to target DNA, in 3T3-L1 preadipocytes, resulting in inhibited adipocyte differentiation in a dose dependent manner, suggesting inhibition of *CEBPA* and *CEBPD*. Betulinic acid was reported to lead to cell death of K562 myelogenous

leukaemia cells through upregulation of BAX and caspase-3 (Wu *et al.*, 2010). It was also found to induce apoptosis in 65% of primary paediatric acute leukaemia cells and leukaemia cell lines, through induction of caspases and release of cytochrome c and *Smac*, mitochondrial apoptosis induction genes (Ehrhardt *et al.*, 2004). In combination with RNA-seq data, both of these compounds could be tested on primary patient material to investigate their efficacy on *IGH-CEBP* BCP-ALL cells. As *CEBPs* are expressed for specific functions at specific time points, their downregulation in BCP-ALL patients could potentially avoid a number of side effects, dependent on drug concentrations.

Future Screening Approaches and Targets in *IGH-CEBP* BCP-ALL

With continuing patient enrolment in UK trails, new *IGH-CEBP* patients are likely to be identified. To expand genomic data, these samples could be analysed with a combination of WGS, WES, SNP and RNA-seq. Using these approaches on a greater number of patients would add depth to the characterisation of this subgroup as recurring mutations could be identified, something that this project has lacked. SNP arrays would provide data on CNAs. They are cheap and were successfully used to identify the recurring *ABL2* deletion in this project. Potentially the most interesting platform for additional work would be RNA-seq, as the vastly different expression profiles of patients 11739 and 23395 have posed many questions as to the prominent mechanism of oncogenesis in the *IGH-CEBP* group. Epigenetic analysis would also be valuable as *CEBP* gene promoters have been shown to be methylated, leading to their inactivation. While this mechanism is unlikely to play a role in *IGH-CEBP* BCP-ALL, the *CEBP* genes are involved indirectly in acetylation and histone modification. An example is posttranscriptional inactivation of *CEBPB* by the histone-lysine N-methyl transferase, H3 lysine-9 specific 3, through binding to the *CEBPs* TAD (Pless *et al.*, 2008).

Final Conclusion

The discoveries made during this PhD support the increasing number of studies identifying the importance of genomic data for patient treatment. Findings such

as the identification of RTK-RAS pathway deregulation in HeH (Paulsson *et al.*, 2015), and the split in haploid BCP-ALL patients between cases with RAS activating mutations and those with *TP53*, *CDKN2A*/*B*, and/or *RB1* deletions (Holmfeldt *et al.*, 2013), show how improved screening is resulting in altered treatment options even in established BCP-ALL subgroups. Additional work is required with a larger patient cohort, but the *IGH-CEBP* subgroup shows several indications that it may be sub-divided further to impart the best possible patient treatment options. Whether these divisions are between individual *CEBP* subgroups based on clinical data, or individual patients based on genetic profiles, and specific CNAs such as *ABL2* deletions, remains to be confirmed. However data acquired in this project suggest that there are current therapeutic options such as dasatinib (*ABL2* deletions) and Plerixafor (CXCL12 overexpression) which could improve patient survival. Rather than going through the long process of developing new therapeutic agents, this body of work implies that treatment options in the *IGH-CEBP* subgroup may be already available.

Chapter 7 Supplementary Data

7.1 Supplementary Tables

Gene	No of probes	Region	Deletion / Deregulation in BCP-ALL
IKZF1	8	7p12.2	29%
CDKN2A/B	3	9p21.3	20%
PAX5	7	9p13.2	32%
EBF1	4	5q33.3	2%
ETV6	6	12p13.2	5%
BTG1	4	12q21.33	9%
RB1	5	13q14.2	5-11%
*PAR1 Region**	1 each	Xp22.33	6%

Supplementary Table 7.1. Table showing probe distribution in the SALSA MLPA P335-A1 ALL IKZF1 kit, location of targeted genes and incidence of genetic lesions in BCP-ALL.

Panel Name	Tube No.	Marker Name	Fluorochrome
Cincinnati Standard Panel	1	Thy1	PE
		CD34	APC
		CD11b	PB
		Death Marker	7AAD
	2	CD14	PE
		CD16	APC-Cy7
		CD33	APC
		Death Marker	7AAD
	3	CD19	PB
		CD10	PE-Cy7
		CD20	PE
		Death Marker	7AAD
Cincinnati Xenograft Panel	1	hCD45	PE-Cy7
		mCD45	APC-Cy7
		CD19	Pacific Blue
		CD13	PE
		CD33	PE
		CD34	APC
		Death Marker	7AAD

Supplementary Table 7.2. Flow cytometry panels used in Cincinnati

Panel Name	*Tube No.*	*Marker Name*	*Fluorochrome*
Newcastle Standard Panel	*1*	*CD34*	*APC*
		CD19	*PE*
		CD33	*PE-Vio770*
		CD10	*APC-Cy7*
		CD11b	*Pacific Blue*
		**DAPI / Zombie Aqua*	*UV/Pacific Blue*
Newcastle Xenograft FACS Panel	*1*	*hCD45*	*PerCP-Cy5.5*
		mCD45	*APC-Cy7*
		CD19	*PE*
		CD33	*PE-Vio770*
		DAPI	*UV*
Newcastle CD34+ Cell Sort FACS Panel	*1*	*hCD45*	*PerCP/Cy5.5*
		CD19	*PE*
		CD33	*PE-Vio770*
		DAPI	*UV*

Supplementary Table 7.3. Flow cytometry panels used in Newcastle. Newcastle standard panel switched from Zombie Aqua to DAPI during the course of experiment 3.*

Partner Gene	**Cytogenetic Break Point**
IGH-CEBPA / CEBPG	t(14;19)(q32;q13)
IGH-CEBPB	t(14;20)(q32;q13)
IGH-CEBPD	t(8;14)(q11;q32)
IGH-CEBPE	inv(14)(q11q32)/t(14;14)(q11;q32)

Supplementary Table 7.4. Cytogenetic breakpoints used to search for potential IGH-CEBP patients.

Patient ID	***IGH* Partner**
1798	*CEBPA*
4175	*CEBPA*

4198	*CEBPA*
4774	*CEBPA*
7143	*CEBPA*
7617	*CEBPA*
11540	*CEBPG*
3455	*CEBPB*
5632	*CEBPB*
10859	*CEBPB*
11682	*CEBPB*
2734	*CEBPD*
3622	*CEBPD*
3759	*CEBPD*
6889	*CEBPD*
20580	*CEBPD*
23395	*CEBPD*
22355	*CEBPD*
7247	*CEBPE*

Supplementary Table 7.5. Patient cohort created by Dr. L. J. Russell.

	Patient Age (%)					
***IGH* Partner**	**<10**	**10-18**	**18-25**	**>25**	**Mean**	**Median**
CEBPA	0(0)	5(50)	1(10)	4(40)	26	17
CEBPB	0(0)	3(38)	0(0)	5(62)	30	31
CEBPD	6(55)	4(36)	1(9)	0(0)	11	9
CEBPE	1(33)	0(0)	0(0)	2(67)	37	45
CEBPG	0(0)	1(100)	0(0)	0(0)	NA	NA
Total	7(21)	13(39)	2(6)	11(33)	22	15

Supplementary Table 7.6. Age of IGH-CEBP translocations patients divided by subgroup and age group (P=0.005).

	WBC x 109/L	
IGH Partner	**<50**	**50+**

CEBPA	9	1
CEBPB	5	3
CEBPD	10	1
CEBPE	3	0
CEBPG	1	0
Total	28	5

Supplementary Table 7.7. WBC of IGH-CEBP translocations patients (P=0.45).

	Number of Abnormalities	
***IGH* Partner**	**0**	**+1**
CEBPA	4(57)	3(43)
CEBPB	0(0)	6(100)
CEBPD	8(73)	3(27)
CEBPE	2(67)	1(33)
CEBPG	1(100)	0(0)
Total	15(54)	13(46)

Supplementary Table 7.8. Copy number abnormalities split by 0 and 1 or greater in the IGH-CEBP cohort (P=0.025).

***IGH* Partner**	0	1	2	3	5	6
CEBPA	4(27)	1(25)	0(0)	1(33)	0(0)	1(50)
CEBPB	0(0)	3(75)	2(67)	1(33)	0(0)	0(0)
CEBPD	8(53)	0(0)	1(33)	0(0)	1(100)	1(50)
CEBPE	2(13)	0(0)	0(0)	1(33)	0(0)	0(0)
CEBPG	1(7)	0(0)	0(0)	0(0)	0(0)	0(0)

Supplementary Table 7.9. Copy number abnormalities split numerical incidence in the IGH-CEBP cohort, percentages in brackets (P=0.05).

10859				
Red	**Green**	**Fusion**	**Gold**	**Average %**
1	1	1	1	68.8
1	1	1	2	18.2
0	0	2	1	4.4
0	0	2	2	3.1
1	1	1	3	1.9
0	1	2	1	1.2

2	1	1	1	0.5
2	1	1	1	0.5
0	1	1	1	0.5
0	0	2	3	0.2
0	1	1	1	0.2
1	1	1	4	0.2
2	1	1	2	0.2

Supplementary Table 7.10. Clonal evolution populations of patient 10859. Probes are CEBPB break apart probe in red and green and IKZF1 in gold.

4774				
Red	**Green**	**Fusion**	**Aqua**	**Average %**
1	1	1	0	54.6
0	0	2	2	20.4
0	0	2	0	9.2
2	1	1	0	3.9
1	1	1	2	3.3
1	0	2	0	3.3
1	1	1	1	2.0
2	2	0	0	1.3
0	0	2	1	1.3
1	0	2	1	0.7

Supplementary Table 7.11. Clonal evolution populations of patient 4774. Probes are CEBPA break apart probe in red and green and CDKN2A/B in aqua.

11739					
Red	**Green**	**Fusion**	**Gold**	**Aqua**	**Average %**
0	0	2	1	0	25.9
0	1	2	1	0	19.2
1	1	1	1	0	9.7
0	0	2	1	1	7.7

0	0	2	2	0	7.7
1	1	1	1	1	6.2
0	1	2	1	1	5.5
0	1	2	2	0	4.5
0	0	2	2	1	3.2
0	0	2	1	2	2.7
1	1	1	1	2	1.5
0	0	2	2	2	1.5
1	1	1	2	1	1.0
1	1	1	2	0	1.0
0	1	2	2	1	0.5
2	0	1	2	0	0.2
1	1	1	3	1	0.2
1	1	1	0	1	0.2
1	1	1	0	1	0.2
0	1	2	3	1	0.2
0	0	2	0	0	0.2
0	2	2	2	2	0.0

Supplementary Table 7.12. Clonal evolution populations of patient 11739. Probes are CEBPB break apart probe in red and green, IKZF1 probe in gold and CDKN2A/B in aqua.

3455				
Red	**Green**	**Fusion**	**Gold**	**Average %**
1	1	1	3	50.3
1	1	1	2	32.3
0	0	2	3	6.5
0	0	2	2	5.8
0	1	1	3	1.3
2	2	0	2	0.6
2	0	1	3	0.6
1	1	1	1	0.6
2	2	2	4	0.3

1	2	2	4	0.3
0	0	2	4	0.3
1	1	2	3	0.3
1	0	1	2	0.3
2	2	0	2	0.3
0	0	2	4	0.0

Supplementary Table 7.13. Clonal evolution populations of patient 3455. Probes are CEBPB break apart probe in red and green and PAX5 in gold.

11682				
Red	**Green**	**Fusion**	**Gold**	**Average %**
1	1	1	1	64.5
0	0	2	1	9.3
1	1	1	2	8.4
1	1	1	0	5.5
0	0	2	2	4.0
2	2	2	2	3.7
0	1	1	1	1.1
0	0	2	0	1.1
0	1	1	2	0.9
2	1	1	1	0.9
2	1	1	2	0.4
1	1	2	3	0.2

Supplementary Table 7.14. Clonal evolution populations of patient 11682. Probes are CEBPB break apart probe in red and green and IKZF1 in gold.

6889				
Red	**Green**	**Fusion**	**Aqua**	**Average %**
2	1	1	0	29.5
1	1	1	0	19.3

3	1	1	0	18.0
2	0	2	0	5.7
1	1	1	1	4.9
0	0	2	2	4.9
0	0	2	0	4.5
0	0	2	1	2.5
1	0	2	1	2.5
2	1	1	1	1.2
2	0	1	0	0.8
2	0	2	2	0.8
1	0	2	0	0.8
1	0	2	2	0.8
3	0	2	0	0.4
0	0	3	0	0.4
1	2	1	0	0.4
1	1	2	1	0.4
1	1	1	1	0.4
1	2	1	0	0.4
1	1	1	2	0.4
3	1	1	1	0.4
2	1	2	1	0.4

Supplementary Table 7.15. Clonal evolution populations of patient 6889. Probes are CEBPD break apart probe in red and green and CDKN2A/B in aqua.

20580					
Red	**Green**	**Fusion**	**Gold**	**Aqua**	**Average%**
0	0	2	2	2	24.1
1	1	1	1	0	17.4
1	1	1	1	1	9.4
2	0	2	2	2	7.1
1	0	2	2	2	4.9
2	1	1	1	0	4.0
1	1	2	1	1	3.1

1	1	1	0	0	3.1
1	1	1	2	1	2.2
0	0	2	1	2	2.2
0	0	2	1	1	2.2
2	2	1	1	1	1.8
2	1	1	2	0	1.3
0	0	2	2	0	1.3
1	1	2	2	2	1.3
2	1	1	1	2	1.3
0	1	1	1	1	1.3
1	0	2	2	1	1.3
2	0	2	2	0	0.9
0	0	2	1	0	0.9
2	1	2	2	2	0.9
1	0	1	1	2	0.9
2	1	1	1	2	0.9
3	2	1	2	0	0.4
1	1	1	2	2	0.4
0	0	2	2	1	0.4
0	0	2	4	2	0.4
2	1	1	2	1	0.4
0	0	2	3	2	0.4
1	1	1	2	0	0.4
2	0	2	3	2	0.4
2	2	1	0	0	0.4
2	2	1	2	0	0.4
2	1	1	2	1	0.4
0	1	1	1	0	0.4

Supplementary Table 7.16. Clonal evolution populations of patient 20580. Probes are CEBPD break apart probe in red and green, IKZF1 probe in gold and CDKN2A/B in aqua.

Patient	LRCG File Name	Diagnostic Sample Aros ID	MAPD	Remission Sample Aros ID	MAPD

20580	A1739	A1739-11	0.18	A1739-40	0.21
19794	A1739	A1739-41	0.20	A1739-42	0.21
19734	A1739	A1739-43	0.19	A1739-44	0.19
11540	A1739	A1739-45	0.20	A1739-46	0.19
22355	A1864	A1864-48	0.40	A1864-49	0.26
24880	A2883	A2883-06	0.22	A2883-07	0.20
2734	A1091	A1091 (file 0210)		NA	
5632	A1091	A1091 (file 0210)		NA	
4175	A1613	A1613-02	0.23	NA	
7143	A1613	A1613-03	0.35	NA	
3622	A1613	A1613-04	0.19	NA	
3759	A1613	A1613-05	0.19	NA	
11682	A1613	A1613-07	0.32	NA	
10859	A1864	A1864-46	0.33	NA	
6889	A1864	A1864-47	0.40	NA	

Supplementary Table 7.17. Table displaying patients analysed by SNP array, data shows patient number, file names and corresponding MAPDH value.

Gene Name	**Occurrence**	**Location**	**CNS**	**Valid Target?**
RHD	x2	Chr 1 p36.11	Deletion (0-1)	No, Blood group D antigen
ABL2	x4	Chr 1 q25.2	Deletion (1)	Yes, related to ABL1 and specific to this subgroup
OR4N4, OR4N3P, OR4M2	x3	Chr 15 q11.2	Deletion (1)	No, region of high variation
REREP3	x3	Chr 15 q11.2	Deletion (1)	No, region of high variation and pseudogene
SCAPER	x5	Chr 15 q24.3	Gain / Deletion	No, region of high variation and present in multiple controls
NPIP	x2	Chr 16 p11.2	Deletion (1)	No, region of high variation
PDXDC1	x2	Chr 16 p13.11	Deletion (1)	No, region of high variation
FAM86DP	x2	Chr 3 p12.3	Deletion (1)	No, Pseudogene and in a region of high variation

UGT2B15, UGT2B17	x4	Chr 4 q13.2	Deletion (0-1)	No, region of high variation
GALNTL6	x2	Chr 4 q34.1	Deletion (0-1)	No, region of high variation
C6orf142 (MLIP)	x3	Chr 6 p12.1	Deletion (1)	No, region of high variation and small intragenic deletion
ADAM5P, ADAM3A	x3	Chr 8 p11.22	Gain / Deletion	No, region of high variation, potential pseudogenes
LOC100132396 (ZNF705B)	x2	Chr 8 p23.1	Gain / Deletion	No, region of high variation
KMT2AT3	x3	Chr 9 p21.3	Deletion (1)	No, same region as CDKN2A/B
KLHL9	x3	Chr 9 p21.3	Deletion (0-1)	No, same region as CDKN2A/B
KIAA1797	x3	Chr 9 p21.3	Deletion (0-1)	No, same region as CDKN2A/B
IFNB1 + Multiple other IFN genes	x3	Chr 9 p21.3	Deletion (0-1)	No, same region as CDKN2A/B
FBXW10	x2	Chr 17 p11.2	Gain / Deletion	No functional relevence
FAM18B1	x2	Chr 17 p11.2	Gain / Deletion	Possibly, highly conserved unknown function
LOC220594 (USP32P2)	x2	Chr 17 p11.2	Gain / Deletion	No, region of high variation
LOC284344	x3	Chr 19 q13.31	Deletion (1)	No, region of high variation
PSG4	x3	Chr 19 q13.31	Deletion (1)	No, region of high variation
PSG9	x2	Chr 19 q13.31	Deletion (1)	No, region of high variation

Supplementary Table 7.18. Potential IGH-CEBP cohort target genes identified using the SNP platform.

Patient	**Abnormality**	**Notes**
20120305_145154	Gain 3	1q Gain
BCR-ABL SNP#36	Gain 3	1q Gain
BCR-ABL SNP#36	Gain 3	1q Gain
BCR-ABL SNP#40	Gain 3	1q Gain
BCR-ABL-SNP#20-250kknsp	Deletion 1	NA

BCR-ABL-SNP#21-250kknsp	Deletion 1	NA
BCR-ABL-SNP#24-250kknsp	Deletion 1	NA
BCR-ABL-SNP#24-250kknsp	Deletion 1	NA
BCR-ABL-SNP#27-250kknsp	Deletion 1	NA
BCR-ABL-SNP#28-250kknsp	Deletion 1	NA
BCR-ABL-SNP#30-250kknsp	Deletion 1	NA
BCR-ABL-SNP#34-250kknsp	Deletion 1	NA
BCR-ABL-SNP#35-250kknsp	Deletion 1	NA
BCR-ABL-SNP#36-250kknsp	Deletion 1	NA
BCR-ABL-SNP#4-250kknsp	Deletion 1	NA
BCR-ABL-SNP#9-250kknsp	Deletion 1	NA
E2A-PBX1 SNP #1	Gain 3	1q Gain
E2A-PBX1 SNP #10	Gain 3	1q Gain
E2A-PBX1 SNP #11	Gain 3	1q Gain
E2A-PBX1 SNP #12	Gain 3	Partial Gain
E2A-PBX1 SNP #13	Gain 3	1q Gain
E2A-PBX1 SNP #14	Gain 3	1q Gain
E2A-PBX1 SNP #15	Gain 3	1q Gain
E2A-PBX1 SNP #16	Gain 3	1q Gain
E2A-PBX1 SNP #17	Gain 3	1q Gain
E2A-PBX1 SNP #2	Gain 3	1q Gain
E2A-PBX1 SNP #3	Gain 3	1q Gain
E2A-PBX1 SNP #4	Gain 3	1q Gain
E2A-PBX1 SNP #5	Gain 3	1q Gain
E2A-PBX1 SNP #6	Gain 3	1q Gain
E2A-PBX1 SNP #7	Gain 3	1q Gain
E2A-PBX1 SNP #8	Gain 3	Partial Gain
E2A-PBX1 SNP #9	Gain 3	1q Gain
E2A-PBX1-SNP#11-250kknsp	Deletion 1	NA
E2A-PBX1-SNP#5-250kknsp	Deletion 1	NA
E2A-PBX1-SNP#6-250kknsp	Deletion 1	NA
Hyperdip50 #10	Gain 3	1q Gain
Pseudodip #10	Deletion 1	1q Deletion

Supplementary Table 7.19. All ABL2 copy number alterations in the Mullighan paediatric ALL 2007 cohort.

11682				
CEBPD	***ABL2***			
Red	***Green***	***Fusion***	***Gold***	***Average %***
1	*1*	*1*	*1*	*47.0*
1	*1*	*1*	*2*	*24.2*

0	*0*	*2*	*2*	*16.2*
1	*1*	*1*	*0*	*8.6*
0	*0*	*2*	*1*	*6.6*
0	*0*	*2*	*0*	*2.0*
2	*2*	*2*	*2*	*0.5*

Supplementary Table 7.20. Table showing FISH scores for CEBP break apart probes and ABL2 copy number probe for CEBPD patient 11682.

10859				
CEBPB	***ABL2***			
Red	***Green***	***Fusion***	***Gold***	***Average %***
1	*1*	*1*	*2*	*57.0*
1	*1*	*1*	*1*	*27.9*
0	*0*	*2*	*2*	*7.8*
0	*0*	*2*	*1*	*2.5*
2	*1*	*1*	*2*	*1.6*
2	*1*	*1*	*1*	*1.2*
0	*0*	*1*	*1*	*0.8*
1	*1*	*1*	*0*	*0.8*
0	*0*	*2*	*0*	*0.4*

Supplementary Table 7.21. Table showing FISH scores for CEBP break apart probes and ABL2 copy number probe for CEBPB patient 10859.

6889				
CEBPD	***ABL2***			
Red	***Green***	***Fusion***	***Gold***	***Average %***
1	*1*	*1*	*2*	*50.4*
1	*1*	*1*	*1*	*24.4*
2	*1*	*1*	*2*	*16.3*
0	*0*	*2*	*2*	*3.7*
0	*0*	*2*	*1*	*2.2*
2	*1*	*1*	*1*	*2.2*
1	*0*	*2*	*2*	*0.7*

Supplementary Table 7.22. Table showing FISH scores for CEBP break apart probes and ABL2 copy number probe for CEBPD patient 6889.

Primer Name	**Position**	**Forward Sequence**	**Reverse Sequence**
ABL2 Set 1	Exon 2	TTTGAATGCCATGAAAAGGA	TCCATTCCCTGTTCTCCATC
ABL2 Set 2	Exon 2	ATCACTTTGCCAGCTGTGTG*	AACCCTTGAATTTGTGGTTCC
ABL2 Set 3	Exon 2	GAAGCTTAAGAAAAGTGACGTGGT	TGCCAATGCCTTAGTTCAAA*
ABL2 Set 1	Exon 3	CTTTGCATCGTCCCTATGGT*	TGAGTGTGTTATCACCACTTGCT*
ABL2 Set 2	Exon 3	AGCTTTGCATCGTCCCTATG	CTGAGTGTGTTATCACCACTTGC

ABL2 Set 1	Exon 5	TTGGCAGAGCTTGTACACCA*	GACGCCAACGTAAACCTCTC*
ABL2 Set 2	Exon 5	TGGGCTGGTGACAACATTAC	GACGCCAACGTAAACCTCTC
ABL2 Set 1	Exon 7	TGTGTACTTTGGAGCCACCAT*	AGTACTCCATTGCAGAAGAAATCTG*
ABL2 Set 2	Exon 7	TTTGGAGCCACCATTTTACA	TTCTCTAAGTACTCCATTGCAGAAGA
ABL2 Set 1	Exon 9	TGGGGTATTGTTGTGGGAAA*	CTCAGGCTGTTCCATTCGAT*
ABL2 Set 2	Exon 11	TAGCTGAGGAGCTTGGGAGA*	CTGGTGCTAAACTGGAAGCA*
B2M	Intron 1-2	TCTAGGCGCCCGCTAAGTT*	TCGCGTGCTGTTTCCTCC*
RPLPO	Intron 2-3	ATAAACGGGCTCAGGCAAGTT*	CGCGCTCTTTTAGAAGCCAG*
TBP	Intron 5-6	TCTCTCTGACCATTGTAGCGGTT*	CCGTGGTTCGTGGCTCTCT*

*Supplementary Table 7.23. Primers used during optimisation of SYBR Green qPCR of ABL2 copy number analysis. Used primer combinations are denoted by *.*

Reg ID	DoB	Age At Diagnosis	Trial ID
23567	27/06/2007	2	IBFM-IGH
2734	18/06/1992	5	ALL97
22355	12/02/2003	6	ALL2003
23168	03/06/2001	8	ALL2003
6889	26/01/1995	8	MRD PILOT
25541	29/06/2004	8	RELAPSE
3622	07/03/1990	9	ALL97
1798	25/08/1984	10	UKALLXI
11540	30/05/1996	10	UKALLXIIR
4175	13/09/1988	11	ALL97
19734	01/05/1995	12	ALL2003
7617	08/10/1991	12	UKALLXIIR
23395		13	ALL2003
5632	12/10/1988	13	ALL97
25855	26/10/1998	14	UKALL2011
11739	28/04/1992	14	UKALLXIIR
24880	11/12/1995	15	ALL2003
3759	01/06/1984	15	ALL97
3455	05/02/1983	15	UKALLXII
19794	27/11/1989	17	ALL2003
20580	26/07/1989	18	ALL2003
4774	02/04/1982	19	UKALLXII
4198	02/02/1972	28	UKALLXII
11682	27/05/1976	30	ALL2003
25458	23/12/1980	31	ALL2003
10859	24/12/1971	34	ALL2003
5588	05/02/1959	43	UKALLXII
7143	14/11/1959	44	UKALLR3
7247	27/09/1958	45	UKALLXIIR
25505	18/08/1960	52	UKALL14
25952	29/12/1957	55	UKALL14

25686	20/06/1953	59	UKALL14
27181	11/04/1949	65	UKALL2011

Supplementary Table 7.24. Patient age and Trials, older patients in newer trials.

Primer Name	**Sequence**
-21M13	TTGTAAAACGACGGCCAGTG
M13 reverse	CACACAGGAAACAGCTATGAC C

Supplementary Table 7.25. Primers used for sequencing of plasmids by DBS Genomics.

RegID	**Subgroup**
20033	IGH-CRLF2 Rearrangement
11538	CRLF2 Rearrangement
20951	IGH-Other Rearrangement
2025	Other Abnormal
2058	Other Abnormal
6637	KMT2A Rearrangement
4679	TCF3 Rearrangement

Supplementary Table 7.26. Control samples for TaqMan qPCR analysis.

RegID	**Subgroup**	**Protein Conc ng/µl**
11672	B-Other	39
22045	B-Other	40
21532	B-Other	36
21226	B-Other	37
11739	IGH-CEBPB	46

Supplementary Table 7.27. BCP-ALL Control patients for Western immunoblotting.

LRCG No.	**Patient ID**	**Subgroup**
LRCG 1	27422	B-other
LRCG 6	20515	B-other
LRCG 14	20683	B-other
LRCG 18	22340	B-other
LRCG 20	10442	B-other
LRCG 21	12356	B-other
LRCG 19	10248	B-other
LRCG 17	21795	B-other
LRCG 7	27460	iAMP21
LRCG 8	25190	iAMP21
LRCG 9	27109	iAMP21

LRCG 16	21567	iAMP21
LRCG 2	7147	JAK-PDGFRB Translocation
LRCG 4	5985	TCR translocation

Supplementary Table 7.28. BCP-ALL Control patients for RNA-seq. Comprised of eight B-other, four iAMP21, and two translocation patients.

Pathway / Function	**Gene**	**CEBPB Role**	**11739 Fold Change**	**Link to Data**	**Paper**
Immune System	CD14	Promoter	0.0	http://onlinelibrary.wiley.com/doi/10.1002/jcp.24513/epdf	Wang *et al* 2007
	CHIT1	Promoter	0.8	http://www.ncbi.nlm.nih.gov/pubmed/17540774	Pham *et al* 2007
	CXCL8	Promoter	-7.5	http://www.ncbi.nlm.nih.gov/pubmed/19734226	John *et al* 2009
	PPARG	Promoter	-0.9	http://www.spandidos-publications.com/mmr/6/5/961	Meng *et al* 2012
Haematopoiesis	ID1	Promoter	-4.0	http://www.ncbi.nlm.nih.gov/pubmed/9001238	Saisanit *et al* 1997
	ID2	Promoter	-3.7	http://www.ncbi.nlm.nih.gov/pubmed/15809228	Karaya *et al* 2005
	GATA2	Repressor	-2.3	http://mcb.asm.org/content/25/2/706.long	Tong *et al* 2005
	ZFPM1	Repressor	-1.6	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261555/	Mancini *et al* 2012
	MYB	Promoter	0.4	http://www.ncbi.nlm.nih.gov/pubmed/11792321	Tahirov *et al* 2007
	SMARCA2	Promoter	0.9	http://www.ncbi.nlm.nih.gov/pubmed/10619021	Kowenz-Leutz *et al* 1999
Pathway / Function	**Gene**	**CEBPD Role**	**23395 Fold Change**	**Link to Data**	**Paper**
Immune System	CD14	NA	0.8		
	CHIT1	NA	0.7		
	CXCL8	NA	3.1		
	PPARG	Promoter	3.7	http://tinyurl.com/j3vq4p9	Cao *et al* 1995
Haematopoiesis	ID1	NA	1.1		
	ID2	NA	2.3		
	GATA2	NA	-0.4		
	ZFPM1	NA	-1.0		
	MYB	NA	1.0		
	SMARCA2	NA	0.5		

Supplementary Table 7.29. Documented expression patterns of CEBPB targets. Data was gathered using BioGRID 3.4, String (known and predicted protein-protein interactions) and pubmed.

Top 15 Upregulated Genes	**Fold Change**	**Function**	**Association**	**Journal Links**
SCN5A	9.2	A key regulator of a gene transcriptional network that controls colon cancer invasion and metastasis.	Cancer, Invasion	
MYF6	8.0	Involved in muscle differentiation (myogenic factor) potentially upregulated in breast cancer.	Differentiation	http://www.ncbi.nlm.nih.gov/pubmed/20651255

BMP8B	7.8	Mediates the survival of pancreatic cancer cells and regulates the progression of pancreatic cancer.	Cancer	
HS3ST4	7.6	Expression of this gene is thought to play a role in HSV-1 pathogenesis.	Disease, Metabolism	
XIST	7.3	Inactivates the X gene.		
LAMP5	6.9	Protein coding gene.		
DFNA5	6.9	Methylated in colorectal cancer, is a tumour suppressor gene.	Cancer, Tumour Suppressor	
DCHS2	6.4	Gastric and colorectal cancer adhesion.	Cancer, Adhesion	
EFEMP1	6.4	This gene encodes a member of the fibulin family of extracellular matrix glycoproteins, also involved in ERK signalling and prostate cancer.	ERK signalling, Cancer	http://www.ncbi.nlm.nih.gov/pubmed/25211630
NCKAP5	6.1	Protein coding gene. Associated with hypersomnia.		
HCG22	5.8	Protein coding gene. Associated with folicular lymphoma.	Cancer	
SPRY4	5.7	Invovled in MAPK singnalling inhibition. Tumour suppressor small cell carcinoma.	Cancer, Tumour Suppressor	
SYNPO	5.6	Actin-associated protein that may play a role in actin-based cell shape and motility.	Adhesion	
CSPG4	5.6	Onocogene in melanoma, carcinoma head and neck and breast.	Cancer, Adhesion	http://www.ncbi.nlm.nih.gov/pubmed/20455858
CHST3	5.5	Encodes an enzyme which is found in the extracellular matrix and most cells which is involved in cell migration and differentiation.	Adhesion, Metabolism	
Top 15 Downregulated Genes	**Fold Change**	**Function**	**Association**	**Journal Links**
MCAM	-8.4	Melanoma adhesion molecule, associated with metastasis in several cancers.	Cancer, Adhesion	
RGS1	-7.8	The gene encodes a member of the regulator of G-protein signalling family. Found to be a poor risk marker in melanoma.	Cancer, Adhesion	http://www.ncbi.nlm.nih.gov/pubmed/18580492
BTG3	-7.6	This family has structurally related proteins that appear to have antiproliferative properties. Breast and gastric cancer target.	Cancer	
CXCL8	-7.5	This chemokine is one of the major mediators of the inflammatory response, Involved in carcinogenesis metastasis and angiogenesis.	Cancer, Angiogenesis	http://www.hindawi.com/journals/isrn/2013/859154/
JUN	-7.5	Bzip family member - invovled in cAMP signalling, inflammation, and several cancers.	Cancer, Inflammation	http://www.nature.com/nrc/journal/v5/n8/full/nrc1687.html
SCN3A	-6.8	Responsible for the generation and propagation of action potentials in neurons		

		and muscle. Involved in cAMP signalling.		
ATF3	-6.6	Encodes a member of the mammalian activation transcription factor/cAMP responsive element-binding (CREB) protein family of transcription factors. Tumour suppressor in prostate cancer.	Cancer, Tumour Suppressor	http://www.nature.com/onc/journal/v34/n38/full/onc2014426a.html
NEU4	-6.6	Has a broad substrate specificity being active on glycoproteins. Contributes to invasive properties of colon cancers.	Cancer	
MYO18B	-6.5	May influence intracellular trafficking when in the cytoplasm. Mutations in this gene are associated with lung cancer, potential tumour suppressor gene.	Cancer, Tumour Suppressor	
FFAR1	-6.4	May be involved in the metabolic regulation of insulin secretion. Onco gene in breast cancer.	Cancer	
MSR1	-6.3	Macrophage receptor, mutations of which lead to cancer.	Cancer, Immune System	
RPS4Y1	-6.1	Ribosome protein coding gene.		
FOS	-6.1	Heterodimerises with JUN to form AP-1.	Inflammation, PI3K cascade	
ARPP21	-6.1	Encodes a cAMP-regulated phosphoprotein.	cAMP signalling	
ZNF492	-6.1	May be involved in transcriptional regulation.	Gene Expression	

Supplementary Table 7.30. Top 15 up and down regulated genes via the RNAseq platform in IGH-CEBPB patient 11739.

Pathway / Function	**Gene**	**CEBPD Role**	**23395 Fold Change**	**Link to Data**	**Paper**
cAMP signalling	CREB1	Promoter	0.3	http://www.jbc.org/content/278/38/36959.long	Kovacs *et al* 2003
	CREBBP	Promoter	0.4	http://www.jbc.org/content/278/38/36959.long	Kovacs *et al* 2003
	EP300	Promoter	-0.7	http://www.ncbi.nlm.nih.gov/pubmed/16397300	Wang *et al* 2006
Immune System	RELA	Promoter	0.8	http://www.ncbi.nlm.nih.gov/pubmed/9570146	Xia *et al* 1997
Inflammation	PPARG	Promoter	3.7	http://www.ncbi.nlm.nih.gov/pubmed/21257317	Tsukada *et al* 2010
	IL6	Promoter	6.2	http://www.nature.com/ni/journal/v10/n4/pdf/ni.1721.pdf	Litvak *et al* 2008
	IL10	Promoter	2.0	http://www.jbc.org/content/early/2002/12/18/jbc.M207448200.full.pdf	Brenner *et al* 2002
Apoptosis	SOD1	Promoter	0.1	http://www.ncbi.nlm.nih.gov/pubmed/20385105	Hour *et al 2010*
Hypoxia Signalling	HIF1A	Promoter	1.0	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3018791/	Balamurugan *et al2010*
Tumour Supressor	FBXW7	Repressor	0.7	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3018791/	Balamurugan *et al2010*
Ubiquitination	PER2	Promoter	1.0	http://ict.sagepub.com/content/8/4/317.long	Gery *et al* 2009

Pathway / Function	Gene	CEBPB Role	11739 Fold Change	Link to Data	Paper
cAMP signalling	CREB1	NA	0.5		
	CREBBP	NA	-0.1		
	EP300	NA	-0.9		
Immune System	RELA	NA	-1.5		
Inflammation	PPARG	Promoter	-0.9	http://www.jbc.org/content/281/12/7960.long	Zuo *et al* 2005
	IL6	Promoter	-2.6	http://www.ncbi.nlm.nih.gov/pubmed/21257317	Tsukada *et al* 2010
	IL10	Promoter	-3.8	http://www.jbc.org/content/early/2002/12/18/jbc.M207448200.full.pdf	Brenner *et al* 2002
Apoptosis	SOD1	NA	-1.1		
Hypoxia Signalling	HIF1A	NA	-1.7		
Tumour Supressor	FBXW7	NA	-0.7		
Ubiquitination	PER2	Promoter	-0.9	http://ict.sagepub.com/content/8/4/317.long	Gery *et al* 2009

Supplementary Table 7.31. Documented expression patterns of CEBPD targets. Data was gathered using BioGRID 3.4, String (known and predicted protein-protein interactions) and pubmed.

Top 15 Upregulated Genes	Fold Change	Function	Association	Journal Links
MMP13	9.4	Degrades collagen type I. Does not act on gelatin or casein. Breast cancer marker promostes angiogenesis.	Cancer, Angiogenesis	
POSTN	8.0	Enhances incorporation of BMP1 in the fibronectin matrix of connective tissues. Aids cancer metastasis.	Cancer, Metastasis	
TNC	8.0	Extracellular matrix protein implicated in guidance of migrating neurons as well as axons during development, synaptic plasticity as well as neuronal regeneration. Promostes metastasis.	Cancer, Metastasis	
MYF6	7.7	Involved in muscle differentiation (myogenic factor) potentially upregulated in breast cancer.		
COL3A1	7.7	Fibrillar collagen that is found in extensible connective tissues such as skin, lung, uterus, intestine and the vascular system.	PI3K-Akt	
IBSP	7.6	Potentially important to cell-matrix interaction. Involved in breast and prostate invasion and growth.	Cancer, Metastasis	http://www.ncbi.nlm.nih.gov/gene/3381
LUM	7.3	Important in collagen and other connective molecule formations. Overexpressed in multiple cancers and is linked to a poor outcome.	Cancer, Metastasis	http://www.ncbi.nlm.nih.gov/pubmed/17671699
COL1A2	7.2	Expressed in majority of collagens. Hypermethylatin of this gene is linked to outcome in head and neck cancer.	Cancer, PI3K-Akt	http://www.ncbi.nlm.nih.gov/pubmed/22674299
FN1	7.1	Fibronectin 1 acts as a potential biomarker for radioresistance.	Cancer	https://www.ncbi.nlm.nih.gov/pubmed/20930522
GJA1	7.1	Gap junction protein that acts as a regulator of bladder capacity. Multiple cancer links through metastasis.	Cancer, Metastasis	
IGFBP5	7.0	Gene associated with neuroblastoma and breast cancer metastasis.	Cancer, Metastasis	
STC1	7.0	Stimulates renal phosphate reabsorption. Involved in breast cancer metastasis.	Cancer, Metastasis	
ZNF804A	7.0	Zinc finger protein.		

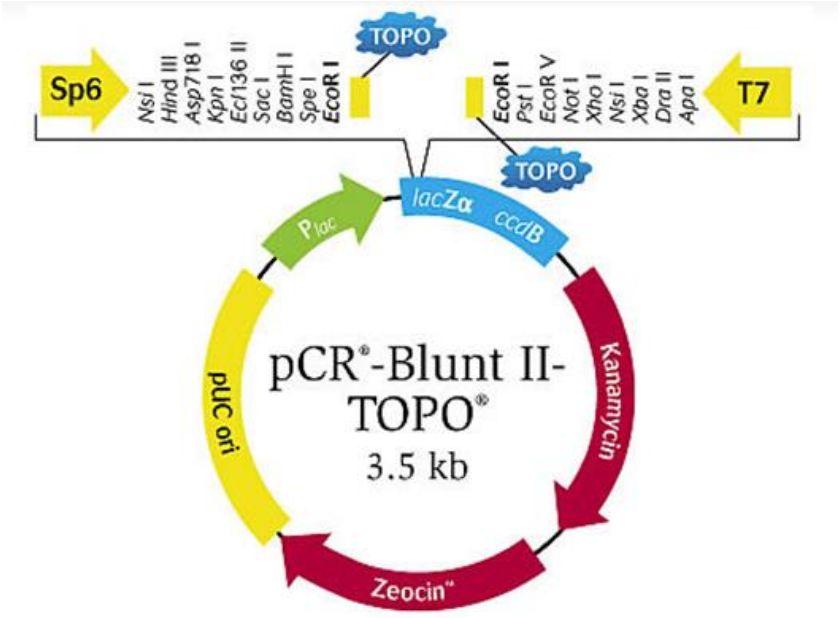
COL11A1	6.9	Collagen gene. Involved in ovarian cancer and tumour progression.	Cancer	
CXCL12	6.8	Macrophage recruitment chemokine involved in inflammation.	Inflammation, Immune System	
Top 15 Downregulated Genes	**Fold Change**	**Function**	**Association**	**Journal Links**
ZNF793	-6.1	May be involved in transcriptional regulation.	Transcriptional Regulation	
CALN1	-5.6	May play a role in the physiology of neurons and is potentially important in memory and learning.		
ZNF135	-5.5	Zinc finger protein involved in gene expression. Linked with renal pelvis carcinoma.	Gene Expression	
ZNF582-AS1	-5.2	Linc-RNA		
CLECL1	-5.2	May function in mediating immune cell-cell interactions. May act as a T-cell costimulatory molecule.	Immune System	
LOC100128398	-5.1	ncRNA gene.		
PNMA3	-5.0	Gene shares homology with retroviral Gag proteins.		
CSMD1	-5.0	Potential suppressor of squamous cell carcinoma.	Cancer, Tumour Suppressor	
EVC2	-5.0	Positive regulator of the hedgehog signalling pathway (By similarity). Plays a critical role in bone formation and skeletal development.	Sonic Hedgehog Signalling	
MYO7B	-4.9	Myosins are actin-based motor molecules with ATPase activity.		
GPA33	-4.8	May play a role in cell-cell recognition and signalling.		
BMPR1B	-4.8	Involved in endochondral bone formation and embryogenesis.		
RGMA	-4.7	This gene performs several functions in the developing and adult nervous system. Tumour suppressor in prostate cancer preventing metastasis.	Cancer, Tumour Suppressor	http://www.ncbi.nlm.nih.gov/pubmed/26721439
SEMA6A	-4.7	Important role in cell-cell signalling. Required for normal granule cell migration in the developing cerebellum.	Cell-cell Signalling	
PLEKHG4B	-4.7	Protein coding gene.		

Supplementary Table 7.32. Top 15 up and down regulated genes via the RNAseq platform in IGH-CEBPD patient 22395.

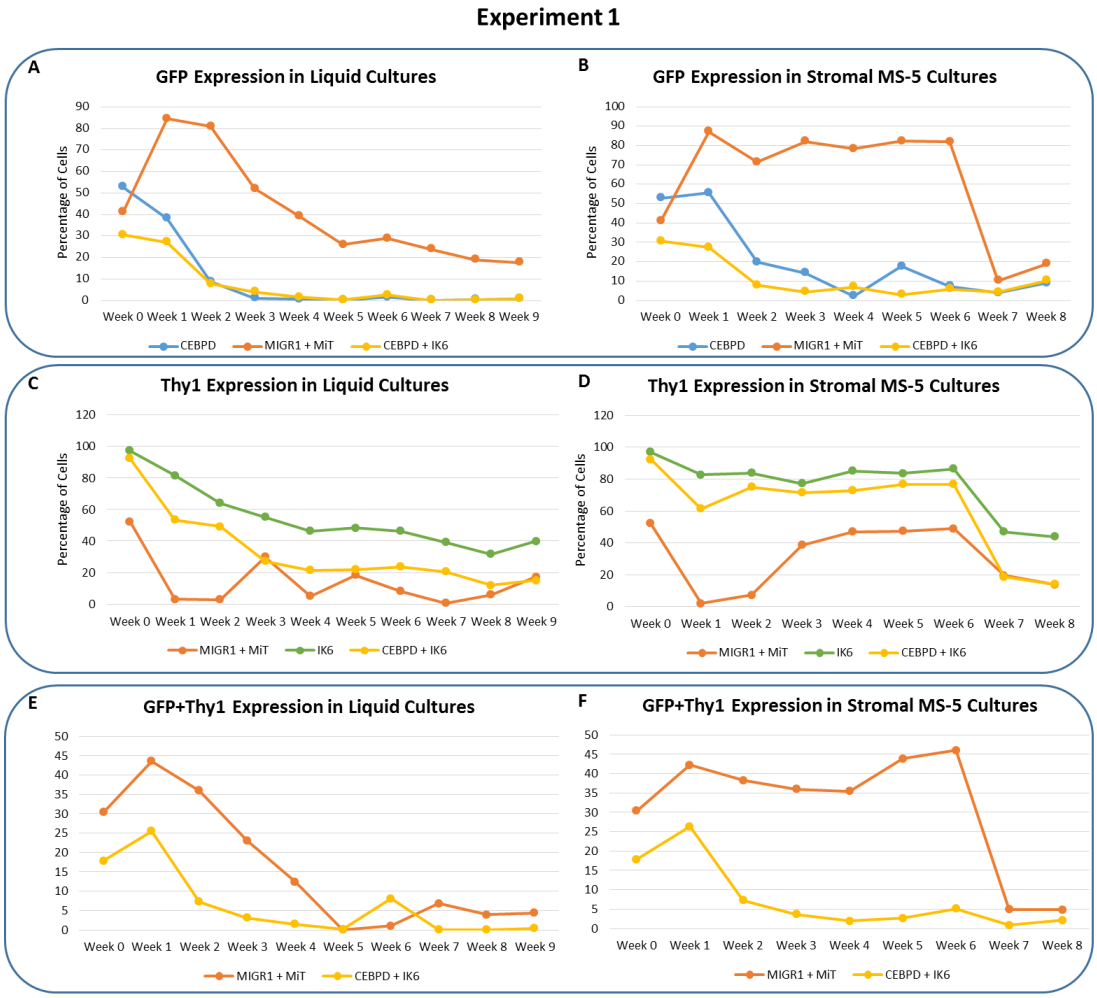
7.2 Supplementary Figures

	CEBPA			CEBPG	CEBPB			CEBPD							
	4175	7143	24880	11540	5632	10859	11682	2734	3622	3759	6889	19734	19794	20580	22355
IKAROS															
PAX5															
CDKN2A															
CDKN2B															
CRLF2															
CSF2RA															
ETV6															
IL3RA															
PAR1															
EBF1															
RB1															

Supplementary Figure 7.1. Concurrence between MLPA and SNP, green blocks shows positive concurrence, red blocks show a miss match.



Supplementary Figure 7.2. pCR-4TOPO construct, image from https://www.thermofisher.com/order/catalog/product/K457502.



Supplementary Figure 7.3 Transduction levels of Thy1 transduced and Thy1+GFP transduced in CD34+ cells in Experiment 1 in two culture conditions.

7.3 Supplementary Sequences

Homo sapiens CCAAT/enhancer binding protein (C/EBP), delta, mRNA (cDNA clone MGC:132769 IMAGE:8144112), complete cds

http://www.ncbi.nlm.nih.gov/nuccore/BC105109.1

ACAGCCTCGCTTGGACGCAGAGCCCGGCCCGACGCCGCCATGAGCGCCG
CGCTCTTCAGCCTGGACGGCCCGGCGCGCGGCGCGCCCTGGCCTGCGG
AGCCTGCGCCCTTCTACGAACCGGGCCGGGCGGGCAAGCCGGGCCGCG
GGGCCGAGCCAGGGGCCCTAGGCGAGCCAGGCGCCGCCGCCCCCGCCA
TGTACGACGACGAGAGCGCCATCGACTTCAGCGCCTACATCGACTCCATG
GCCGCCGTGCCCACCCTGGAGCTGTGCCACGACGAGCTCTTCGCCGACC
TCTTCAACAGCAATCACAAGGCGGGCGGCGCGGGGCCCCTGGAGCTTCT
TCCCGGCGGCCCCGCGCGCCCCTTGGGCCCGGGCCCTGCCGCTCCCCG
CCTGCTCAAGCGCGAGCCCGACTGGGGCGACGGCGACGCGCCCGGCTC

GCTGTTGCCCGCGCAGGTGGCCGCGTGCGCACAGACCGTGGTGAGCTTG
GCGGCCGCAGGGCAGCCCACCCCGCCCACGTCGCCGGAGCCGCCGCGC
AGCAGCCCCAGGCAGACCCCCGCGCCCGGCCCCGCCCGGGAGAAGAGC
GCCGGCAAGAGGGGCCCGGACCGCGGCAGCCCCGAGTACCGGCAGCGG
CGCGAGCGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGC
GGCGCAACCAGGAGATGCAGCAGAAGTTGGTGGAGCTGTCGGCTGAGAA
CGAGAAGCTGCACCAGCGCGTGGAGCAGCTCACGCGGGACCTGGCCGG
CCTCCGGCAGTTCTTCAAGCAGCTGCCCAGCCCGCCCTTCCTGCCGGCC
GCCGGGACAGCAGACTGCCGGTAACGCGCGGCCGGGGCGGGAGAGACT
CAGCAA

Supplementary Sequence 7.1. CEBPD cDNA insert sequence.

MIGR1 retroviral vector sequence

Cgatctctcgaggttaacgaattccgcccccccccccctaacgttactggccgaagccgcttggaataaggccg
gtgtgcgtttgtctatatgttattttccaccatattgccgtcttttggcaatgtgagggcccggaaacctggccctgtct
tcttgacgagcattcctaggggtctttcccctctcgccaaaggaatgcaaggtctgttgaatgtcgtgaaggaag
cagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccttgcaggcagcggaacccccccacc
tggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccag
tgccacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaa
ggatgcccagaaggtaccccattgtatgggatctgatctggggcctcggtgcacatgctttacatgtgtttagtcg
aggttaaaaaaacgtctaggccccccgaaccacggggacgtggttttcctttgaaaaacacgatgataatatg
gccacaaccatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacgg
cgacgtaaacggccacaagttcagcgtgtctggcgagggcgagggcgatgccacctacggcaagctgacc
ctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggc
gtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggc
tacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcga
gggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgggg
cacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatca
aggcgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcaga
acaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagc
aaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcgg
catggacgagctgtacaagtaatgaattaattaagaattatcaagcttatcgatttcgccgacctgcagccaag
cttatcgataaaataaaagattttatttagtctccagaaaaaggggggaatgaaagaccccacctgtaggtttg
gcaagctagcttaagtaacgccattttgcaaggcatggaaaatacataactgagaatagagaagttcagatc

aaggttaggaacagagagacagcagaatatgggccaaacaggatatctgtggtaagcagttcctgccccgg
ctcagggccaagaacagatggtccccagatgcggtcccgccctcagcagtttctagagaaccatcagatgttt
ccagggtgccccaaggacctgaaatgaccctgtgccttatttgaactaaccaatcagttcgcttctcgcttctgttc
gcgcgcttctgctccccgagctcaataaaagagcccacaacccctcactcggcgcgccagtcctccgataga
ctgcgtcgcccgggtacccgtgtatccaataaaccctcttgcagttgcatccgacttgtggtctcgctgttccttgg
gagggtctcctctgagtgattgactacccgtcagcgggggtctttcatgggtaacagtttcttgaagttggagaac
aacattctgagggtaggagtcgaatattaagtaatcctgactcaattagccactgttttgaatccacatactccaa
tactcctgaaatagttcattatggacagcgcaaaaagagctggggagaattgtgaaattgttatccgctcacaat
tccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacatt
aattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacg
cgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcgg
ctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagg
aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttc
cataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacag
gactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccg
gatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgt
aggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaac
tatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca
gagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagt
atttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaac
caccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatc
ctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatca
aaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttg
gtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctg
actccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcga
gacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtg
gtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaat
agtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctcc
ggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctcc
gatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcat
gccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgacc
gagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattg
gaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtg
cacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgcc
gcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcattt

atcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgca
catttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgta
tcacgaggccctttcgtctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacg
gtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgg
gtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaatac
cgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagtt
gggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccacgctctcccttatgcgactc
ctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgca
aggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctca
tgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctg
tggcgccggtgatgccggccacgatgcgtccggcgtagaggcgattagtccaatttgttaaagacaggatatc
agtggtccaggctctagttttgactcaacaatatcaccagctgaagcctatagagtacgagccatagataaaat
aaaagattttatttagtctccagaaaaaggggggaatgaaagaccccacctgtaggtttggcaagctagctta
agtaacgccattttgcaaggcatggaaaatacataactgagaatagagaagttcagatcaaggttaggaaca
gagagacagcagaatatgggccaaacaggatatctgtggtaagcagttcctgccccggctcagggccaag
aacagatggtccccagatgcggtcccgccctcagcagtttctagagaaccatcagatgtttccagggtgcccc
aaggacctgaaatgaccctgtgccttatttgaactaaccaatcagttcgcttctcgcttctgttcgcgcgcttctgct
ccccgagctcaataaaagagcccacaacccctcactcggcgcgccagtcctccgatagactgcgtcgcccg
ggtacccgtattcccaataaagcctcttgctgtttgcatccgaatcgtggactcgctgatccttgggagggtctcct
cagattgattgactgcccacctcgggggtctttcatttggaggttccaccgagatttggagacccctgcccaggg
accaccgacccccccgccgggaggtaagctggccagcggtcgtttcgtgtctgtctctgtctttgtgcgtgtttgtg
ccggcatctaatgtttgcgcctgcgtctgtactagttagctaactagctctgtatctggcggacccgtggtggaact
gacgagttctgaacacccggccgcaaccctgggagacgtcccagggactttgggggccgtttttgtggcccg
acctgaggaagggagtcgatgtggaatccgaccccgtcaggatatgtggttctggtaggagacgagaaccta
aaacagttcccgcctccgtctgaatttttgctttcggtttggaaccgaagccgcgcgtcttgtctgctgcagcgctg
cagcatcgttctgtgttgtctctgtctgactgtgtttctgtatttgtctgaaaattagggccagactgttaccactcccttt
aagtttgaccttaggtcactggaaagatgtcgagcggatcgctcacaaccagtcggtagatgtcaagaagag
acgttgggttaccttctgctctgcagaatggccaacctttaacgtcggatggccgcgagacggcacctttaacc
gagacctcatcacccaggttaagatcaaggtcttttcacctggcccgcatggacacccagaccaggtccccta
catcgtgacctgggaagccttggcttttgaccccctccctgggtcaagccctttgtacaccctaagcctccgcct
cctcttcctccatccgccccgtctctcccccttgaacctcctcgttcgaccccgcctcgatcctccctttatccagcc
ctcactccttctctaggcgccggaatt

Supplementary Sequence 7.2. MIGR1 retroviral vector sequence.

MIVR1 retroviral vector sequence

Cgatctctcgaggttaacgaattccgccccccccccctaacgttactggccgaagccgcttggaataaggccg
gtgtgcgtttgtctatatgttattttccaccatattgccgtcttttggcaatgtgagggcccggaaacctggccctgtct
tcttgacgagcattcctaggggtctttcccctctcgccaaaggaatgcaaggtctgttgaatgtcgtgaaggaag
cagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccttgcaggcagcggaacccccccacc
tggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccag
tgccacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaa
ggatgcccagaaggtaccccattgtatgggatctgatctggggcctcggtgcacatgctttacatgtgtttagtcg
aggttaaaaaaacgtctaggccccccgaaccacggggacgtggttttcctttgaaaaacacgatgataatatg
gccacaaccatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacgg
cgacgtaaacggccacaagttcagcgtgtctggcgagggcgagggcgatgccacctacggcaagctgacc
ctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggc
gtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggc
tacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcga
gggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgggg
cacaagctggagtacaactacaacagccacaacgtcTATATCACCGCCGACAAGCAGAAG
AACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccga
caaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctg
ctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaatgaattaattaagaatt
atcaagcttatcgatttcgccgacctgcagccaagcttatcgataaaataaaagattttatttagtctccagaaaa
aggggggaatgaaagaccccacctgtaggtttggcaagctagcttaagtaacgccattttgcaaggcatgga
aaatacataactgagaatagagaagttcagatcaaggttaggaacagagagacagcagaatatgggccaa
acaggatatctgtggtaagcagttcctgccccggctcagggccaagaacagatggtccccagatgcggtccc
gccctcagcagtttctagagaaccatcagatgtttccagggtgccccaaggacctgaaatgaccctgtgcctta
tttgaactaaccaatcagttcgcttctcgcttctgttcgcgcgcttctgctccccgagctcaataaaagagcccac
aacccctcactcggcgcgccagtcctccgatagactgcgtcgcccgggtacccgtgtatccaataaaccctctt
gcagttgcatccgacttgtggtctcgctgttccttgggagggtctcctctgagtgattgactacccgtcagcgggg
gtctttcatgggtaacagtttcttgaagttggagaacaacattctgagggtaggagtcgaatattaagtaatcctg
actcaattagccactgttttgaatccacatactccaatactcctgaaatagttcattatggacagcgcaaaaaga
gctggggagaattgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaa
agcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaa
cctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccg

cttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggta
atacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggcc
aggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatc
gacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctcc
ctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgct
ttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccc
cccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatc
gccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaa
gtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcgg
aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagca
gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacg
aaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatg
aagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacct
atctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggaggg
cttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaata
aaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaatt
gttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgt
ggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccc
atgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcac
tcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactc
aaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccg
cgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatctta
ccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtt
tctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttg
aatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgt
atttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaacc
attattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtctcgcgcgtttcggtgatgacg
gtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcaga
caagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcag
attgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcagg
cgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagct
ggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaa
acgacggccagtgccacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccg
ttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacgggg
cctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggt

gatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgt
agaggcgattagtccaatttgttaaagacaggatatcagtggtccaggctctagttttgactcaacaatatcacc
agctgaagcctatagagtacgagccatagataaaataaaagattttatttagtctccagaaaaaggggggaat
gaaagaccccacctgtaggtttggcaagctagcttaagtaacgccattttgcaaggcatggaaaatacataac
tgagaatagagaagttcagatcaaggttaggaacagagagacagcagaatatgggccaaacaggatatct
gtggtaagcagttcctgccccggctcagggccaagaacagatggtccccagatgcggtcccgccctcagca
gtttctagagaaccatcagatgtttccagggtgccccaaggacctgaaatgaccctgtgccttatttgaactaac
caatcagttcgcttctcgcttctgttcgcgcgcttctgctccccgagctcaataaaagagcccacaacccctcact
cggcgcgccagtcctccgatagactgcgtcgcccgggtacccgtattcccaataaagcctcttgctgtttgcatc
cgaatcgtggactcgctgatccttgggagggtctcctcagattgattgactgcccacctcgggggtctttcatttgg
aggttccaccgagatttggagacccctgcccagggaccaccgacccccccgccgggaggtaagctggcca
gcggtcgtttcgtgtctgtctctgtctttgtgcgtgtttgtgccggcatctaatgtttgcgcctgcgtctgtactagttagc
taactagctctgtatctggcggacccgtggtggaactgacgagttctgaacacccggccgcaaccctgggag
acgtcccagggactttgggggccgtttttgtggcccgacctgaggaagggagtcgatgtggaatccgaccccg
tcaggatatgtggttctggtaggagacgagaacctaaaacagttcccgcctccgtctgaatttttgctttcggtttg
gaaccgaagccgcgcgtcttgtctgctgcagcgctgcagcatcgttctgtgttgtctctgtctgactgtgtttctgtat
ttgtctgaaaattagggccagactgttaccactcccttaagtttgaccttaggtcactggaaagatgtcgagcgg
atcgctcacaaccagtcggtagatgtcaagaagagacgttgggttaccttctgctctgcagaatggccaacctttt
aacgtcggatggccgcgagacggcacctttaaccgagacctcatcacccaggttaagatcaaggtcttttcac
ctggcccgcatggacacccagaccaggtcccctacatcgtgacctgggaagccttggcttttgaccccctcc
ctgggtcaagccctttgtacaccctaagcctccgcctcctcttcctccatccgccccgtctctcccccttgaacctc
ctcgttcgaccccgcctcgatcctccctttatccagccctcactccttctctaggcgccggaatt

Supplementary Sequence 7.3 MIVR1 retroviral vector sequence.

IK6 cDNA Sequence

atggatgcggatgaaggccaggatatgagccaggtgagcggcaaagaaagcccgccggtgagcgatacc
ccggatgaaggcgatgaaccgatgccgattccggaagatctgagcaccaccagcggcggccagcagagc
agcaaaagcgatcgcgtggtgggcgataaaggcctgagcgataccccgtatgatagcagcgcgagctatg
aaaaagaaaacgaaatgatgaaaagccatgtgatggatcaggcgattaacaacgcgattaactatctgggc
gcggaaagcctgcgcccgctggtgcagaccccgccgggcggcagcgaagtggtgccggtgattagcccga
tgtatcagctgcataaaccgctggcggaaggcaccccgcgcagcaaccatagcgcgcaggatagcgcggt
ggaaaacctgctgctgctgagcaaagcgaaactggtgccgagcgaacgcgaagcgagcccgagcaaca
gctgccaggatagcaccgataccgaaagcaacaacgaagaacagcgcagcggcctgatttatctgaccaa

ccatattgcgccgcatgcgcgcaacggcctgagcctgaaagaagaacatcgcgcgtatgatctgctgcgcgc
ggcgagcgaaaacagccaggatgcgctgcgcgtggtgagcaccagcggcgaacagatgaaagtgtataa
atgcgaacattgccgcgtgctgtttctggatcatgtgatgtataccattcatatgggctgccatggctttcgcgatcc
gtttgaatgcaacatgtgcggctatcatagccaggatcgctatgaatttagcagccatattacccgcggcgaac
atcgctttcatatgagc

Supplementary Sequence 7.4 IK6 cDNA insert sequence.

ngagctcaggacaggtagaaagaatgaatagaacaataaaagagacccttactaaattgaccttagagact
ggcttaaaagattggagacgcctcctatctctggctttgttaagagccagaaatacgcccaaccgttttcggctc
accccatatgaaatcctttatgggggaccccccccttttgtcaaccttgctcaattccttctcccctccgatcctaa
gactgatttacaagcccgactaaaagggctgcaaggcgtgcaggcccaaatctggacacccctggccgaat
tgtaccggccaggacatccacaaactagccacccatttcaggtgggagactccgtgtacgtccggcggcacc
gctctcaaggattggagcctcgttggaagggaccttacatcgtcctgctgaccacgcccaccgccataaaggtt
gacgggatcgccgcctggattcacgcatcgcacgccaaggcagccccaaaaaccccctggaccagaaact
cccaaaacctggaagctccgccgttcggagaaccctcttaagataagactctcccgtgtctgactgctaatcca
ccttgtccctgtactaacccaaaatgaaactcccaacaggaatggtcattttatgtagcctaataatagttcgggc
agggtttgacgacccccgcaaggctatcgcattagtacaaaaacaacatggtaaaccatgcgaatgcagcg
gagggcaggtatccgaggccccaccgaactccatccaacaggtaacttgcccaggcaagacggcctactta
atgaccaaccaaaaatggaaatgcagagtcactccaaaaatctcacctagcgggggagaactccagaact
gcccctgtaacactttccaggactcgatgcacagttcttgttatactgaataccggcaatgcaggcgaattaata
agacatactacacggccaccttgcttaaaatacggtctgggagcctcaacgaggtacagatattacaaaacc
ccaatcagctcctacagtccccttgtaggggctctataaatcagcccgtttgctggagtgccacagcccccatcc
atatctccgatggtggaggacccctcgatactaagagagtgtggacagtccaaaaaaggctagaacaaattc
ataaggctatgactcctgaacttcaataccacccctagccctgcccaaagtcagagatgaccttagccttgatg
cacggacttttgatatcctgaataccacttttaggttactccagatgtccaattttagccttgcccaagattgttggct
ctgtttaaaactaggtacccctacccctcttgcgatacccactccctctttaacctactccctagcagactccctag
cgaatgcctcctgtcagattatacctcccctcttggttcaaccgatgcagttctccaactcgtcctgtttatcttcccct
ttcattaacgatacggaacaaatagacttaggtgcagtcacctttactaactgcacctctgtagccaatgtcagta
gtcctttatgtgccctaaacgggtcagtcttcctctgtggaaataacatggcatacacctatttaccccaaaactg
gaccagactttgcgtccaagcctccctcctccccgacattgacatcaacccgggggatgagccagtccccatt
cctgccattgatcattatatacatagacctaaacgagctgtacagttcatccctttactagctggactgggaatca
ccgcagcattcaccaccggagctacaggcctaggtgtctccgtcacccagtatacaaaattatcccatcagtta
atatctgatgtccaagtcttatccggtaccatacaagatttacaagaccaggtagactcgttagctgaagtagttc
tccaaaataggaggggactggacctactaacggcagaacaaggaggaatttgtttagccttacaagaaaaa

tgctgtttttatgctaacaagtcaggaattgtgagaaacaaaataagaaccctacaagaagaattacaaaaac
gcagggaaagcctggcaaccaaccctctctggaccgggctgcagggctttcttccgtacctcctacctctcctg
ggacccctactcaccctcctactcatactaaccattgggccatgcgttttcagtcgcctcatggccttcattaatga
tagacttaatgttgtacatgccatggtgctggcccagcaataccaagcactcaaagctgaggaagaagctca
ggattgagcttccgggacaaaagcaggggggaatgagaagtcagaacccccccacctttgctacataaataa
ccgctttcatttcgcttctgtaaaacgcttatgcgccccaccctagccggaaagtccccagccgctacgcaacc
cgggccccgagttgcatcagccgttcgcaacccgggctccgagttgcatcagccgaaagaaacttcatttcca
agctt

Supplementary Sequence 7.5 FEV RD114 envelope sequence.

Chapter 8 Bibliography

Agrawal, S., Hofmann, W.K., Tidow, N., Ehrich, M., van den Boom, D., Koschmieder, S., Berdel, W.E., Serve, H. and Muller-Tidow, C. (2007) 'The C/EBPdelta tumor suppressor is silenced by hypermethylation in acute myeloid leukemia', *Blood*, 109(9), pp. 3895-905.

Akasaka, T., Balasas, T., Russell, L.J., Sugimoto, K.J., Majid, A., Walewska, R., Karran, E.L., Brown, D.G., Cain, K., Harder, L., Gesk, S., Martin-Subero, J.I., Atherton, M.G., Bruggemann, M., Calasanz, M.J., Davies, T., Haas, O.A., Hagemeijer, A., Kempski, H., Lessard, M., Lillington, D.M., Moore, S., Nguyen-Khac, F., Radford-Weiss, I., Schoch, C., Struski, S., Talley, P., Welham, M.J., Worley, H., Strefford, J.C., Harrison, C.J., Siebert, R. and Dyer, M.J. (2007) 'Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL)', *Blood*, 109(8), pp. 3451-61.

Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990) 'A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family', *Embo J*, 9(6), pp. 1897-906.

Alberich-Jorda, M., Wouters, B., Balastik, M., Shapiro-Koss, C., Zhang, H., Di Ruscio, A., Radomska, H.S., Ebralidze, A.K., Amabile, G., Ye, M., Zhang, J., Lowers, I., Avellino, R., Melnick, A., Figueroa, M.E., Valk, P.J., Delwel, R. and Tenen, D.G. (2012) 'C/EBPgamma deregulation results in differentiation arrest in acute myeloid leukemia', *J Clin Invest*, 122(12), pp. 4490-504.

Angel, P. and Karin, M. (1991) 'The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation', *Biochim Biophys Acta*, 1072(2-3), pp. 129-57.

Antonson, P., Stellan, B., Yamanaka, R. and Xanthopoulos, K.G. (1996) 'A novel human CCAAT/enhancer binding protein gene, C/EBPepsilon, is expressed in cells of lymphoid and myeloid lineages and is localized on chromosome 14q11.2 close to the T-cell receptor alpha/delta locus', *Genomics*, 35(1), pp. 30-8.

Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P. and Akashi, K. (2007) 'Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages', *Cell Stem Cell*, 1(4), pp. 416-27.

Asou, H., Gombart, A.F., Takeuchi, S., Tanaka, H., Tanioka, M., Matsui, H., Kimura, A., Inaba, T. and Koeffler, H.P. (2003) 'Establishment of the acute myeloid leukemia cell line Kasumi-6 from a patient with a dominant-negative mutation in the DNA-binding region of the C/EBPalpha gene', *Genes Chromosomes Cancer*, 36(2), pp. 167-74.

Balamurugan, K. and Sterneck, E. (2013) 'The many faces of C/EBPdelta and their relevance for inflammation and cancer', *Int J Biol Sci*, 9(9), pp. 917-33.

Balamurugan, K., Wang, J.M., Tsai, H.H., Sharan, S., Anver, M., Leighty, R. and Sterneck, E. (2010) 'The tumour suppressor C/EBPdelta inhibits FBXW7 expression and promotes mammary tumour metastasis', *EMBO J*, 29(24), pp. 4106-17.

Bao, Z.H., Li, G.L. and Yu, J.H. (2007) '[Expression of cyclooxygenase-2 in bone marrow cells of chronic leukemia and its significance]', *Zhongguo Shi Yan Xue Ye Xue Za Zhi*, 15(5), pp. 923-6.

Barabe, F., Kennedy, J.A., Hope, K.J. and Dick, J.E. (2007) 'Modeling the initiation and progression of human acute leukemia in mice', *Science*, 316(5824), pp. 600-4.
Barbaro, V., Testa, A., Di Iorio, E., Mavilio, F., Pellegrini, G. and De Luca, M. (2007) 'C/EBPdelta regulates cell cycle and self-renewal of human limbal stem cells', *J Cell Biol*, 177(6), pp. 1037-49.
Barber, K.E., Harrison, C.J., Broadfield, Z.J., Stewart, A.R.M., Wright, S.L., Martineau, M., Strefford, J.C. and Moorman, A.V. (2007) 'Molecular cytogenetic characterization of TCF3 (E2A)/19p13.3 rearrangements in B-cell precursor acute lymphoblastic leukemia', *Genes, Chromosomes and Cancer*, 46(5), pp. 478-486.
Barker, E. and Planelles, V. (2003) 'Vectors derived from the human immunodeficiency virus, HIV-1', *Front Biosci*, 8, pp. d491-510.
Bateman, C.M., Colman, S.M., Chaplin, T., Young, B.D., Eden, T.O., Bhakta, M., Gratias, E.J., van Wering, E.R., Cazzaniga, G., Harrison, C.J., Hain, R., Ancliff, P., Ford, A.M., Kearney, L. and Greaves, M. (2010) 'Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia', *Blood*, 115(17), pp. 3553-3558.
Bedi, R., Du, J., Sharma, A.K., Gomes, I. and Ackerman, S.J. (2009) 'Human C/EBP-epsilon activator and repressor isoforms differentially reprogram myeloid lineage commitment and differentiation', *Blood*, 113(2), pp. 317-27.
Bell, A.J., Jr., Fegen, D., Ward, M. and Bank, A. (2010) 'RD114 envelope proteins provide an effective and versatile approach to pseudotype lentiviral vectors', *Exp Biol Med (Maywood)*, 235(10), pp. 1269-76.
Benekli, M., Xia, Z., Donohue, K.A., Ford, L.A., Pixley, L.A., Baer, M.R., Baumann, H. and Wetzler, M. (2002) 'Constitutive activity of signal transducer and activator of transcription 3 protein in acute myeloid leukemia blasts is associated with short disease-free survival', *Blood*, 99(1), pp. 252-7.
Benito, J.M., Godfrey, L., Kojima, K., Hogdal, L., Wunderlich, M., Geng, H., Marzo, I., Harutyunyan, K.G., Golfman, L., North, P., Kerry, J., Ballabio, E., Chonghaile, T.N., Gonzalo, O., Qiu, Y., Jeremias, I., Debose, L., O'Brien, E., Ma, H., Zhou, P., Jacamo, R., Park, E., Coombes, K.R., Zhang, N., Thomas, D.A., O'Brien, S., Kantarjian, H.M., Leverson, J.D., Kornblau, S.M., Andreeff, M., Muschen, M., Zweidler-McKay, P.A., Mulloy, J.C., Letai, A., Milne, T.A. and Konopleva, M. (2015) 'MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199', *Cell Rep*, 13(12), pp. 2715-27.
Bergsagel, P.L., Kuehl, W.M., Zhan, F., Sawyer, J., Barlogie, B. and Shaughnessy, J., Jr. (2005) 'Cyclin D dysregulation: an early and unifying pathogenic event in multiple myeloma', *Blood*, 106(1), pp. 296-303.
Bernardi, R., Grisendi, S. and Pandolfi, P.P. (2002) 'Modelling haematopoietic malignancies in the mouse and therapeutical implications', *Oncogene*, 21(21), pp. 3445-58.
Bernstein, D.L., Le Lay, J.E., Ruano, E.G. and Kaestner, K.H. (2015) 'TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts', *J Clin Invest*, 125(5), pp. 1998-2006.
Bolognesi, D.P. (1993) 'Human immunodeficiency virus vaccines', *Adv Virus Res*, 42, pp. 103-48.
Bond, J., Jones, C., Haughton, M., DeMicco, C., Kipling, D. and Wynford-Thomas, D. (2004) 'Direct evidence from siRNA-directed "knock down" that p16(INK4a) is required for human fibroblast senescence and for limiting ras-induced epithelial cell proliferation', *Exp Cell Res*, 292(1), pp. 151-6.

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.
Brueckner, L.M., Hess, E.M., Schwab, M. and Savelyeva, L. (2013) 'Instability at the FRA8I common fragile site disrupts the genomic integrity of the KIAA0146, CEBPD and PRKDC genes in colorectal cancer', *Cancer Lett*, 336(1), pp. 85-95.
Buchschacher, G.L., Jr. (2001) 'Introduction to retroviruses and retroviral vectors', *Somat Cell Mol Genet*, 26(1-6), pp. 1-11.
Buck, M., Turler, H. and Chojkier, M. (1994) 'LAP (NF-IL-6), a tissue-specific transcriptional activator, is an inhibitor of hepatoma cell proliferation', *EMBO J*, 13(4), pp. 851-60.
Burmeister, T., Schwartz, S., Bartram, C.R., Gokbuget, N., Hoelzer, D. and Thiel, E. (2008) 'Patients' age and BCR-ABL frequency in adult B-precursor ALL: a retrospective analysis from the GMALL study group', *Blood*, 112(3), pp. 918-9.
Calkhoven, C.F., Muller, C. and Leutz, A. (2000) 'Translational control of C/EBPalpha and C/EBPbeta isoform expression', *Genes Dev*, 14(15), pp. 1920-32.
Cancer Research Worldwide Cancer Stats (2011). Available at: http://publications.cancerresearchuk.org/publicationformat/formatstats/statsworldwide20051.html.
Cao, X., Tanis, K.Q., Koleske, A.J. and Colicelli, J. (2008) 'Enhancement of ABL kinase catalytic efficiency by a direct binding regulator is independent of other regulatory mechanisms', *J Biol Chem*, 283(46), pp. 31401-7.
Carroll, M., Ohno-Jones, S., Tamura, S., Buchdunger, E., Zimmermann, J., Lydon, N.B., Gilliland, D.G. and Druker, B.J. (1997) 'CGP 57148, a tyrosine kinase inhibitor, inhibits the growth of cells expressing BCR-ABL, TEL-ABL, and TEL-PDGFR fusion proteins', *Blood*, 90(12), pp. 4947-52.
Case-Gould, M.J. (1984) *Endotoxin in Vertebrate Cell Culture: Its Measurement and Significance. In Uses and Standardization of Vertebrate Cell Lines*.
Cazzaniga, G., van Delft, F.W., Lo Nigro, L., Ford, A.M., Score, J., Iacobucci, I., Mirabile, E., Taj, M., Colman, S.M., Biondi, A. and Greaves, M. (2011) 'Developmental origins and impact of BCR-ABL1 fusion and IKZF1 deletions in monozygotic twins with Ph+ acute lymphoblastic leukemia', *Blood*, 118(20), pp. 5559-64.
Chalandon, Y., Jiang, X., Hazlewood, G., Loutet, S., Conneally, E., Eaves, A. and Eaves, C. (2002) 'Modulation of p210(BCR-ABL) activity in transduced primary human hematopoietic cells controls lineage programming', *Blood*, 99(9), pp. 3197-204.
Chao, M.P., Alizadeh, A.A., Tang, C., Jan, M., Weissman-Tsukamoto, R., Zhao, F., Park, C.Y., Weissman, I.L. and Majeti, R. (2011) 'Therapeutic antibody targeting of CD47 eliminates human acute lymphoblastic leukemia', *Cancer Res*, 71(4), pp. 1374-84.
Chapiro, E., Radford-Weiss, I., Cung, H.A., Dastugue, N., Nadal, N., Taviaux, S., Barin, C., Struski, S., Talmant, P., Vandenberghe, P., Mozziconacci, M.J., Tigaud, I., Lefebvre, C., Penther, D., Bastard, C., Lippert, E., Mugneret, F., Romana, S., Bernard, O.A., Harrison, C.J., Russell, L.J., Nguyen-Khac, F. and Groupe Francophone de Cytogenetique, H. (2013) 'Chromosomal translocations involving the IGH@ locus in B-cell precursor acute lymphoblastic leukemia: 29 new cases and a review of the literature', *Cancer Genet*, 206(5), pp. 162-73.

Chapiro, E., Russell, L., Lainey, E., Kaltenbach, S., Ragu, C., Della-Valle, V., Hanssens, K., Macintyre, E.A., Radford-Weiss, I., Delabesse, E., Cave, H., Mercher, T., Harrison, C.J., Nguyen-Khac, F., Dubreuil, P. and Bernard, O.A. (2010) 'Activating mutation in the TSLPR gene in B-cell precursor lymphoblastic leukemia', *Leukemia*, 24(3), pp. 642-645.
Chapiro, E., Russell, L., Radford-Weiss, I., Bastard, C., Lessard, M., Struski, S., Cave, H., Fert-Ferrer, S., Barin, C., Maarek, O., Della-Valle, V., Strefford, J.C., Berger, R., Harrison, C.J., Bernard, O.A., Nguyen-Khac, F. and the Groupe Francophone de Cytogénétique, H. (2006) 'Overexpression of CEBPA resulting from the translocation t(14;19)(q32;q13) of human precursor B acute lymphoblastic leukemia', *Blood*, 108(10), pp. 3560-3563.
Chen, F.L., Yang, Z.H., Liu, Y., Li, L.X., Liang, W.C., Wang, X.C., Zhou, W.B., Yang, Y.H. and Hu, R.M. (2008a) 'Berberine inhibits the expression of TNFalpha, MCP-1, and IL-6 in AcLDL-stimulated macrophages through PPARgamma pathway', *Endocrine*, 33(3), pp. 331-7.
Chen, S., Dumitrescu, T.P., Smithgall, T.E. and Engen, J.R. (2008b) 'Abl N-terminal cap stabilization of SH3 domain dynamics', *Biochemistry*, 47(21), pp. 5795-803.
Cheng, Y. and Li, G. (2012) 'Role of the ubiquitin ligase Fbw7 in cancer progression', *Cancer Metastasis Rev*, 31(1-2), pp. 75-87.
Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D. and Jaenisch, R. (2000) 'Retroviral expression in embryonic stem cells and hematopoietic stem cells', *Mol Cell Biol*, 20(20), pp. 7419-26.
Chiaretti, S. and Foa, R. (2009) 'T-cell acute lymphoblastic leukemia', *Haematologica*, 94(2), pp. 160-2.
Chiarini, F., Lonetti, A., Evangelisti, C., Buontempo, F., Orsini, E., Evangelisti, C., Cappellini, A., Neri, L.M., McCubrey, J.A. and Martelli, A.M. (2016) 'Advances in understanding the acute lymphoblastic leukemia bone marrow microenvironment: From biology to therapeutic targeting', *Biochim Biophys Acta*, 1863(3), pp. 449-63.
Chiorazzi, N., Rai, K.R. and Ferrarini, M. (2005) 'Chronic lymphocytic leukemia', *N Engl J Med*, 352(8), pp. 804-15.
Chiu, C.H., Lin, W.D., Huang, S.Y. and Lee, Y.H. (2004) 'Effect of a C/EBP gene replacement on mitochondrial biogenesis in fat cells', *Genes Dev*, 18(16), pp. 1970-5.
Chumakov, A.M., Grillier, I., Chumakova, E., Chih, D., Slater, J. and Koeffler, H.P. (1997) 'Cloning of the novel human myeloid-cell-specific C/EBP-epsilon transcription factor', *Mol Cell Biol*, 17(3), pp. 1375-86.
Churchman, M.L., Low, J., Qu, C., Paietta, E.M., Kasper, L.H., Chang, Y., Payne-Turner, D., Althoff, M.J., Song, G., Chen, S.C., Ma, J., Rusch, M., McGoldrick, D., Edmonson, M., Gupta, P., Wang, Y.D., Caufield, W., Freeman, B., Li, L., Panetta, J.C., Baker, S., Yang, Y.L., Roberts, K.G., McCastlain, K., Iacobucci, I., Peters, J.L., Centonze, V.E., Notta, F., Dobson, S.M., Zandi, S., Dick, J.E., Janke, L., Peng, J., Kodali, K., Pagala, V., Min, J., Mayasundari, A., Williams, R.T., Willman, C.L., Rowe, J., Luger, S., Dickins, R.A., Guy, R.K., Chen, T. and Mullighan, C.G. (2015) 'Efficacy of Retinoids in IKZF1-Mutated BCR-ABL1 Acute Lymphoblastic Leukemia', *Cancer Cell*, 28(3), pp. 343-56.
Clappier, E., Auclerc, M.F., Rapion, J., Bakkus, M., Caye, A., Khemiri, A., Giroux, C., Hernandez, L., Kabongo, E., Savola, S., Leblanc, T., Yakouben, K., Plat, G., Costa, V., Ferster, A., Girard, S., Fenneteau, O., Cayuela, J.M., Sigaux, F., Dastugue, N., Suciu, S., Benoit, Y., Bertrand, Y., Soulier, J. and Cave, H. (2014) 'An intragenic ERG deletion is a marker of an oncogenic

subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions', *Leukemia*, 28(1), pp. 70-7.
Classon, M. and Harlow, E. (2002) 'The retinoblastoma tumour suppressor in development and cancer', *Nat Rev Cancer*, 2(12), pp. 910-7.
Cobaleda, C., Schebesta, A., Delogu, A. and Busslinger, M. (2007) 'Pax5: the guardian of B cell identity and function', *Nat Immunol*, 8(5), pp. 463-70.
Codrington, R., O'Connor, H.E., Jalali, G.R., Carrara, P., Papaioannou, M., Hart, S.M., Hoffbrand, A.V., Potter, M., Prentice, H.G., Harrison, C.J. and Foroni, L. (2000) 'Analysis of ETV6/AML1 abnormalities in acute lymphoblastic leukaemia: incidence, alternative spliced forms and minimal residual disease value', *Br J Haematol*, 111(4), pp. 1071-9.
Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1997) 'The Interactions of Retroviruses and their Hosts', in Coffin, J.M., Hughes, S.H. and Varmus, H.E. (eds.) *Retroviruses*. Cold Spring Harbor (NY).
Cohen, T.V., Klarmann, K.D., Sakchaisri, K., Cooper, J.P., Kuhns, D., Anver, M., Johnson, P.F., Williams, S.C., Keller, J.R. and Stewart, C.L. (2008) 'The lamin B receptor under transcriptional control of C/EBPepsilon is required for morphological but not functional maturation of neutrophils', *Hum Mol Genet*, 17(19), pp. 2921-33.
Colicelli, J. (2010) 'ABL tyrosine kinases: evolution of function, regulation, and specificity', *Sci Signal*, 3(139), p. re6.
Colotta, F., Allavena, P., Sica, A., Garlanda, C. and Mantovani, A. (2009) 'Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability', *Carcinogenesis*, 30(7), pp. 1073-81.
Conter, V., Bartram, C.R., Valsecchi, M.G., Schrauder, A., Panzer-Grumayer, R., Moricke, A., Arico, M., Zimmermann, M., Mann, G., De Rossi, G., Stanulla, M., Locatelli, F., Basso, G., Niggli, F., Barisone, E., Henze, G., Ludwig, W.D., Haas, O.A., Cazzaniga, G., Koehler, R., Silvestri, D., Bradtke, J., Parasole, R., Beier, R., van Dongen, J.J., Biondi, A. and Schrappe, M. (2010) 'Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study', *Blood*, 115(16), pp. 3206-14.
Cooper, C., Henderson, A., Artandi, S., Avitahl, N. and Calame, K. (1995) 'Ig/EBP (C/EBPγ) is a transdominant negative inhibitor of C/EBP family transcriptional activators', *Nucleic Acids Res*, 23(21), pp. 4371-4377.
Cooper, L.A., Gutman, D.A., Chisolm, C., Appin, C., Kong, J., Rong, Y., Kurc, T., Van Meir, E.G., Saltz, J.H., Moreno, C.S. and Brat, D.J. (2012) 'The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma', *Am J Pathol*, 180(5), pp. 2108-19.
D'Souza, V., Dey, A., Habib, D. and Summers, M.F. (2004) 'NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus', *J Mol Biol*, 337(2), pp. 427-42.
Dang, J., Wei, L., de Ridder, J., Su, X., Rust, A.G., Roberts, K.G., Payne-Turner, D., Cheng, J., Ma, J., Qu, C., Wu, G., Song, G., Huether, R.G., Schulman, B., Janke, L., Zhang, J., Downing, J.R., van der Weyden, L., Adams, D.J. and Mullighan, C.G. (2015) 'PAX5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia', *Blood*, 125(23), pp. 3609-17.
Daniel, R. and Smith, J.A. (2008) 'Integration site selection by retroviral vectors: molecular mechanism and clinical consequences', *Hum Gene Ther*, 19(6), pp. 557-68.

de Boer, C.J., Kluin-Nelemans, J.C., Dreef, E., Kester, M.G., Kluin, P.M., Schuuring, E. and van Krieken, J.H. (1996) 'Involvement of the CCND1 gene in hairy cell leukemia', *Ann Oncol*, 7(3), pp. 251-6.
De Braekeleer, E., Douet-Guilbert, N., Morel, F., Le Bris, M.J., Basinko, A. and De Braekeleer, M. (2012) 'ETV6 fusion genes in hematological malignancies: a review', *Leuk Res*, 36(8), pp. 945-61.
de Oliveira Cavassin, G.G., De Lucca, F.L., Delgado Andre, N., Covas, D.T., Pelegrinelli Fungaro, M.H., Voltarelli, J.C. and Watanabe, M.A. (2004) 'Molecular investigation of the stromal cell-derived factor-1 chemokine in lymphoid leukemia and lymphoma patients from Brazil', *Blood Cells Mol Dis*, 33(1), pp. 90-3.
de Oliveira, C.E., Cavassin, G.G., Perim Ade, L., Nasser, T.F., de Oliveira, K.B., Fungaro, M.H., Carneiro, J.L. and Watanabe, M.A. (2007) 'Stromal cell-derived factor-1 chemokine gene variant in blood donors and chronic myelogenous leukemia patients', *J Clin Lab Anal*, 21(1), pp. 49-54.
de Vasconcellos, J.F., Laranjeira, A.B., Zanchin, N.I., Otubo, R., Vaz, T.H., Cardoso, A.A., Brandalise, S.R. and Yunes, J.A. (2011) 'Increased CCL2 and IL-8 in the bone marrow microenvironment in acute lymphoblastic leukemia', *Pediatr Blood Cancer*, 56(4), pp. 568-77.
DeKoter, R.P. and Singh, H. (2000) 'Regulation of B lymphocyte and macrophage development by graded expression of PU.1', *Science*, 288(5470), pp. 1439-41.
Den Boer, M.L., van Slegtenhorst, M., De Menezes, R.X., Cheok, M.H., Buijs-Gladdines, J.G., Peters, S.T., Van Zutven, L.J., Beverloo, H.B., Van der Spek, P.J., Escherich, G., Horstmann, M.A., Janka-Schaub, G.E., Kamps, W.A., Evans, W.E. and Pieters, R. (2009) 'A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study', *Lancet Oncol*, 10(2), pp. 125-34.
Dias, S., Mansson, R., Gurbuxani, S., Sigvardsson, M. and Kee, B.L. (2008) 'E2A proteins promote development of lymphoid-primed multipotent progenitors', *Immunity*, 29(2), pp. 217-27.
Din, O.S. and Woll, P.J. (2008) 'Treatment of gastrointestinal stromal tumor: focus on imatinib mesylate', *Ther Clin Risk Manag*, 4(1), pp. 149-62.
Dokter, W.H., Tuyt, L., Sierdsema, S.J., Esselink, M.T. and Vellenga, E. (1995) 'The spontaneous expression of interleukin-1 beta and interleukin-6 is associated with spontaneous expression of AP-1 and NF-kappa B transcription factor in acute myeloblastic leukemia cells', *Leukemia*, 9(3), pp. 425-32.
Dores, G.M., Devesa, S.S., Curtis, R.E., Linet, M.S. and Morton, L.M. (2012) 'Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007', *Blood*, 119(1), pp. 34-43.
Draper, G.J., Kroll, M.E. and Stiller, C.A. (1994) 'Childhood cancer', *Cancer Surv*, 19-20, pp. 493-517.
Drexler, H.G., MacLeod, R.A., Borkhardt, A. and Janssen, J.W. (1995) 'Recurrent chromosomal translocations and fusion genes in leukemia-lymphoma cell lines', *Leukemia*, 9(3), pp. 480-500.
Dropulic, B. (2011) 'Lentiviral vectors: their molecular design, safety, and use in laboratory and preclinical research', *Hum Gene Ther*, 22(6), pp. 649-57.
Duprez, E., Wagner, K., Koch, H. and Tenen, D.G. (2003) 'C/EBPbeta: a major PML-RARA-responsive gene in retinoic acid-induced differentiation of APL cells', *EMBO J*, 22(21), pp. 5806-16.

Eaton, E.M. and Sealy, L. (2003) 'Modification of CCAAT/enhancer-binding protein-beta by the small ubiquitin-like modifier (SUMO) family members, SUMO-2 and SUMO-3', *J Biol Chem*, 278(35), pp. 33416-21.
Ehrhardt, H., Fulda, S., Fuhrer, M., Debatin, K.M. and Jeremias, I. (2004) 'Betulinic acid-induced apoptosis in leukemia cells', *Leukemia*, 18(8), pp. 1406-12.
Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A., Laks, E., Biele, J., Shumansky, K., Rosner, J., McPherson, A., Nielsen, C., Roth, A.J., Lefebvre, C., Bashashati, A., de Souza, C., Siu, C., Aniba, R., Brimhall, J., Oloumi, A., Osako, T., Bruna, A., Sandoval, J.L., Algara, T., Greenwood, W., Leung, K., Cheng, H., Xue, H., Wang, Y., Lin, D., Mungall, A.J., Moore, R., Zhao, Y., Lorette, J., Nguyen, L., Huntsman, D., Eaves, C.J., Hansen, C., Marra, M.A., Caldas, C., Shah, S.P. and Aparicio, S. (2015) 'Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution', *Nature*, 518(7539), pp. 422-6.
Ensor, H.M., Schwab, C., Russell, L.J., Richards, S.M., Morrison, H., Masic, D., Jones, L., Kinsey, S.E., Vora, A.J., Mitchell, C.D., Harrison, C.J. and Moorman, A.V. (2011) 'Demographic, clinical, and outcome features of children with acute lymphoblastic leukemia and CRLF2 deregulation: results from the MRC ALL97 clinical trial', *Blood*, 117(7), pp. 2129-36.
Enssle, J., Jordan, I., Mauer, B. and Rethwilm, A. (1996) 'Foamy virus reverse transcriptase is expressed independently from the Gag protein', *Proc Natl Acad Sci U S A*, 93(9), pp. 4137-41.
Estrov, Z., Manna, S.K., Harris, D., Van, Q., Estey, E.H., Kantarjian, H.M., Talpaz, M. and Aggarwal, B.B. (1999) 'Phenylarsine oxide blocks interleukin-1beta-induced activation of the nuclear transcription factor NF-kappaB, inhibits proliferation, and induces apoptosis of acute myelogenous leukemia cells', *Blood*, 94(8), pp. 2844-53.
Eyre, T., Schwab, C.J., Kinstrie, R., McGuire, A.K., Strefford, J., Peniket, A., Mead, A., Littlewood, T., Holyoake, T.L., Copland, M., Moorman, A.V., Harrison, C.J. and Vyas, P. (2012) 'Episomal amplification of NUP214-ABL1 fusion gene in B-cell acute lymphoblastic leukemia', *Blood*, 120(22), pp. 4441-3.
Faderl, S., Talpaz, M., Estrov, Z. and Kantarjian, H.M. (1999) 'Chronic myelogenous leukemia: biology and therapy', *Ann Intern Med*, 131(3), pp. 207-19.
Familiades, J., Bousquet, M., Lafage-Pochitaloff, M., Bene, M.C., Beldjord, K., De Vos, J., Dastugue, N., Coyaud, E., Struski, S., Quelen, C., Prade-Houdellier, N., Dobbelstein, S., Cayuela, J.M., Soulier, J., Grardel, N., Preudhomme, C., Cave, H., Blanchet, O., Lheritier, V., Delannoy, A., Chalandon, Y., Ifrah, N., Pigneux, A., Brousset, P., Macintyre, E.A., Huguet, F., Dombret, H., Broccardo, C. and Delabesse, E. (2009) 'PAX5 mutations occur frequently in adult B-cell progenitor acute lymphoblastic leukemia and PAX5 haploinsufficiency is associated with BCR-ABL1 and TCF3-PBX1 fusion genes: a GRAALL study', *Leukemia*, 23(11), pp. 1989-98.
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.
Felice, B., Cattoglio, C., Cittaro, D., Testa, A., Miccio, A., Ferrari, G., Luzi, L., Recchia, A. and Mavilio, F. (2009) 'Transcription factor binding sites are genetic determinants of retroviral integration in the human genome', *PLoS One*, 4(2), p. e4571.

Felice, M.S., Gallego, M.S., Alonso, C.N., Alfaro, E.M., Guitter, M.R., Bernasconi, A.R., Rubio, P.L., Zubizarreta, P.A. and Rossi, J.G. (2011) 'Prognostic impact of t(1;19)/ TCF3-PBX1 in childhood acute lymphoblastic leukemia in the context of Berlin-Frankfurt-Munster-based protocols', *Leuk Lymphoma*, 52(7), pp. 1215-21.
Fenrick, R., Amann, J.M., Lutterbach, B., Wang, L., Westendorf, J.J., Downing, J.R. and Hiebert, S.W. (1999) 'Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein', *Mol Cell Biol*, 19(10), pp. 6566-74.
Ferretti, E., Cocco, C., Airoldi, I. and Pistoia, V. (2012) 'Targeting acute myeloid leukemia cells with cytokines', *J Leukoc Biol*, 92(3), pp. 567-75.
Fier, K., Slámová, L., Dobiášová, A., Starková, J., Froňková, E., Ž Aliová, M., Polgárová, K., Figueroa, M.E., Kalina, T., Zuna, J., Trka, J., Starý, J., Hruš ák, O. and Mejstříková, E. (2014) 'Molecular Background of BCP-ALL Cases with an Early Switch to Monocytic Lineage', *Blood*, 124(21), pp. 3562-3562.
Fong, S., Itahana, Y., Sumida, T., Singh, J., Coppe, J.P., Liu, Y., Richards, P.C., Bennington, J.L., Lee, N.M., Debs, R.J. and Desprez, P.Y. (2003) 'Id-1 as a molecular target in therapy for breast cancer cell invasion and metastasis', *Proc Natl Acad Sci U S A*, 100(23), pp. 13543-8.
Ford, A.M., Ridge, S.A., Cabrera, M.E., Mahmoud, H., Steel, C.M., Chan, L.C. and Greaves, M. (1993) 'In utero rearrangements in the trithorax-related oncogene in infant leukaemias', *Nature*, 363(6427), pp. 358-360.
Forestier, E., Gauffin, F., Andersen, M.K., Autio, K., Borgstrom, G., Golovleva, I., Gustafsson, B., Heim, S., Heinonen, K., Heyman, M., Hovland, R., Johannsson, J.H., Kerndrup, G., Rosenquist, R., Schoumans, J., Swolin, B., Johansson, B., Nordgren, A., Nordic Society of Pediatric, H., Oncology, Swedish Cytogenetic Leukemia Study, G. and Group, N.L.C.S. (2008) 'Clinical and cytogenetic features of pediatric dic(9;20)(p13.2;q11.2)-positive B-cell precursor acute lymphoblastic leukemias: a Nordic series of 24 cases and review of the literature', *Genes Chromosomes Cancer*, 47(2), pp. 149-58.
Fornace, A.J., Jr., Nebert, D.W., Hollander, M.C., Luethy, J.D., Papathanasiou, M., Fargnoli, J. and Holbrook, N.J. (1989) 'Mammalian genes coordinately regulated by growth arrest signals and DNA-damaging agents', *Mol Cell Biol*, 9(10), pp. 4196-203.
Freed, E.O. (2001) 'HIV-1 replication', *Somat Cell Mol Genet*, 26(1-6), pp. 13-33.
Frelin, C., Imbert, V., Griessinger, E., Peyron, A.C., Rochet, N., Philip, P., Dageville, C., Sirvent, A., Hummelsberger, M., Berard, E., Dreano, M., Sirvent, N. and Peyron, J.F. (2005) 'Targeting NF-kappaB activation via pharmacologic inhibition of IKK2-induced apoptosis of human acute myeloid leukemia cells', *Blood*, 105(2), pp. 804-11.
Freytag, S.O. and Geddes, T.J. (1992) 'Reciprocal regulation of adipogenesis by Myc and C/EBP alpha', *Science*, 256(5055), pp. 379-82.
Friedman, A.D. (2007) 'Transcriptional control of granulocyte and monocyte development', *Oncogene*, 26(47), pp. 6816-28.
Fröhling, S., Schlenk, R.F., Stolze, I., Bihlmayr, J., Benner, A., Kreitmeier, S., Tobis, K., Döhner, H. and Döhner, K. (2004) 'CEBPA Mutations in Younger Adults With Acute Myeloid Leukemia and Normal Cytogenetics: Prognostic Relevance and Analysis of Cooperating Mutations', *Journal of Clinical Oncology*, 22(4), pp. 624-633.
Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E. and Busslinger, M. (2004) 'Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene', *Genes Dev*, 18(4), pp. 411-22.

Gales, D., Clark, C., Manne, U. and Samuel, T. (2013) 'The Chemokine CXCL8 in Carcinogenesis and Drug Response', *ISRN Oncol*, 2013, p. 859154.
Gambacorti-Passerini, C., Antolini, L., Mahon, F.X., Guilhot, F., Deininger, M., Fava, C., Nagler, A., Della Casa, C.M., Morra, E., Abruzzese, E., D'Emilio, A., Stagno, F., le Coutre, P., Hurtado-Monroy, R., Santini, V., Martino, B., Pane, F., Piccin, A., Giraldo, P., Assouline, S., Durosinmi, M.A., Leeksma, O., Pogliani, E.M., Puttini, M., Jang, E., Reiffers, J., Valsecchi, M.G. and Kim, D.W. (2011) 'Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib', *J Natl Cancer Inst*, 103(7), pp. 553-61.
Gaubatz, S., Lindeman, G.J., Ishida, S., Jakoi, L., Nevins, J.R., Livingston, D.M. and Rempel, R.E. (2000) 'E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control', *Mol Cell*, 6(3), pp. 729-35.
Gery, S., Gombart, A.F., Fung, Y.K. and Koeffler, H.P. (2004) 'C/EBPepsilon interacts with retinoblastoma and E2F1 during granulopoiesis', *Blood*, 103(3), pp. 828-35.
Gery, S., Tanosaki, S., Hofmann, W.K., Koppel, A. and Koeffler, H.P. (2005) 'C/EBPdelta expression in a BCR-ABL-positive cell line induces growth arrest and myeloid differentiation', *Oncogene*, 24(9), pp. 1589-97.
Giles, F.J., Krawczyk, J., O'Dwyer, M., Swords, R. and Freeman, C. (2014) 'The role of inflammation in leukaemia', *Adv Exp Med Biol*, 816, pp. 335-60.
Giordano, T.J. (2014) 'The cancer genome atlas research network: a sight to behold', *Endocr Pathol*, 25(4), pp. 362-5.
Golub, T.R., Barker, G.F., Bohlander, S.K., Hiebert, S.W., Ward, D.C., Bray-Ward, P., Morgan, E., Raimondi, S.C., Rowley, J.D. and Gilliland, D.G. (1995) 'Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia', *Proc Natl Acad Sci U S A*, 92(11), pp. 4917-21.
Gombart, A.F., Shiohara, M., Kwok, S.H., Agematsu, K., Komiyama, A. and Koeffler, H.P. (2001) 'Neutrophil-specific granule deficiency: homozygous recessive inheritance of a frameshift mutation in the gene encoding transcription factor CCAAT/enhancer binding protein–ε', *Blood*, 97(9), pp. 2561-2567.
Gomis, R.R., Alarcon, C., Nadal, C., Van Poznak, C. and Massague, J. (2006) 'C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells', *Cancer Cell*, 10(3), pp. 203-14.
Graham, F.L. and van der Eb, A.J. (1973) 'A new technique for the assay of infectivity of human adenovirus 5 DNA', *Virology*, 52(2), pp. 456-67.
Greaves, M. (2006) 'Infection, immune responses and the aetiology of childhood leukaemia', *Nat Rev Cancer*, 6(3), pp. 193-203.
Greaves, M.F. (1997) 'Aetiology of acute leukaemia', *Lancet*, 349(9048), pp. 344-9.
Greaves, M.F., Maia, A.T., Wiemels, J.L. and Ford, A.M. (2003) 'Leukemia in twins: lessons in natural history', *Blood*, 102(7), pp. 2321-33.
Greuber, E.K. and Pendergast, A.M. (2012) 'Abl family kinases regulate FcgammaR-mediated phagocytosis in murine macrophages', *J Immunol*, 189(11), pp. 5382-92.
Greuber, E.K., Smith-Pearson, P., Wang, J. and Pendergast, A.M. (2013) 'Role of ABL family kinases in cancer: from leukaemia to solid tumours', *Nat Rev Cancer*, 13(8), pp. 559-71.
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.
Grivennikov, S.I. and Karin, M. (2010) 'Dangerous liaisons: STAT3 and NF-kappaB collaboration and crosstalk in cancer', *Cytokine Growth Factor Rev*, 21(1), pp. 11-9.
Grotta, S., D'Elia, G., Scavelli, R., Genovese, S., Surace, C., Sirleto, P., Cozza, R., Romanzo, A., De Ioris, M.A., Valente, P., Tomaiuolo, A.C., Lepri, F.R., Franchin, T., Ciocca, L., Russo, S., Locatelli, F. and Angioni, A. (2015) 'Advantages of a next generation sequencing targeted approach for the molecular diagnosis of retinoblastoma', *BMC Cancer*, 15, p. 841.
Guerzoni, C., Bardini, M., Mariani, S.A., Ferrari-Amorotti, G., Neviani, P., Panno, M.L., Zhang, Y., Martinez, R., Perrotti, D. and Calabretta, B. (2006) 'Inducible activation of CEBPB, a gene negatively regulated by BCR/ABL, inhibits proliferation and promotes differentiation of BCR/ABL-expressing cells', *Blood*, 107(10), pp. 4080-4089.
Gulbagci, N.T., Li, L., Ling, B., Gopinadhan, S., Walsh, M., Rossner, M., Nave, K.A. and Taneja, R. (2009) 'SHARP1/DEC2 inhibits adipogenic differentiation by regulating the activity of C/EBP', *EMBO Rep*, 10(1), pp. 79-86.
Gutsch, R., Kandemir, J.D., Pietsch, D., Cappello, C., Meyer, J., Simanowski, K., Huber, R. and Brand, K. (2011) 'CCAAT/enhancer-binding protein beta inhibits proliferation in monocytic cells by affecting the retinoblastoma protein/E2F/cyclin E pathway but is not directly required for macrophage morphology', *J Biol Chem*, 286(26), pp. 22716-29.
Guzman, M.L., Neering, S.J., Upchurch, D., Grimes, B., Howard, D.S., Rizzieri, D.A., Luger, S.M. and Jordan, C.T. (2001) 'Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells', *Blood*, 98(8), pp. 2301-7.
Guzman, M.L., Swiderski, C.F., Howard, D.S., Grimes, B.A., Rossi, R.M., Szilvassy, S.J. and Jordan, C.T. (2002) 'Preferential induction of apoptosis for primary human leukemic stem cells', *Proc Natl Acad Sci U S A*, 99(25), pp. 16220-5.
Haas, S.C., Huber, R., Gutsch, R., Kandemir, J.D., Cappello, C., Krauter, J., Duyster, J., Ganser, A. and Brand, K. (2010) 'ITD- and FL-induced FLT3 signal transduction leads to increased C/EBPbeta-LIP expression and LIP/LAP ratio by different signalling modules', *Br J Haematol*, 148(5), pp. 777-90.
Hackanson, B., Bennett, K.L., Brena, R.M., Jiang, J., Claus, R., Chen, S.S., Blagitko-Dorfs, N., Maharry, K., Whitman, S.P., Schmittgen, T.D., Lubbert, M., Marcucci, G., Bloomfield, C.D. and Plass, C. (2008) 'Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia', *Cancer Res*, 68(9), pp. 3142-51.
Hagedorn, M., Delugin, M., Abraldes, I., Allain, N., Belaud-Rotureau, M.A., Turmo, M., Prigent, C., Loiseau, H., Bikfalvi, A. and Javerzat, S. (2007) 'FBXW7/hCDC4 controls glioma cell proliferation in vitro and is a prognostic marker for survival in glioblastoma patients', *Cell Div*, 2, p. 9.
Harewood, L., Robinson, H., Harris, R., Al-Obaidi, M.J., Jalali, G.R., Martineau, M., Moorman, A.V., Sumption, N., Richards, S., Mitchell, C. and Harrison, C.J. (2003) 'Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases', *Leukemia*, 17(3), pp. 547-553.
Harrison, C.J. (2009) 'Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia', *Br J Haematol*, 144(2), pp. 147-56.
Harrison, C.J. (2011) 'Acute lymphoblastic leukemia', *Clin Lab Med*, 31(4), pp. 631-47, ix.

Harrison, C.J. (2013) 'Targeting signaling pathways in acute lymphoblastic leukemia: new insights', *Hematology Am Soc Hematol Educ Program*, 2013, pp. 118-25.
Harrison, C.J., Moorman, A.V., Schwab, C., Carroll, A.J., Raetz, E.A., Devidas, M., Strehl, S., Nebral, K., Harbott, J., Teigler-Schlegel, A., Zimmerman, M., Dastuge, N., Baruchel, A., Soulier, J., Auclerc, M.F., Attarbaschi, A., Mann, G., Stark, B., Cazzaniga, G., Chilton, L., Vandenberghe, P., Forestier, E., Haltrich, I., Raimondi, S.C., Parihar, M., Bourquin, J.P., Tchinda, J., Haferlach, C., Vora, A., Hunger, S.P., Heerema, N.A., Haas, O.A. and Ponte di Legno International Workshop in Childhood Acute Lymphoblastic, L. (2014) 'An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome', *Leukemia*, 28(5), pp. 1015-21.
Harvey, R.C., Mullighan, C.G., Chen, I.M., Wharton, W., Mikhail, F.M., Carroll, A.J., Kang, H., Liu, W., Dobbin, K.K., Smith, M.A., Carroll, W.L., Devidas, M., Bowman, W.P., Camitta, B.M., Reaman, G.H., Hunger, S.P., Downing, J.R. and Willman, C.L. (2010a) 'Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia', *Blood*, 115(26), pp. 5312-5321.
Harvey, R.C., Mullighan, C.G., Wang, X., Dobbin, K.K., Davidson, G.S., Bedrick, E.J., Chen, I.M., Atlas, S.R., Kang, H., Ar, K., Wilson, C.S., Wharton, W., Murphy, M., Devidas, M., Carroll, A.J., Borowitz, M.J., Bowman, W.P., Downing, J.R., Relling, M., Yang, J., Bhojwani, D., Carroll, W.L., Camitta, B., Reaman, G.H., Smith, M., Hunger, S.P. and Willman, C.L. (2010b) 'Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome', *Blood*, 116(23), pp. 4874-84.
Hasemann, M.S., Lauridsen, F.K., Waage, J., Jakobsen, J.S., Frank, A.K., Schuster, M.B., Rapin, N., Bagger, F.O., Hoppe, P.S., Schroeder, T. and Porse, B.T. (2014) 'C/EBPalpha is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors', *PLoS Genet*, 10(1), p. e1004079.
Heavey, B., Charalambous, C., Cobaleda, C. and Busslinger, M. (2003) 'Myeloid lineage switch of Pax5 mutant but not wild-type B cell progenitors by C/EBP[alpha] and GATA factors', *Embo J*, 22(15), pp. 3887-3897.
Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B., Bacchetta, R., Tsalenko, A., Dellinger, D., Bruhn, L. and Porteus, M.H. (2015) 'Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells', *Nat Biotechnol*, 33(9), pp. 985-9.
Hertzberg, L., Vendramini, E., Ganmore, I., Cazzaniga, G., Schmitz, M., Chalker, J., Shiloh, R., Iacobucci, I., Shochat, C., Zeligson, S., Cario, G., Stanulla, M., Strehl, S., Russell, L.J., Harrison, C.J., Bornhauser, B., Yoda, A., Rechavi, G., Bercovich, D., Borkhardt, A., Kempski, H., te Kronnie, G., Bourquin, J.-P., Domany, E. and Izraeli, S. (2010) 'Down syndrome acute lymphoblastic leukemia, a highly heterogeneous disease in which aberrant expression of CRLF2 is associated with mutated JAK2: a report from the International BFM Study Group', *Blood*, 115(5), pp. 1006-1017.
Hess, J., Angel, P. and Schorpp-Kistner, M. (2004) 'AP-1 subunits: quarrel and harmony among siblings', *J Cell Sci*, 117(Pt 25), pp. 5965-73.

Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S., Imanishi, J., Akashi, K. and Tenen, D.G. (2006) 'C/EBPbeta is required for 'emergency' granulopoiesis', *Nat Immunol*, 7(7), pp. 732-9.
Hollis, A., Sperl, B., Graber, M. and Berg, T. (2012) 'The natural product betulinic acid inhibits C/EBP family transcription factors', *Chembiochem*, 13(2), pp. 302-7.
Holmfeldt, L., Wei, L., Diaz-Flores, E., Walsh, M., Zhang, J., Ding, L., Payne-Turner, D., Churchman, M., Andersson, A., Chen, S.C., McCastlain, K., Becksfort, J., Ma, J., Wu, G., Patel, S.N., Heatley, S.L., Phillips, L.A., Song, G., Easton, J., Parker, M., Chen, X., Rusch, M., Boggs, K., Vadodaria, B., Hedlund, E., Drenberg, C., Baker, S., Pei, D., Cheng, C., Huether, R., Lu, C., Fulton, R.S., Fulton, L.L., Tabib, Y., Dooling, D.J., Ochoa, K., Minden, M., Lewis, I.D., To, L.B., Marlton, P., Roberts, A.W., Raca, G., Stock, W., Neale, G., Drexler, H.G., Dickins, R.A., Ellison, D.W., Shurtleff, S.A., Pui, C.H., Ribeiro, R.C., Devidas, M., Carroll, A.J., Heerema, N.A., Wood, B., Borowitz, M.J., Gastier-Foster, J.M., Raimondi, S.C., Mardis, E.R., Wilson, R.K., Downing, J.R., Hunger, S.P., Loh, M.L. and Mullighan, C.G. (2013) 'The genomic landscape of hypodiploid acute lymphoblastic leukemia', *Nat Genet*, 45(3), pp. 242-52.
Hour, T.C., Lai, Y.L., Kuan, C.I., Chou, C.K., Wang, J.M., Tu, H.Y., Hu, H.T., Lin, C.S., Wu, W.J., Pu, Y.S., Sterneck, E. and Huang, A.M. (2010) 'Transcriptional up-regulation of SOD1 by CEBPD: a potential target for cisplatin resistant human urothelial carcinoma cells', *Biochem Pharmacol*, 80(3), pp. 325-34.
Hsu, C.L., King-Fleischman, A.G., Lai, A.Y., Matsumoto, Y., Weissman, I.L. and Kondo, M. (2006) 'Antagonistic effect of CCAAT enhancer-binding protein-alpha and Pax5 in myeloid or lymphoid lineage choice in common lymphoid progenitors', *Proc Natl Acad Sci U S A*, 103(3), pp. 672-7.
Huggins, C.J., Malik, R., Lee, S., Salotti, J., Thomas, S., Martin, N., Quinones, O.A., Alvord, W.G., Olanich, M.E., Keller, J.R. and Johnson, P.F. (2013) 'C/EBPgamma suppresses senescence and inflammatory gene expression by heterodimerizing with C/EBPbeta', *Mol Cell Biol*, 33(16), pp. 3242-58.
Ikezoe, T., Gery, S., Yin, D., O'Kelly, J., Binderup, L., Lemp, N., Taguchi, H. and Koeffler, H.P. (2005) 'CCAAT/enhancer-binding protein delta: a molecular target of 1,25-dihydroxyvitamin D3 in androgen-responsive prostate cancer LNCaP cells', *Cancer Res*, 65(11), pp. 4762-8.
Investigators, U.K.C.C.S. (2002) 'The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 2: gamma radiation', *Br J Cancer*, 86(11), pp. 1727-31.
Jaffe, E.S., Blattner, W.A., Blayney, D.W., Bunn, P.A., Jr., Cossman, J., Robert-Guroff, M. and Gallo, R.C. (1984) 'The pathologic spectrum of adult T-cell leukemia/lymphoma in the United States. Human T-cell leukemia/lymphoma virus-associated lymphoid malignancies', *Am J Surg Pathol*, 8(4), pp. 263-75.
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.
Jeffries, S.J., Jones, L., Harrison, C.J. and Russell, L.J. (2014) 'IGH@ translocations co-exist with other primary rearrangements in B-cell precursor acute lymphoblastic leukemia', *Haematologica*, 99(8), pp. 1334-42.
Jemal, A., Thomas, A., Murray, T. and Thun, M. (2002) 'Cancer Statistics, 2002', *CA Cancer J Clin*, 52(1), pp. 23-47.

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.
Johnson, P.F. (1993) 'Identification of C/EBP basic region residues involved in DNA sequence recognition and half-site spacing preference', *Molecular and Cellular Biology*, 13(11), pp. 6919-6930.
Johnson, P.F. (2005) 'Molecular stop signs: regulation of cell-cycle arrest by C/EBP transcription factors', *J Cell Sci*, 118(12), pp. 2545-2555.
Jones, L.C., Lin, M.L., Chen, S.S., Krug, U., Hofmann, W.K., Lee, S., Lee, Y.H. and Koeffler, H.P. (2002) 'Expression of C/EBPbeta from the C/ebpalpha gene locus is sufficient for normal hematopoiesis in vivo', *Blood*, 99(6), pp. 2032-6.
Juarez, J.G., Thien, M., Dela Pena, A., Baraz, R., Bradstock, K.F. and Bendall, L.J. (2009) 'CXCR4 mediates the homing of B cell progenitor acute lymphoblastic leukaemia cells to the bone marrow via activation of p38MAPK', *Br J Haematol*, 145(4), pp. 491-9.
Kathrein, K.L., Lorenz, R., Innes, A.M., Griffiths, E. and Winandy, S. (2005) 'Ikaros induces quiescence and T-cell differentiation in a leukemia cell line', *Mol Cell Biol*, 25(5), pp. 1645-54.
Kennedy, J.A. and Barabe, F. (2008) 'Investigating human leukemogenesis: from cell lines to in vivo models of human leukemia', *Leukemia*, 22(11), pp. 2029-40.
Khanna-Gupta, A. (2008) 'Sumoylation and the function of CCAAT enhancer binding protein alpha (C/EBP alpha)', *Blood Cells Mol Dis*, 41(1), pp. 77-81.
Kinlen, L.J. (1995) 'Epidemiological evidence for an infective basis in childhood leukaemia', *Br J Cancer*, 71(1), pp. 1-5.
Ko, C.Y., Chang, L.H., Lee, Y.C., Sterneck, E., Cheng, C.P., Chen, S.H., Huang, A.M., Tseng, J.T. and Wang, J.M. (2012) 'CCAAT/enhancer binding protein delta (CEBPD) elevating PTX3 expression inhibits macrophage-mediated phagocytosis of dying neuron cells', *Neurobiol Aging*, 33(2), pp. 422 e11-25.
Koblas, T., Zacharovova, K., Berkova, Z., Mindlova, M., Girman, P., Dovolilova, E., Karasova, L. and Saudek, F. (2007) 'Isolation and characterization of human CXCR4-positive pancreatic cells', *Folia Biol (Praha)*, 53(1), pp. 13-22.
Kornblau, S.M., McCue, D., Singh, N., Chen, W., Estrov, Z. and Coombes, K.R. (2010) 'Recurrent expression signatures of cytokines and chemokines are present and are independently prognostic in acute myelogenous leukemia and myelodysplasia', *Blood*, 116(20), pp. 4251-61.
Kowenz-Leutz, E. and Leutz, A. (1999) 'A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes', *Mol Cell*, 4(5), pp. 735-43.
Kuiper, R.P., Waanders, E., van der Velden, V.H., van Reijmersdal, S.V., Venkatachalam, R., Scheijen, B., Sonneveld, E., van Dongen, J.J., Veerman, A.J., van Leeuwen, F.N., van Kessel, A.G. and Hoogerbrugge, P.M. (2010) 'IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL', *Leukemia*, 24(7), pp. 1258-64.
Kuo, C.L., Chi, C.W. and Liu, T.Y. (2005) 'Modulation of apoptosis by berberine through inhibition of cyclooxygenase-2 and Mcl-1 expression in oral cancer cells', *In Vivo*, 19(1), pp. 247-52.
Kwon, K., Hutter, C., Sun, Q., Bilic, I., Cobaleda, C., Malin, S. and Busslinger, M. (2008) 'Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development', *Immunity*, 28(6), pp. 751-62.

Laiosa, C.V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. and Graf, T. (2006) 'Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors', *Immunity*, 25(5), pp. 731-44.
Laker, C., Meyer, J., Schopen, A., Friel, J., Heberlein, C., Ostertag, W. and Stocking, C. (1998) 'Host cis-mediated extinction of a retrovirus permissive for expression in embryonal stem cells during differentiation', *J Virol*, 72(1), pp. 339-48.
Lam, E.W., Glassford, J., Banerji, L., Thomas, N.S., Sicinski, P. and Klaus, G.G. (2000) 'Cyclin D3 compensates for loss of cyclin D2 in mouse B-lymphocytes activated via the antigen receptor and CD40', *J Biol Chem*, 275(5), pp. 3479-84.
Lammens, T., Li, J., Leone, G. and De Veylder, L. (2009) 'Atypical E2Fs: new players in the E2F transcription factor family', *Trends Cell Biol*, 19(3), pp. 111-8.
Lasorella, A., Benezra, R. and Iavarone, A. (2014) 'The ID proteins: master regulators of cancer stem cells and tumour aggressiveness', *Nat Rev Cancer*, 14(2), pp. 77-91.
Lekstrom-Himes, J. and Xanthopoulos, K.G. (1998) 'Biological Role of the CCAAT/Enhancer-binding Protein Family of Transcription Factors', *Journal of Biological Chemistry*, 273(44), pp. 28545-28548.
Leung, W., Campana, D., Yang, J., Pei, D., Coustan-Smith, E., Gan, K., Rubnitz, J.E., Sandlund, J.T., Ribeiro, R.C., Srinivasan, A., Hartford, C., Triplett, B.M., Dallas, M., Pillai, A., Handgretinger, R., Laver, J.H. and Pui, C.H. (2011) 'High success rate of hematopoietic cell transplantation regardless of donor source in children with very high-risk leukemia', *Blood*, 118(2), pp. 223-30.
Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C.C., Ecker, J.R. and Bushman, F.D. (2006) 'Retroviral DNA integration: viral and cellular determinants of target-site selection', *PLoS Pathog*, 2(6), p. e60.
Li, R. and Pendergast, A.M. (2011) 'Arg kinase regulates epithelial cell polarity by targeting beta1-integrin and small GTPase pathways', *Curr Biol*, 21(18), pp. 1534-42.
Li, Y., Schwab, C., Ryan, S.L., Papaemmanuil, E., Robinson, H.M., Jacobs, P., Moorman, A.V., Dyer, S., Borrow, J., Griffiths, M., Heerema, N.A., Carroll, A.J., Talley, P., Bown, N., Telford, N., Ross, F.M., Gaunt, L., McNally, R.J., Young, B.D., Sinclair, P., Rand, V., Teixeira, M.R., Joseph, O., Robinson, B., Maddison, M., Dastugue, N., Vandenberghe, P., Haferlach, C., Stephens, P.J., Cheng, J., Van Loo, P., Stratton, M.R., Campbell, P.J. and Harrison, C.J. (2014) 'Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia', *Nature*, 508(7494), pp. 98-102.
Lin, C.C., Lin, S.Y., Chung, J.G., Lin, J.P., Chen, G.W. and Kao, S.T. (2006) 'Down-regulation of cyclin B1 and up-regulation of Wee1 by berberine promotes entry of leukemia cells into the G2/M-phase of the cell cycle', *Anticancer Res*, 26(2A), pp. 1097-104.
Lin, F.T., MacDougald, O.A., Diehl, A.M. and Lane, M.D. (1993) 'A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity', *Proceedings of the National Academy of Sciences*, 90(20), pp. 9606-9610.
Lin, S., Tsai, S.C., Lee, C.C., Wang, B.W., Liou, J.Y. and Shyu, K.G. (2004) 'Berberine inhibits HIF-1alpha expression via enhanced proteolysis', *Mol Pharmacol*, 66(3), pp. 612-9.
Litvak, V., Ramsey, S.A., Rust, A.G., Zak, D.E., Kennedy, K.A., Lampano, A.E., Nykter, M., Shmulevich, I. and Aderem, A. (2009) 'Function of C/EBPdelta in a

regulatory circuit that discriminates between transient and persistent TLR4-induced signals', *Nat Immunol*, 10(4), pp. 437-43.
Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15(12), p. 550.
Lundin, C., Heldrup, J., Ahlgren, T., Olofsson, T. and Johansson, B. (2009) 'B-cell precursor t(8;14)(q11;q32)-positive acute lymphoblastic leukemia in children is strongly associated with Down syndrome or with a concomitant Philadelphia chromosome', *Eur J Haematol*, 82(1), pp. 46-53.
Ma, S., Pathak, S., Mandal, M., Trinh, L., Clark, M.R. and Lu, R. (2010) 'Ikaros and Aiolos inhibit pre-B-cell proliferation by directly suppressing c-Myc expression', *Mol Cell Biol*, 30(17), pp. 4149-58.
Majeti, R., Chao, M.P., Alizadeh, A.A., Pang, W.W., Jaiswal, S., Gibbs, K.D., Jr., van Rooijen, N. and Weissman, I.L. (2009) 'CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells', *Cell*, 138(2), pp. 286-99.
Mancini, M., Scappaticci, D., Cimino, G., Nanni, M., Derme, V., Elia, L., Tafuri, A., Vignetti, M., Vitale, A., Cuneo, A., Castoldi, G., Saglio, G., Pane, F., Mecucci, C., Camera, A., Specchia, G., Tedeschi, A., Di Raimondo, F., Fioritoni, G., Fabbiano, F., Marmont, F., Ferrara, F., Cascavilla, N., Todeschini, G., Nobile, F., Kropp, M.G., Leoni, P., Tabilio, A., Luppi, M., Annino, L., Mandelli, F. and Foà, R. (2005) 'A comprehensive genetic classification of adult acute lymphoblastic leukemia (ALL): analysis of the GIMEMA 0496 protocol', *Blood*, 105(9), pp. 3434-3441.
Mani, R.S. and Chinnaiyan, A.M. (2010) 'Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences', *Nat Rev Genet*, 11(12), pp. 819-29.
Mansson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thoren, L., Adolfsson, J., Buza-Vidas, N., Qian, H., Soneji, S., Enver, T., Sigvardsson, M. and Jacobsen, S.E. (2007) 'Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors', *Immunity*, 26(4), pp. 407-19.
Mantena, S.K., Sharma, S.D. and Katiyar, S.K. (2006) 'Berberine, a natural product, induces G1-phase cell cycle arrest and caspase-3-dependent apoptosis in human prostate carcinoma cells', *Mol Cancer Ther*, 5(2), pp. 296-308.
Marschalek, R. (2011) 'Mechanisms of leukemogenesis by MLL fusion proteins', *Br J Haematol*, 152(2), pp. 141-54.
Maude, S.L., Tasian, S.K., Vincent, T., Hall, J.W., Sheen, C., Roberts, K.G., Seif, A.E., Barrett, D.M., Chen, I.M., Collins, J.R., Mullighan, C.G., Hunger, S.P., Harvey, R.C., Willman, C.L., Fridman, J.S., Loh, M.L., Grupp, S.A. and Teachey, D.T. (2012) 'Targeting JAK1/2 and mTOR in murine xenograft models of Ph-like acute lymphoblastic leukemia', *Blood*, 120(17), pp. 3510-8.
McNagny, K.M., Sieweke, M.H., Doderlein, G., Graf, T. and Nerlov, C. (1998) 'Regulation of eosinophil-specific gene expression by a C/EBP-Ets complex and GATA-1', *EMBO J*, 17(13), pp. 3669-80.
Mechtcheriakova, D., Svoboda, M., Meshcheryakova, A. and Jensen-Jarolim, E. (2012) 'Activation-induced cytidine deaminase (AID) linking immunity, chronic inflammation, and cancer', *Cancer Immunol Immunother*, 61(9), pp. 1591-8.
Mellentin, J.D., Murre, C., Donlon, T.A., McCaw, P.S., Smith, S.D., Carroll, A.J., McDonald, M.E., Baltimore, D. and Cleary, M.L. (1989) 'The gene for enhancer

binding proteins E12/E47 lies at the t(1;19) breakpoint in acute leukemias', *Science*, 246(4928), pp. 379-82.
Messinger, Y.H., Higgins, R.R., Devidas, M., Hunger, S.P., Carroll, A.J. and Heerema, N.A. (2012) 'Pediatric acute lymphoblastic leukemia with a t(8;14)(q11.2;q32): B-cell disease with a high proportion of Down syndrome: a Children's Oncology Group study', *Cancer Genet*, 205(9), pp. 453-8.
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.
Miller, A.L., Wang, Y., Mooseker, M.S. and Koleske, A.J. (2004) 'The Abl-related gene (Arg) requires its F-actin-microtubule cross-linking activity to regulate lamellipodial dynamics during fibroblast adhesion', *J Cell Biol*, 165(3), pp. 407-19.
Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R. and Bushman, F.D. (2004) 'Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences', *PLoS Biol*, 2(8), p. E234.
Moorman, A.V. (2012a) 'The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia', *Blood Rev*, 26(3), pp. 123-135.
Moorman, A.V. (2012b) 'The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia', *Blood Rev*, 26(3), pp. 123-35.
Moorman, A.V., Chilton, L., Wilkinson, J., Ensor, H.M., Bown, N. and Proctor, S.J. (2010a) 'A population-based cytogenetic study of adults with acute lymphoblastic leukemia', *Blood*, 115(2), pp. 206-14.
Moorman, A.V., Enshaei, A., Schwab, C., Wade, R., Chilton, L., Elliott, A., Richardson, S., Hancock, J., Kinsey, S.E., Mitchell, C.D., Goulden, N., Vora, A. and Harrison, C.J. (2014) 'A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia', *Blood*, 124(9), pp. 1434-44.
Moorman, A.V., Ensor, H.M., Richards, S.M., Chilton, L., Schwab, C., Kinsey, S.E., Vora, A., Mitchell, C.D. and Harrison, C.J. (2010b) 'Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial', *Lancet Oncol*, 11(5), pp. 429-38.
Moorman, A.V., Harrison, C.J., Buck, G.A., Richards, S.M., Secker-Walker, L.M., Martineau, M., Vance, G.H., Cherry, A.M., Higgins, R.R., Fielding, A.K., Foroni, L., Paietta, E., Tallman, M.S., Litzow, M.R., Wiernik, P.H., Rowe, J.M., Goldstone, A.H. and Dewald, G.W. (2007a) 'Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial', *Blood*, 109(8), pp. 3189-97.

Moorman, A.V., Richards, S.M., Robinson, H.M., Strefford, J.C., Gibson, B.E.S., Kinsey, S.E., Eden, T.O.B., Vora, A.J., Mitchell, C.D., Harrison, C.J. and on behalf of the, U.K.M.R.C.N.C.R.I.C.L.W.P. (2007b) 'Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21)', *Blood*, 109(6), pp. 2327-2330.
Moorman, A.V., Robinson, H., Schwab, C., Richards, S.M., Hancock, J., Mitchell, C.D., Goulden, N., Vora, A. and Harrison, C.J. (2013) 'Risk-directed treatment intensification significantly reduces the risk of relapse among children and adolescents with acute lymphoblastic leukemia and intrachromosomal amplification of chromosome 21: a comparison of the MRC ALL97/99 and UKALL2003 trials', *J Clin Oncol*, 31(27), pp. 3389-96.
Mori, H., Colman, S.M., Xiao, Z., Ford, A.M., Healy, L.E., Donaldson, C., Hows, J.M., Navarrete, C. and Greaves, M. (2002) 'Chromosome translocations and covert leukemic clones are generated during normal fetal development', *Proc Natl Acad Sci U S A*, 99(12), pp. 8242-7.
Moricke, A., Zimmermann, M., Reiter, A., Henze, G., Schrauder, A., Gadner, H., Ludwig, W.D., Ritter, J., Harbott, J., Mann, G., Klingebiel, T., Zintl, F., Niemeyer, C., Kremens, B., Niggli, F., Niethammer, D., Welte, K., Stanulla, M., Odenwald, E., Riehm, H. and Schrappe, M. (2010) 'Long-term results of five consecutive trials in childhood acute lymphoblastic leukemia performed by the ALL-BFM study group from 1981 to 2000', *Leukemia*, 24(2), pp. 265-84.
Morosetti, R., Park, D.J., Chumakov, A.M., Grillier, I., Shiohara, M., Gombart, A.F., Nakamaki, T., Weinberg, K. and Koeffler, H.P. (1997) 'A Novel, Myeloid Transcription Factor, C/EBPε, Is Upregulated During Granulocytic, But Not Monocytic, Differentiation', *Blood*, 90(7), pp. 2591-2600.
Muchardt, C. and Yaniv, M. (2001) 'When the SWI/SNF complex remodels...the cell cycle', *Oncogene*, 20(24), pp. 3067-75.
Muller, C., Bremer, A., Schreiber, S., Eichwald, S. and Calkhoven, C.F. (2010) 'Nucleolar retention of a translational C/EBPalpha isoform stimulates rDNA transcription and cell size', *EMBO J*, 29(5), pp. 897-909.
Mullighan, C.G., Collins-Underwood, J.R., Phillips, L.A.A., Loudin, M.G., Liu, W., Zhang, J., Ma, J., Coustan-Smith, E., Harvey, R.C., Willman, C.L., Mikhail, F.M., Meyer, J., Carroll, A.J., Williams, R.T., Cheng, J., Heerema, N.A., Basso, G., Pession, A., Pui, C.-H., Raimondi, S.C., Hunger, S.P., Downing, J.R., Carroll, W.L. and Rabin, K.R. (2009a) 'Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia', *Nat Genet*, 41(11), pp. 1243-1246.
Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., Su, X., Pui, C.H., Relling, M.V., Evans, W.E., Shurtleff, S.A. and Downing, J.R. (2007) 'Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia', *Nature*, 446(7137), pp. 758-64.
Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.H., Relling, M.V., Shurtleff, S.A. and Downing, J.R. (2008a) 'BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros', *Nature*, 453(7191), pp. 110-4.
Mullighan, C.G., Phillips, L.A., Su, X., Ma, J., Miller, C.B., Shurtleff, S.A. and Downing, J.R. (2008b) 'Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia', *Science*, 322(5906), pp. 1377-80.
Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., Schulman, B.A., Harvey, R.C., Chen, I.M., Clifford, R.J., Carroll, W.L., Reaman, G., Bowman, W.P., Devidas, M., Gerhard, D.S., Yang,

W., Relling, M.V., Shurtleff, S.A., Campana, D., Borowitz, M.J., Pui, C.H., Smith, M., Hunger, S.P., Willman, C.L., Downing, J.R. and Children's Oncology, G. (2009b) 'Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia', *N Engl J Med*, 360(5), pp. 470-80.
Mulloy, J.C., Cammenga, J., Berguido, F.J., Wu, K., Zhou, P., Comenzo, R.L., Jhanwar, S., Moore, M.A. and Nimer, S.D. (2003) 'Maintaining the self-renewal and differentiation potential of human CD34+ hematopoietic cells using a single genetic element', *Blood*, 102(13), pp. 4369-76.
Munzert, G., Kirchner, D., Ottmann, O., Bergmann, L. and Schmid, R.M. (2004) 'Constitutive NF-kappab/Rel activation in philadelphia chromosome positive (Ph+) acute lymphoblastic leukemia (ALL)', *Leuk Lymphoma*, 45(6), pp. 1181-4.
Nakajima, H., Watanabe, N., Shibata, F., Kitamura, T., Ikeda, Y. and Handa, M. (2006) 'N-terminal region of CCAAT/enhancer-binding protein epsilon is critical for cell cycle arrest, apoptosis, and functional maturation during myeloid differentiation', *J Biol Chem*, 281(20), pp. 14494-502.
Nebral, K., Denk, D., Attarbaschi, A., Konig, M., Mann, G., Haas, O.A. and Strehl, S. (2009) 'Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia', *Leukemia*, 23(1), pp. 134-43.
Nemazee, D. (2006) 'Receptor editing in lymphocyte development and central tolerance', *Nat Rev Immunol*, 6(10), pp. 728-40.
Nerlov, C. (2004) 'C/EBP[alpha] mutations in acute myeloid leukaemias', *Nat Rev Cancer*, 4(5), pp. 394-400.
Nerlov, C. (2007) 'The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control', *Trends Cell Biol*, 17(7), pp. 318-24.
Nerlov, C. and Graf, T. (1998) 'PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors', *Genes Dev*, 12(15), pp. 2403-12.
Nerlov, C. and Ziff, E.B. (1995) 'CCAAT/enhancer binding protein-alpha amino acid motifs with dual TBP and TFIIB binding ability co-operate to activate transcription in both yeast and mammalian cells', *Embo J*, 14(17), pp. 4318-28.
Newman, J.R. and Keating, A.E. (2003) 'Comprehensive identification of human bZIP interactions with coiled-coil arrays', *Science*, 300(5628), pp. 2097-101.
Nowell, P.C. and Hungerford, D.A. (1960) 'Chromosome studies on normal and leukemic human leukocytes', *J Natl Cancer Inst*, 25, pp. 85-109.
Nutt, S.L., Urbanek, P., Rolink, A. and Busslinger, M. (1997) 'Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus', *Genes Dev*, 11(4), pp. 476-91.
O'Rourke, J.P., Newbound, G.C., Hutt, J.A. and DeWille, J. (1999) 'CCAAT/enhancer-binding protein delta regulates mammary epithelial cell G0 growth arrest and apoptosis', *J Biol Chem*, 274(23), pp. 16582-9.
Ohtani, N., Zebedee, Z., Huot, T.J., Stinson, J.A., Sugimoto, M., Ohashi, Y., Sharrocks, A.D., Peters, G. and Hara, E. (2001) 'Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence', *Nature*, 409(6823), pp. 1067-70.
Ossipow, V., Descombes, P. and Schibler, U. (1993) 'CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials', *Proceedings of the National Academy of Sciences*, 90(17), pp. 8219-8223.
Ottmann, O.G. and Pfeifer, H. (2009) 'Management of Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL)', *Hematology Am Soc Hematol Educ Program*, pp. 371-81.

Pabst, T., Mueller, B.U., Zhang, P., Radomska, H.S., Narravula, S., Schnittger, S., Behre, G., Hiddemann, W. and Tenen, D.G. (2001) 'Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia', *Nat Genet*, 27(3), pp. 263-70.
Pan, Y.C., Li, C.F., Ko, C.Y., Pan, M.H., Chen, P.J., Tseng, J.T., Wu, W.C., Chang, W.C., Huang, A.M., Sterneck, E. and Wang, J.M. (2010) 'CEBPD reverses RB/E2F1-mediated gene repression and participates in HMDB-induced apoptosis of cancer cells', *Clin Cancer Res*, 16(23), pp. 5770-80.
Papaemmanuil, E., Hosking, F.J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S.E., Lightfoot, T., Roman, E., Irving, J.A., Allan, J.M., Tomlinson, I.P., Taylor, M., Greaves, M. and Houlston, R.S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet*, 41(9), pp. 1006-10.
Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., Martincorena, I., Hinton, J., Gundem, G., van Delft, F.W., Nik-Zainal, S., Jones, D.R., Ramakrishna, M., Titley, I., Stebbings, L., Leroy, C., Menzies, A., Gamble, J., Robinson, B., Mudie, L., Raine, K., O'Meara, S., Teague, J.W., Butler, A.P., Cazzaniga, G., Biondi, A., Zuna, J., Kempski, H., Muschen, M., Ford, A.M., Stratton, M.R., Greaves, M. and Campbell, P.J. (2014) 'RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia', *Nat Genet*, 46(2), pp. 116-25.
Parkin, D.M., Stiller, C.A., Draper, G.J. and Bieber, C.A. (1988) 'The international incidence of childhood cancer', *Int J Cancer*, 42(4), pp. 511-20.
Parkin, S.E., Baer, M., Copeland, T.D., Schwartz, R.C. and Johnson, P.F. (2002) 'Regulation of CCAAT/Enhancer-binding Protein (C/EBP) Activator Proteins by Heterodimerization with C/EBPγ (Ig/EBP)', *Journal of Biological Chemistry*, 277(26), pp. 23563-23572.
Pastwa, E. and Blasiak, J. (2003) 'Non-homologous DNA end joining', *Acta Biochim Pol*, 50(4), pp. 891-908.
Paulsson, K., Forestier, E., Lilljebjorn, H., Heldrup, J., Behrendtz, M., Young, B.D. and Johansson, B. (2010) 'Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia', *Proc Natl Acad Sci U S A*, 107(50), pp. 21719-24.
Paulsson, K., Lilljebjorn, H., Biloglav, A., Olsson, L., Rissler, M., Castor, A., Barbany, G., Fogelstrand, L., Nordgren, A., Sjogren, H., Fioretos, T. and Johansson, B. (2015) 'The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia', *Nat Genet*, 47(6), pp. 672-6.
Pawar, S.A., Sarkar, T.R., Balamurugan, K., Sharan, S., Wang, J., Zhang, Y., Dowdy, S.F., Huang, A.M. and Sterneck, E. (2010) 'C/EBP{delta} targets cyclin D1 for proteasome-mediated degradation via induction of CDC27/APC3 expression', *Proc Natl Acad Sci U S A*, 107(20), pp. 9210-5.
Pham, T.P., Kwon, J. and Shin, J. (2011) 'Berberine exerts anti-adipogenic activity through up-regulation of C/EBP inhibitors, CHOP and DEC2', *Biochem Biophys Res Commun*, 413(2), pp. 376-82.
Pieraets, S., Cox, L., Gielen, O. and Cools, J. (2012) 'Development of a siRNA and shRNA screening system based on a kinase fusion protein', *RNA*, 18(6), pp. 1296-306.
Pierini, V., Nofrini, V., La Starza, R., Barba, G., Vitale, A., Di Raimondo, F., Matteucci, C., Crescenzi, B., Elia, L., Gorello, P., Storlazzi, C.T. and Mecucci, C. (2011) 'Double CEBPE-IGH rearrangement due to chromosome duplication

and cryptic insertion in an adult with B-cell acute lymphoblastic leukemia', *Cancer Genet*, 204(10), pp. 563-8.
Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., Hovi, L., LeBlanc, T., Szczepanski, T., Ferster, A., Janka, G., Rubnitz, J., Silverman, L., Stary, J., Campbell, M., Li, C.-K., Mann, G., Suppiah, R., Biondi, A., Vora, A. and Valsecchi, M.G. (2007) 'A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial', *The Lancet*, 370(9583), pp. 240-250.
Placencio, V.R. and DeClerck, Y.A. (2015) 'Plasminogen Activator Inhibitor-1 in Cancer: Rationale and Insight for Future Therapeutic Testing', *Cancer Res*, 75(15), pp. 2969-74.
Pless, O., Kowenz-Leutz, E., Knoblich, M., Lausen, J., Beyermann, M., Walsh, M.J. and Leutz, A. (2008) 'G9a-mediated lysine methylation alters the function of CCAAT/enhancer-binding protein-beta', *J Biol Chem*, 283(39), pp. 26357-63.
Podust, L.M., Krezel, A.M. and Kim, Y. (2001) 'Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA', *J Biol Chem*, 276(1), pp. 505-13.
Polager, S. and Ginsberg, D. (2008) 'E2F - at the crossroads of life and death', *Trends Cell Biol*, 18(11), pp. 528-35.
Poli, V., Mancini, F.P. and Cortese, R. (1990) 'IL-6DBP, a nuclear protein involved in interleukin-6 signal transduction, defines a new family of leucine zipper proteins related to C/EBP', *Cell*, 63(3), pp. 643-53.
Popernack, P.M., Truong, L.T., Kamphuis, M. and Henderson, A.J. (2001) 'Ectopic expression of CCAAT/enhancer binding protein beta (C/EBPbeta) in long-term bone marrow cultures induces granulopoiesis and alters stromal cell function', *J Hematother Stem Cell Res*, 10(5), pp. 631-42.
Preston, D.L., Kusumi, S., Tomonaga, M., Izumi, S., Ron, E., Kuramoto, A., Kamada, N., Dohy, H., Matsuo, T., Matsui, T. and et al. (1994) 'Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987', *Radiat Res*, 137(2 Suppl), pp. S68-97.
Preudhomme, C., Sagot, C., Boissel, N., Cayuela, J.-M., Tigaud, I., de Botton, S., Thomas, X., Raffoux, E., Lamandin, C., Castaigne, S., Fenaux, P. and Dombret, H. (2002) 'Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA)', *Blood*, 100(8), pp. 2717-2723.
Pui, C.H. and Howard, S.C. (2008) 'Current management and challenges of malignant disease in the CNS in paediatric leukaemia', *Lancet Oncol*, 9(3), pp. 257-68.
Pulte, D., Gondos, A. and Brenner, H. (2008) 'Improvements in survival of adults diagnosed with acute myeloblastic leukemia in the early 21st century', *Haematologica*, 93(4), pp. 594-600.
Radich, J.P., Dai, H., Mao, M., Oehler, V., Schelter, J., Druker, B., Sawyers, C., Shah, N., Stock, W., Willman, C.L., Friend, S. and Linsley, P.S. (2006) 'Gene expression changes associated with progression and response in chronic myeloid leukemia', *Proc Natl Acad Sci U S A*, 103(8), pp. 2794-9.
Radomska, H.S., Bassères, D.S., Zheng, R., Zhang, P., Dayaram, T., Yamamoto, Y., Sternberg, D.W., Lokker, N., Giese, N.A., Bohlander, S.K., Schnittger, S., Delmotte, M.-H., Davis, R.J., Small, D., Hiddemann, W., Gilliland, D.G. and Tenen, D.G. (2006) 'Block of C/EBPα function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations', *J Exp Med*, 203(2), pp. 371-381.

Raghavan, M., Lillington, D.M., Skoulakis, S., Debernardi, S., Chaplin, T., Foot, N.J., Lister, T.A. and Young, B.D. (2005) 'Genome-Wide Single Nucleotide Polymorphism Analysis Reveals Frequent Partial Uniparental Disomy Due to Somatic Recombination in Acute Myeloid Leukemias', *Cancer Res*, 65(2), pp. 375-378.
Ramji, D.P. and Foka, P. (2002) 'CCAAT/enhancer-binding proteins: structure, function and regulation', *Biochem J*, 365(Pt 3), pp. 561-75.
Rand, V., Parker, H., Russell, L.J., Schwab, C., Ensor, H., Irving, J., Jones, L., Masic, D., Minto, L., Morrison, H., Ryan, S., Robinson, H., Sinclair, P., Moorman, A.V., Strefford, J.C. and Harrison, C.J. (2011) 'Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia', *Blood*, 117(25), pp. 6848-6855.
Rayess, H., Wang, M.B. and Srivatsan, E.S. (2012) 'Cellular senescence and tumor suppressor gene p16', *Int J Cancer*, 130(8), pp. 1715-25.
Reddy, V.A., Iwama, A., Iotzova, G., Schulz, M., Elsasser, A., Vangala, R.K., Tenen, D.G., Hiddemann, W. and Behre, G. (2002) 'Granulocyte inducer C/EBPalpha inactivates the myeloid master regulator PU.1: possible role in lineage commitment decisions', *Blood*, 100(2), pp. 483-90.
Robert-Richard, E., Ged, C., Ortet, J., Santarelli, X., Lamrissi-Garcia, I., de Verneuil, H. and Mazurier, F. (2006) 'Human cell engraftment after busulfan or irradiation conditioning of NOD/SCID mice', *Haematologica*, 91(10), p. 1384.
Roberts, K.G., Li, Y., Payne-Turner, D., Harvey, R.C., Yang, Y.L., Pei, D., McCastlain, K., Ding, L., Lu, C., Song, G., Ma, J., Becksfort, J., Rusch, M., Chen, S.C., Easton, J., Cheng, J., Boggs, K., Santiago-Morales, N., Iacobucci, I., Fulton, R.S., Wen, J., Valentine, M., Cheng, C., Paugh, S.W., Devidas, M., Chen, I.M., Reshmi, S., Smith, A., Hedlund, E., Gupta, P., Nagahawatte, P., Wu, G., Chen, X., Yergeau, D., Vadodaria, B., Mulder, H., Winick, N.J., Larsen, E.C., Carroll, W.L., Heerema, N.A., Carroll, A.J., Grayson, G., Tasian, S.K., Moore, A.S., Keller, F., Frei-Jones, M., Whitlock, J.A., Raetz, E.A., White, D.L., Hughes, T.P., Guidry Auvil, J.M., Smith, M.A., Marcucci, G., Bloomfield, C.D., Mrozek, K., Kohlschmidt, J., Stock, W., Kornblau, S.M., Konopleva, M., Paietta, E., Pui, C.H., Jeha, S., Relling, M.V., Evans, W.E., Gerhard, D.S., Gastier-Foster, J.M., Mardis, E., Wilson, R.K., Loh, M.L., Downing, J.R., Hunger, S.P., Willman, C.L., Zhang, J. and Mullighan, C.G. (2014a) 'Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia', *N Engl J Med*, 371(11), pp. 1005-15.
Roberts, K.G., Morin, R.D., Zhang, J., Hirst, M., Zhao, Y., Su, X., Chen, S.C., Payne-Turner, D., Churchman, M.L., Harvey, R.C., Chen, X., Kasap, C., Yan, C., Becksfort, J., Finney, R.P., Teachey, D.T., Maude, S.L., Tse, K., Moore, R., Jones, S., Mungall, K., Birol, I., Edmonson, M.N., Hu, Y., Buetow, K.E., Chen, I.M., Carroll, W.L., Wei, L., Ma, J., Kleppe, M., Levine, R.L., Garcia-Manero, G., Larsen, E., Shah, N.P., Devidas, M., Reaman, G., Smith, M., Paugh, S.W., Evans, W.E., Grupp, S.A., Jeha, S., Pui, C.H., Gerhard, D.S., Downing, J.R., Willman, C.L., Loh, M., Hunger, S.P., Marra, M.A. and Mullighan, C.G. (2012) 'Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia', *Cancer Cell*, 22(2), pp. 153-66.
Roberts, K.G. and Mullighan, C.G. (2015) 'Genomics in acute lymphoblastic leukaemia: insights and treatment implications', *Nat Rev Clin Oncol*, 12(6), pp. 344-57.
Roberts, K.G., Pei, D., Campana, D., Payne-Turner, D., Li, Y., Cheng, C., Sandlund, J.T., Jeha, S., Easton, J., Becksfort, J., Zhang, J., Coustan-Smith, E.,

Raimondi, S.C., Leung, W.H., Relling, M.V., Evans, W.E., Downing, J.R., Mullighan, C.G. and Pui, C.H. (2014b) 'Outcomes of children with BCR-ABL1-like acute lymphoblastic leukemia treated with risk-directed therapy based on the levels of minimal residual disease', *J Clin Oncol*, 32(27), pp. 3012-20.
Robinson, H.M., Harrison, C.J., Moorman, A.V., Chudoba, I. and Strefford, J.C. (2007) 'Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage–fusion–bridge cycle', *Genes, Chromosomes and Cancer*, 46(4), pp. 318-326.
Roman, C., Platero, J.S., Shuman, J. and Calame, K. (1990) 'Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP', *Genes Dev*, 4(8), pp. 1404-15.
Romana, S.P., Le Coniat, M. and Berger, R. (1994) 't(12;21): a new recurrent translocation in acute lymphoblastic leukemia', *Genes Chromosomes Cancer*, 9(3), pp. 186-91.
Romana, S.P., Poirel, H., Leconiat, M., Flexor, M.A., Mauchauffe, M., Jonveaux, P., Macintyre, E.A., Berger, R. and Bernard, O.A. (1995) 'High frequency of t(12;21) in childhood B-lineage acute lymphoblastic leukemia', *Blood*, 86(11), pp. 4263-9.
Romano, M.F., Petrella, A., Bisogni, R., Turco, M.C. and Venuta, S. (2003) 'Effect of NF-kappaB/Rel inhibition on spontaneous vs chemotherapy-induced apoptosis in AML and normal cord blood CD34+ cells', *Leukemia*, 17(6), pp. 1190-2.
Ron, D. and Habener, J.F. (1992) 'CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription', *Genes Dev*, 6(3), pp. 439-53.
Rosenbauer, F. and Tenen, D.G. (2007) 'Transcription factors in myeloid development: balancing differentiation with transformation', *Nat Rev Immunol*, 7(2), pp. 105-17.
Rosenwald, A., Wright, G., Wiestner, A., Chan, W.C., Connors, J.M., Campo, E., Gascoyne, R.D., Grogan, T.M., Muller-Hermelink, H.K., Smeland, E.B., Chiorazzi, M., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Henrickson, S., Yang, L., Powell, J., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Montserrat, E., Bosch, F., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Fisher, R.I., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Holte, H., Delabie, J. and Staudt, L.M. (2003) 'The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma', *Cancer Cell*, 3(2), pp. 185-97.
Rowe, J.M., Buck, G., Burnett, A.K., Chopra, R., Wiernik, P.H., Richards, S.M., Lazarus, H.M., Franklin, I.M., Litzow, M.R., Ciobanu, N., Prentice, H.G., Durrant, J., Tallman, M.S., Goldstone, A.H., Ecog and Party, M.N.A.L.W. (2005) 'Induction therapy for adults with acute lymphoblastic leukemia: results of more than 1500 patients from the international ALL trial: MRC UKALL XII/ECOG E2993', *Blood*, 106(12), pp. 3760-7.
Russell, L.J., Akasaka, T., Majid, A., Sugimoto, K.J., Loraine Karran, E., Nagel, I., Harder, L., Claviez, A., Gesk, S., Moorman, A.V., Ross, F., Mazzullo, H., Strefford, J.C., Siebert, R., Dyer, M.J. and Harrison, C.J. (2008) 't(6;14)(p22;q32): a new recurrent IGH@ translocation involving ID4 in B-cell precursor acute lymphoblastic leukemia (BCP-ALL)', *Blood*, 111(1), pp. 387-91.
Russell, L.J., Capasso, M., Vater, I., Akasaka, T., Bernard, O.A., Calasanz, M.J., Chandrasekaran, T., Chapiro, E., Gesk, S., Griffiths, M., Guttery, D.S.,

Haferlach, C., Harder, L., Heidenreich, O., Irving, J., Kearney, L., Nguyen-Khac, F., Machado, L., Minto, L., Majid, A., Moorman, A.V., Morrison, H., Rand, V., Strefford, J.C., Schwab, C., Tönnies, H., Dyer, M.J.S., Siebert, R. and Harrison, C.J. (2009) 'Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia', *Blood*, 114(13), pp. 2688-2698.
Russell, L.J., Enshaei, A., Jones, L., Erhorn, A., Masic, D., Bentley, H., Laczko, K.S., Fielding, A.K., Goldstone, A.H., Goulden, N., Mitchell, C.D., Wade, R., Vora, A., Moorman, A.V. and Harrison, C.J. (2014) 'IGH@ translocations are prevalent in teenagers and young adults with acute lymphoblastic leukemia and are associated with a poor outcome', *J Clin Oncol*, 32(14), pp. 1453-62.
Sadasivam, S. and DeCaprio, J.A. (2013) 'The DREAM complex: master coordinator of cell cycle-dependent gene expression', *Nat Rev Cancer*, 13(8), pp. 585-95.
Sahin, A.O. and Buitenhuis, M. (2012) 'Molecular mechanisms underlying adhesion and migration of hematopoietic stem cells', *Cell Adh Migr*, 6(1), pp. 39-48.
Saisanit, S. and Sun, X.H. (1997) 'Regulation of the pro-B-cell-specific enhancer of the Id1 gene involves the C/EBP family of proteins', *Mol Cell Biol*, 17(2), pp. 844-50.
Salemi, D., Cammarata, G., Agueli, C., Augugliaro, L., Corrado, C., Bica, M.G., Raimondo, S., Marfia, A., Randazzo, V., Dragotto, P., Di Raimondo, F., Alessandro, R., Fabbiano, F. and Santoro, A. (2015) 'miR-155 regulative network in FLT3 mutated acute myeloid leukemia', *Leuk Res*, 39(8), pp. 883-96.
Sanford, D.C. and DeWille, J.W. (2005) 'C/EBPdelta is a downstream mediator of IL-6 induced growth inhibition of prostate cancer cells', *Prostate*, 63(2), pp. 143-54.
Sarkar, T.R., Sharan, S., Wang, J., Pawar, S.A., Cantwell, C.A., Johnson, P.F., Morrison, D.K., Wang, J.M. and Sterneck, E. (2012) 'Identification of a Src tyrosine kinase/SIAH2 E3 ubiquitin ligase pathway that regulates C/EBPdelta expression and contributes to transformation of breast tumor cells', *Mol Cell Biol*, 32(2), pp. 320-32.
Schepers, H., Eggen, B.J., Schuringa, J.J. and Vellenga, E. (2006) 'Constitutive activation of NF-kappa B is not sufficient to disturb normal steady-state hematopoiesis', *Haematologica*, 91(12), pp. 1710-1.
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.
Schrappe, M., Reiter, A., Ludwig, W.D., Harbott, J., Zimmermann, M., Hiddemann, W., Niemeyer, C., Henze, G., Feldges, A., Zintl, F., Kornhuber, B., Ritter, J., Welte, K., Gadner, H. and Riehm, H. (2000) 'Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM Study Group', *Blood*, 95(11), pp. 3310-22.
Schrem, H., Klempnauer, J. and Borlak, J. (2004) 'Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation', *Pharmacol Rev*, 56(2), pp. 291-330.
Schwab, C.J., Chilton, L., Morrison, H., Jones, L., Al-Shehhi, H., Erhorn, A., Russell, L.J., Moorman, A.V. and Harrison, C.J. (2013) 'Genes commonly

deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features', *Haematologica*.
Schwab, C.J., Jones, L.R., Morrison, H., Ryan, S.L., Yigittop, H., Schouten, J.P. and Harrison, C.J. (2010a) 'Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia', *Genes, Chromosomes and Cancer*, 49(12), pp. 1104-1113.
Schwab, C.J., Jones, L.R., Morrison, H., Ryan, S.L., Yigittop, H., Schouten, J.P. and Harrison, C.J. (2010b) 'Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia', *Genes Chromosomes Cancer*, 49(12), pp. 1104-13.
Schwartzberg, P.L., Stall, A.M., Hardin, J.D., Bowdish, K.S., Humaran, T., Boast, S., Harbison, M.L., Robertson, E.J. and Goff, S.P. (1991) 'Mice homozygous for the ablm1 mutation show poor viability and depletion of selected B and T cell populations', *Cell*, 65(7), pp. 1165-75.
Schwieger, M., Schuler, A., Forster, M., Engelmann, A., Arnold, M.A., Delwel, R., Valk, P.J., Lohler, J., Slany, R.K., Olson, E.N. and Stocking, C. (2009) 'Homing and invasiveness of MLL/ENL leukemic cells is regulated by MEF2C', *Blood*, 114(12), pp. 2476-88.
Scientific, F. (1996) *Heat Inactivation— Are You Wasting Your Time?* (1). Art to Science. [Online]. Available at: http://fscimage.fishersci.com/cmsassets/downloads/segment/Scientific/pdf/Cell_Culture/Application_Notes/heat_inactivation.pdf?&storeId=10652.
Screpanti, I., Romani, L., Musiani, P., Modesti, A., Fattori, E., Lazzaro, D., Sellitto, C., Scarpa, S., Bellavia, D., Lattanzio, G. and et al. (1995) 'Lymphoproliferative disorder and imbalanced T-helper response in C/EBP beta-deficient mice', *EMBO J*, 14(9), pp. 1932-41.
Seabright, M. (1971) 'A rapid banding technique for human chromosomes', *Lancet*, 2(7731), pp. 971-2.
Sebastian, T., Malik, R., Thomas, S., Sage, J. and Johnson, P.F. (2005) 'C/EBPbeta cooperates with RB:E2F to implement Ras(V12)-induced cellular senescence', *EMBO J*, 24(18), pp. 3301-12.
Serrano, M., Lee, H., Chin, L., Cordon-Cardo, C., Beach, D. and DePinho, R.A. (1996) 'Role of the INK4a locus in tumor suppression and cell mortality', *Cell*, 85(1), pp. 27-37.
Sherr, C.J. (2012) 'Ink4-Arf locus in cancer and aging', *Wiley Interdiscip Rev Dev Biol*, 1(5), pp. 731-41.
Shimizu, T., Marusawa, H., Endo, Y. and Chiba, T. (2012) 'Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis', *Cancer Sci*, 103(7), pp. 1201-6.
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.
Sicinska, E., Aifantis, I., Le Cam, L., Swat, W., Borowski, C., Yu, Q., Ferrando, A.A., Levin, S.D., Geng, Y., von Boehmer, H. and Sicinski, P. (2003)

'Requirement for cyclin D3 in lymphocyte development and T cell leukemias', *Cancer Cell*, 4(6), pp. 451-61.
Sigvardsson, M., Clark, D.R., Fitzsimmons, D., Doyle, M., Akerblad, P., Breslin, T., Bilke, S., Li, R., Yeamans, C., Zhang, G. and Hagman, J. (2002) 'Early B-cell factor, E2A, and Pax-5 cooperate to activate the early B cell-specific mb-1 promoter', *Mol Cell Biol*, 22(24), pp. 8539-51.
Skinner, J., Mee, T.J., Blackwell, R.P., Maslanyj, M.P., Simpson, J., Allen, S.G., Day, N.E., Cheng, K.K., Gilman, E., Williams, D., Cartwright, R., Craft, A., Birch, J.M., Eden, O.B., McKinney, P.A., Deacon, J., Peto, J., Beral, V., Roman, E., Elwood, P., Alexander, F.E., Mott, M., Chilvers, C.E., Muir, K., Doll, R., Taylor, C.M., Greaves, M., Goodhead, D., Fry, F.A., Adams, G., Law, G. and United Kingdom Childhood Cancer Study, I. (2002) 'Exposure to power frequency electric fields and the risk of childhood cancer in the UK', *Br J Cancer*, 87(11), pp. 1257-66.
Smith, M.L., Cavenagh, J.D., Lister, T.A. and Fitzgibbon, J. (2004) 'Mutation of CEBPA in Familial Acute Myeloid Leukemia', *New England Journal of Medicine*, 351(23), pp. 2403-2407.
Strefford, J.C., van Delft, F.W., Robinson, H.M., Worley, H., Yiannikouris, O., Selzer, R., Richmond, T., Hann, I., Bellotti, T., Raghavan, M., Young, B.D., Saha, V. and Harrison, C.J. (2006) 'Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21', *Proc Natl Acad Sci U S A*, 103(21), pp. 8167-72.
Studzinski, G.P., Wang, X., Ji, Y., Wang, Q., Zhang, Y., Kutner, A. and Harrison, J.S. (2005) 'The rationale for deltanoids in therapy for myeloid leukemia: role of KSR-MAPK-C/EBP pathway', *J Steroid Biochem Mol Biol*, 97(1-2), pp. 47-55.
Sun, X.H. (1994) 'Constitutive expression of the Id1 gene impairs mouse B cell development', *Cell*, 79(5), pp. 893-900.
Sutcliffe, M.J., Shuster, J.J., Sather, H.N., Camitta, B.M., Pullen, J., Schultz, K.R., Borowitz, M.J., Gaynon, P.S., Carroll, A.J. and Heerema, N.A. (2005) 'High concordance from independent studies by the Children's Cancer Group (CCG) and Pediatric Oncology Group (POG) associating favorable prognosis with combined trisomies 4, 10, and 17 in children with NCI Standard-Risk B-precursor Acute Lymphoblastic Leukemia: a Children's Oncology Group (COG) initiative', *Leukemia*, 19(5), pp. 734-40.
Tanaka, T., Yoshida, N., Kishimoto, T. and Akira, S. (1997) 'Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene', *EMBO J*, 16(24), pp. 7432-43.
Tang, Q.Q., Otto, T.C. and Lane, M.D. (2003) 'Mitotic clonal expansion: a synchronous process required for adipogenesis', *Proc Natl Acad Sci U S A*, 100(1), pp. 44-9.
Tasian, S.K., Doral, M.Y., Borowitz, M.J., Wood, B.L., Chen, I.M., Harvey, R.C., Gastier-Foster, J.M., Willman, C.L., Hunger, S.P., Mullighan, C.G. and Loh, M.L. (2012) 'Aberrant STAT5 and PI3K/mTOR pathway signaling occurs in human CRLF2-rearranged B-precursor acute lymphoblastic leukemia', *Blood*, 120(4), pp. 833-42.
Teh, M.-T., Blaydon, D., Chaplin, T., Foot, N.J., Skoulakis, S., Raghavan, M., Harwood, C.A., Proby, C.M., Philpott, M.P., Young, B.D. and Kelsell, D.P. (2005) 'Genomewide Single Nucleotide Polymorphism Microarray Mapping in Basal Cell Carcinomas Unveils Uniparental Disomy as a Key Somatic Event', *Cancer Res*, 65(19), pp. 8597-8603.

Thangaraju, M., Rudelius, M., Bierie, B., Raffeld, M., Sharan, S., Hennighausen, L., Huang, A.M. and Sterneck, E. (2005) 'C/EBPdelta is a crucial regulator of pro-apoptotic gene expression during mammary gland involution', *Development*, 132(21), pp. 4675-85.
Theocharides, A.P., Dobson, S.M., Laurenti, E., Notta, F., Voisin, V., Cheng, P.Y., Yuan, J.S., Guidos, C.J., Minden, M.D., Mullighan, C.G., Torlakovic, E. and Dick, J.E. (2015) 'Dominant-negative Ikaros cooperates with BCR-ABL1 to induce human acute myeloid leukemia in xenografts', *Leukemia*, 29(1), pp. 177-87.
Thomassin, H., Hamel, D., Bernier, D., Guertin, M. and Belanger, L. (1992) 'Molecular cloning of two C/EBP-related proteins that bind to the promoter and the enhancer of the alpha 1-fetoprotein gene. Further analysis of C/EBP beta and C/EBP gamma', *Nucleic Acids Res*, 20(12), pp. 3091-8.
Thompson, E.C., Cobb, B.S., Sabbattini, P., Meixlsperger, S., Parelho, V., Liberg, D., Taylor, B., Dillon, N., Georgopoulos, K., Jumaa, H., Smale, S.T., Fisher, A.G. and Merkenschlager, M. (2007) 'Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits', *Immunity*, 26(3), pp. 335-44.
Trinchieri, G. (2012) 'Cancer and inflammation: an old intuition with rapidly evolving new concepts', *Annu Rev Immunol*, 30, pp. 677-706.
Tsukada, J., Saito, K., Waterman, W.R., Webb, A.C. and Auron, P.E. (1994) 'Transcription factors NF-IL6 and CREB recognize a common essential site in the human prointerleukin 1 beta gene', *Mol Cell Biol*, 14(11), pp. 7285-97.
Tsukada, J., Yoshida, Y., Kominato, Y. and Auron, P.E. (2011) 'The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation', *Cytokine*, 54(1), pp. 6-19.
Tybulewicz, V.L., Crawford, C.E., Jackson, P.K., Bronson, R.T. and Mulligan, R.C. (1991) 'Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene', *Cell*, 65(7), pp. 1153-63.
Tzeng, Y.S., Li, H., Kang, Y.L., Chen, W.C., Cheng, W.C. and Lai, D.M. (2011) 'Loss of Cxcl12/Sdf-1 in adult mice decreases the quiescent state of hematopoietic stem/progenitor cells and alters the pattern of hematopoietic regeneration after myelosuppression', *Blood*, 117(2), pp. 429-39.
Uy, G.L., Rettig, M.P., Motabi, I.H., McFarland, K., Trinkaus, K.M., Hladnik, L.M., Kulkarni, S., Abboud, C.N., Cashen, A.F., Stockerl-Goldstein, K.E., Vij, R., Westervelt, P. and DiPersio, J.F. (2012) 'A phase 1/2 study of chemosensitization with the CXCR4 antagonist plerixafor in relapsed or refractory acute myeloid leukemia', *Blood*, 119(17), pp. 3917-24.
Van Etten, R.A. (2007) 'Aberrant cytokine signaling in leukemia', *Oncogene*, 26(47), pp. 6738-49.
van Gent, D.C. and van der Burg, M. (2007) 'Non-homologous end-joining, a sticky affair', *Oncogene*, 26(56), pp. 7731-40.
van Zelm, M.C., van der Burg, M., de Ridder, D., Barendregt, B.H., de Haas, E.F.E., Reinders, M.J.T., Lankester, A.C., Révész, T., Staal, F.J.T. and van Dongen, J.J.M. (2005) 'Ig Gene Rearrangement Steps Are Initiated in Early Human Precursor B Cell Subsets and Correlate with Specific Transcription Factor Expression', *The Journal of Immunology*, 175(9), pp. 5912-5922.
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.
Vegesna, V., Takeuchi, S., Hofmann, W.K., Ikezoe, T., Tavor, S., Krug, U., Fermin, A.C., Heaney, A., Miller, C.W. and Koeffler, H.P. (2002) 'C/EBP-beta, C/EBP-delta, PU.1, AML1 genes: mutational analysis in 381 samples of hematopoietic and solid malignancies', *Leuk Res*, 26(5), pp. 451-7.
Vettermann, C., Herrmann, K. and Jack, H.M. (2006) 'Powered by pairing: the surrogate light chain amplifies immunoglobulin heavy chain signaling and pre-selects the antibody repertoire', *Semin Immunol*, 18(1), pp. 44-55.
Vettermann, C. and Schlissel, M.S. (2010) 'Allelic exclusion of immunoglobulin genes: models and mechanisms', *Immunol Rev*, 237(1), pp. 22-42.
Vinson, C.R., Hai, T. and Boyd, S.M. (1993) 'Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design', *Genes Dev*, 7(6), pp. 1047-58.
Wagner, K., Zhang, P., Rosenbauer, F., Drescher, B., Kobayashi, S., Radomska, H.S., Kutok, J.L., Gilliland, D.G., Krauter, J. and Tenen, D.G. (2006) 'Absence of the transcription factor CCAAT enhancer binding protein alpha results in loss of myeloid identity in bcr/abl-induced malignancy', *Proc Natl Acad Sci U S A*, 103(16), pp. 6338-43.
Wall, L., Destroismaisons, N., Delvoye, N. and Guy, L.G. (1996) 'CAAT/enhancer-binding proteins are involved in beta-globin gene expression and are differentially expressed in murine erythroleukemia and K562 cells', *J Biol Chem*, 271(28), pp. 16477-84.
Wang, H., Iakova, P., Wilde, M., Welm, A., Goode, T., Roesler, W.J. and Timchenko, N.A. (2001) 'C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4', *Mol Cell*, 8(4), pp. 817-28.
Wang, J., Sarkar, T.R., Zhou, M., Sharan, S., Ritt, D.A., Veenstra, T.D., Morrison, D.K., Huang, A.M. and Sterneck, E. (2010) 'CCAAT/enhancer binding protein delta (C/EBPdelta, CEBPD)-mediated nuclear import of FANCD2 by IPO4 augments cellular response to DNA damage', *Proc Natl Acad Sci U S A*, 107(37), pp. 16131-6.
Wang, J.F., Park, I.W. and Groopman, J.E. (2000) 'Stromal cell-derived factor-1alpha stimulates tyrosine phosphorylation of multiple focal adhesion proteins and induces migration of hematopoietic progenitor cells: roles of phosphoinositide-3 kinase and protein kinase C', *Blood*, 95(8), pp. 2505-13.
Wang, J.M., Ko, C.Y., Chen, L.C., Wang, W.L. and Chang, W.C. (2006) 'Functional role of NF-IL6beta and its sumoylation and acetylation modifications in promoter activation of cyclooxygenase 2 gene', *Nucleic Acids Res*, 34(1), pp. 217-31.
Wang, N.D., Finegold, M.J., Bradley, A., Ou, C.N., Abdelsayed, S.V., Wilde, M.D., Taylor, L.R., Wilson, D.R. and Darlington, G.J. (1995) 'Impaired energy homeostasis in C/EBP alpha knockout mice', *Science*, 269(5227), pp. 1108-12.
Wang, Q.F. and Friedman, A.D. (2002) 'CCAAT/enhancer-binding proteins are required for granulopoiesis independent of their induction of the granulocyte colony-stimulating factor receptor', *Blood*, 99(8), pp. 2776-85.
Wang, X., Fan, M., Kalis, S., Wei, L. and Scharff, M.D. (2014) 'A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation', *Nat Commun*, 5, p. 4137.
Warnock, J.N., Daigre, C. and Al-Rubeai, M. (2011) 'Introduction to viral vectors', *Methods Mol Biol*, 737, pp. 1-25.
Werner, M., Kraunus, J., Baum, C. and Brocker, T. (2004) 'B-cell-specific transgene expression using a self-inactivating retroviral vector with human

CD19 promoter and viral post-transcriptional regulatory element', *Gene Ther*, 11(12), pp. 992-1000.
Weston, B.W., Hayden, M.A., Roberts, K.G., Bowyer, S., Hsu, J., Fedoriw, G., Rao, K.W. and Mullighan, C.G. (2013) 'Tyrosine kinase inhibitor therapy induces remission in a patient with refractory EBF1-PDGFRB-positive acute lymphoblastic leukemia', *J Clin Oncol*, 31(25), pp. e413-6.
Williams, R.T., Roussel, M.F. and Sherr, C.J. (2006) 'Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia', *Proc Natl Acad Sci U S A*, 103(17), pp. 6688-93.
Willis, T.G. and Dyer, M.J. (2000) 'The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies', *Blood*, 96(3), pp. 808-22.
Wu, Q., He, J., Fang, J. and Hong, M. (2010) 'Antitumor effect of betulinic acid on human acute leukemia K562 cells in vitro', *J Huazhong Univ Sci Technolog Med Sci*, 30(4), pp. 453-7.
Wu, S.R., Li, C.F., Hung, L.Y., Huang, A.M., Tseng, J.T., Tsou, J.H. and Wang, J.M. (2011) 'CCAAT/enhancer-binding protein delta mediates tumor necrosis factor alpha-induced Aurora kinase C transcription and promotes genomic instability', *J Biol Chem*, 286(33), pp. 28662-70.
Xavier, A.C. and Taub, J.W. (2010) 'Acute leukemia in children with Down syndrome', *Haematologica*, 95(7), pp. 1043-5.
Xia, H., Nho, R.S., Kahm, J., Kleidon, J. and Henke, C.A. (2004) 'Focal adhesion kinase is upstream of phosphatidylinositol 3-kinase/Akt in regulating fibroblast survival in response to contraction of type I collagen matrices via a beta 1 integrin viability signaling pathway', *J Biol Chem*, 279(31), pp. 33024-34.
Xie, H., Ye, M., Feng, R. and Graf, T. (2004) 'Stepwise Reprogramming of B Cells into Macrophages', *Cell*, 117(5), pp. 663-676.
Yamanaka, R., Kim, G.D., Radomska, H.S., Lekstrom-Himes, J., Smith, L.T., Antonson, P., Tenen, D.G. and Xanthopoulos, K.G. (1997) 'CCAAT/enhancer binding protein epsilon is preferentially up-regulated during granulocytic differentiation and its functional versatility is determined by alternative use of promoters and differential splicing', *Proc Natl Acad Sci U S A*, 94(12), pp. 6462-7.
Yan, C., Zhu, M., Staiger, J., Johnson, P.F. and Gao, H. (2012) 'C5a-regulated CCAAT/enhancer-binding proteins beta and delta are essential in Fcgamma receptor-mediated inflammatory cytokine and chemokine production in macrophages', *J Biol Chem*, 287(5), pp. 3217-30.
Yao, D.M., Qian, J., Lin, J., Wang, Y.L., Chen, Q., Qian, Z., Li, Y., Wang, C.Z. and Yang, J. (2011) 'Aberrant methylation of CCAAT/enhancer binding protein zeta promoter in acute myeloid leukemia', *Leuk Res*, 35(7), pp. 957-60.
Ye, M., Zhang, H., Amabile, G., Yang, H., Staber, P.B., Zhang, P., Levantini, E., Alberich-Jorda, M., Zhang, J., Kawasaki, A. and Tenen, D.G. (2013) 'C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence', *Nat Cell Biol*, 15(4), pp. 385-94.
You, D., Xin, J., Volk, A., Wei, W., Schmidt, R., Scurti, G., Nand, S., Breuer, E.K., Kuo, P.C., Breslin, P., Kini, A.R., Nishimura, M.I., Zeleznik-Le, N.J. and Zhang, J. (2015) 'FAK mediates a compensatory survival signal parallel to PI3K-AKT in PTEN-null T-ALL cells', *Cell Rep*, 10(12), pp. 2055-68.
Zahnow, C.A. (2009) 'CCAAT/enhancer-binding protein beta: its role in breast cancer and associations with receptor tyrosine kinases', *Expert Rev Mol Med*, 11, p. e12.

Zahnow, C.A., Cardiff, R.D., Laucirica, R., Medina, D. and Rosen, J.M. (2001) 'A role for CCAAT/enhancer binding protein beta-liver-enriched inhibitory protein in mammary epithelial cell proliferation', *Cancer Res*, 61(1), pp. 261-9.
Zahnow, C.A., Younes, P., Laucirica, R. and Rosen, J.M. (1997) 'Overexpression of C/EBPbeta-LIP, a naturally occurring, dominant-negative transcription factor, in human breast cancer', *J Natl Cancer Inst*, 89(24), pp. 1887-91.
Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonthaler, H.B., Kenner, L., Tschachler, E. and Wagner, E.F. (2008) 'Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease', *Arthritis Res Ther*, 10(1), p. 201.
Zhang, D.-E., Zhang, P., Wang, N.-d., Hetherington, C.J., Darlington, G.J. and Tenen, D.G. (1997) 'Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice', *Proceedings of the National Academy of Sciences*, 94(2), pp. 569-574.
Zhang, J., Adrian, F.J., Jahnke, W., Cowan-Jacob, S.W., Li, A.G., Iacob, R.E., Sim, T., Powers, J., Dierks, C., Sun, F., Guo, G.R., Ding, Q., Okram, B., Choi, Y., Wojciechowski, A., Deng, X., Liu, G., Fendrich, G., Strauss, A., Vajpai, N., Grzesiek, S., Tuntland, T., Liu, Y., Bursulaya, B., Azam, M., Manley, P.W., Engen, J.R., Daley, G.Q., Warmuth, M. and Gray, N.S. (2010) 'Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors', *Nature*, 463(7280), pp. 501-6.
Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., Akashi, K. and Tenen, D.G. (2004) 'Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha', *Immunity*, 21(6), pp. 853-63.
Zhang, T., He, Y.M., Wang, J.S., Shen, J., Xing, Y.Y. and Xi, T. (2011) 'Ursolic acid induces HL60 monocytic differentiation and upregulates C/EBPbeta expression by ERK pathway activation', *Anticancer Drugs*, 22(2), pp. 158-65.
Ziemin-van der Poel, S., McCabe, N.R., Gill, H.J., Espinosa, R., 3rd, Patel, Y., Harden, A., Rubinelli, P., Smith, S.D., LeBeau, M.M., Rowley, J.D. and et al. (1991) 'Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias', *Proc Natl Acad Sci U S A*, 88(23), pp. 10735-9.